LNTP: An End-to-End Online Prediction Model for Network Traffic

Lianming Zhang, Huan Zhang, Qian Tang, Pingping Dong, Zhen Zhao, Yehua Wei, Jing Mei, and Kaiping Xue

Abstract

As network data keeps getting bigger, deep learning is coming to play a key role in network design and management. Meanwhile, accurate network traffic prediction is of critical importance for network management that is implemented to improve the quality of service (QoS) for users. However, the performance of existing network traffic prediction methods is still poor due to three challenges: complicated characteristics of network traffic, dynamics of traffic patterns caused by different network applications, and a complex set of variations like burstiness. In this article, we propose a long short-term memory (LSTM) based network traffic prediction (LNTP) model, which aims to forecast network traffic timely and accurately. The model can be divided into two parts, namely, wavelet transform and LSTM. The working process of LNTP falls into three stages, i.e., data acquisition, model training, and online learning and prediction. In addition, to avoid the negative incentives to models caused by the burstiness and adapt to the changing trend of the network traffic, a weight optimization algorithm of the neural network named sliding window gradient descent (SWGD), is also proposed. Extensive experiments based on two real-world network traffic datasets demonstrate that our model outperforms the state-of-the-art network traffic prediction models by more than 29 percent.

Introduction

With the development of network technology and the rise of various kinds of Internet services, the demand for network traffic has grown rapidly around the world. According to the technical report from Internet live stats [1], the number of Internet users worldwide surpassed 4 billion by June 1, 2018, and network traffic reaches 67 TB in one second at 11:13:37 on December 1, 2018. Accurate and timely network traffic prediction is important for bandwidth allocation, congestion control, admission control [2] and privacy-preserving routing [3]. Some research efforts have been made to improve the performance of traffic forecasting, and existing prediction models can be classified into three categories, namely time series based models, machine learning models, and fusion models. Shu *et al.* [4] proposed a prediction model based on the seasonal autoregressive integrated moving average (ARIMA) models to forecast the wireless traffic. Generalized autoregressive conditional heteroskedasticity (GARCH) [5] was proposed which is a non-linear time series model and can capture the burstiness of the network traffic. Due to the non-linear characteristics of network traffic, linear models cannot fit network traffic well. Thus, some non-linear models based on machine learning have been proposed and applied to network traffic prediction. A deep traffic predictor (DeepTP) based on deep learning was proposed in [6]. The model consists of two parts: the feature extraction model is used to extract the spatial dependence and the features contained in some external information. The time series model is used to fit the distribution characteristics of flow data over time. Combining Markov chains with tensors to implement predictions, Liu *et al*. [7] focused on proposing a multivariate multi-order Markov transition to realize multi-modal accurate predictions.

However, the characteristics of network traffic are becoming more and more complex, making it difficult for a single prediction method to capture all kinds of characteristics of the traffic data. Therefore, scholars have proposed a series of fusion models for network traffic prediction. Dai *et al.* [8] proposed empirical mode decomposition (EMD) based on multi-model prediction (EMD-MMP) for network traffic prediction which first decomposes the network traffic series into different modes with different frequencies by EMD. Then, different components are predicted by auto-regressive and moving average (ARMA) and support vector regression (SVR) methods separately. Nie *et al.* [9] proposed a network traffic prediction method based on a deep belief network (DBN) and a Gaussian model (DBNG). The method first adopts a discrete wavelet transform to extract the low-pass component of network traffic. Then it utilizes DBN and the Gaussian model to model the extracted low-pass component and the rest high-pass component, respectively. Huang *et al.* [10] proposed a traffic forecasting model based on modified ensemble EMD (MEEMD) and quantum neural network (QNN). The MEEMD method is employed to decompose the traffic data sequence into intrinsic mode function (IMF) component. The QNN is adopted to forecast the IMF components. Chen *et al*. [11] proposed a new hybrid network traffic prediction method (EPSVM) primarily based on EMD, particle swarm optimization (PSO), and SVM. The EPSVM utilizes EMD to eliminate the impact of noise signals. SVM is applied to model training and fitting, and the parameters of SVM are optimized by PSO.

Although existing fusion models are superior to the single models, predicting different compo-

Digital Object Identifier:
10.1109/MNET.011.1900647

Lianming Zhang, Huan Zhang, Qian Tang, Pingping Dong (corresponding author), Zhen Zhao, Yehua Wei, and Jing Mei are with Hunan Normal University; Kaiping Xue is with the University of Science and Technology of China.

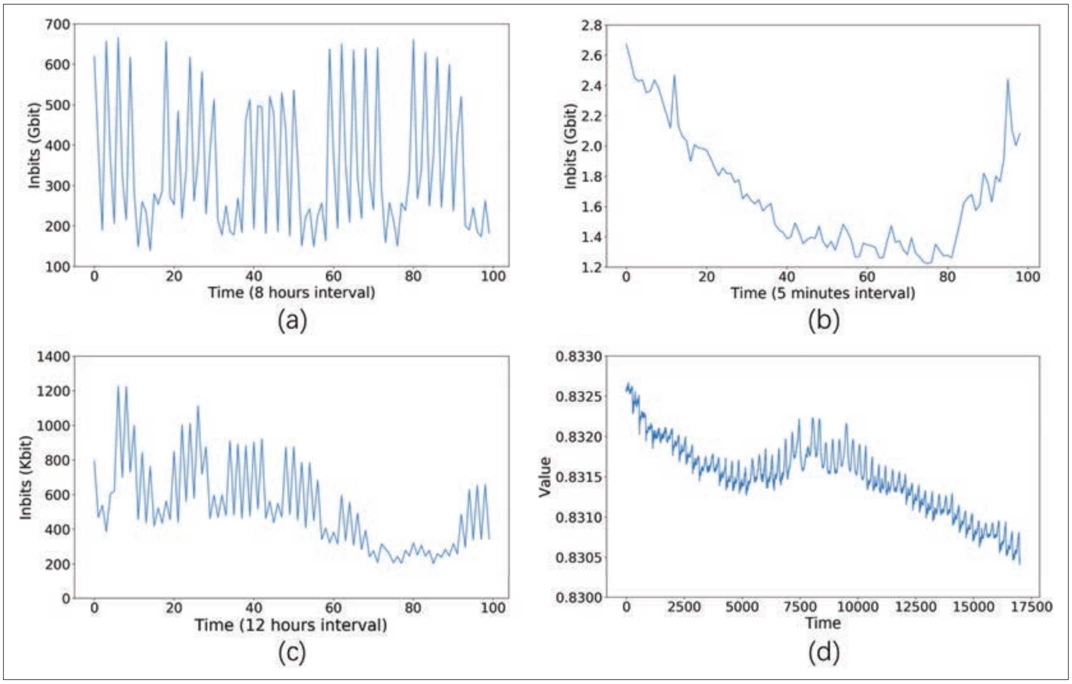

FIGURE 1. Real network traffic. The original data comes from two private ISPs, namely, the United Kingdom academic network backbone and the European cities network backbone: a) real network traffic with 8 hours as sampling interval; b) real network traffic with 5 minutes as the sampling interval; c) real network traffic with 12 hours as the sampling interval; d) changes of a certain weight value in LSTM during the online learning process with SGD.

nents with different models may ignore the correlation among the components decomposed by the original traffic, which will influence the accuracy of the prediction results. The recurrent neural network (RNN) has been widely used in image classification, natural language processing, speech recognition, audio processing, and machine vision. Long short-term memory (LSTM) network, which is based on RNN and overcomes the natural defects of RNN in terms of gradient explosion and gradient disappearance, is more and more widely used in time series problems. This article aims to improve the prediction accuracy of the network traffic and proposes an LSTM based on the network traffic prediction model (LNTP). The main contributions of the article are as follows.

A novel end-to-end deep learning-based online traffic prediction architecture named LNTP is proposed. LNTP can effectively capture the various characteristics contained in the network traffic data by utilizing both wavelet transform and improved LSTM. The components decomposed from the original network traffic data obtained by wavelet transform are synchronously input into LSTM. Through the training by LSTM, the model cannot only learn the internal laws of each component but also capture the connections between the various components, and thus it can learn more intrinsic features of network traffic data.

To improve the accuracy of the prediction model, the proposed LNTP contains not only the training stage but also defines the calculation process during the online learning phase to adapt to the real-time dynamics of network traffic, and thus maintain high prediction accuracy for a long time.

A weight optimization algorithm named SWGD (sliding window gradient descent) is proposed. As the number of training samples is limited, the weight fluctuation caused by network traffic burstiness during the online learning process is particularly obvious. SWGD can effectively avoid the negative incentives to the proposed models caused by the burstiness of the network traffic and thus LNTP can adapt to the changing trend of the network traffic and achieve high accuracy.

The remainder of this article is organized as follows. We first introduce the challenges of our investigated problem. We then describe the details of LNTP, online learning, and SWGD Algorithm. Extensive experiments are then conducted. Finally, we conclude the article.

Problems and Challenges

Network traffic can be predicted with time series based models by capturing the correlation, periodicity, randomness, and other characteristics inherent in the traffic data. Real network traffic is given in Fig. 1, and it comes from two private Internet service providers (ISPs), namely, the United Kingdom academic network backbone and the European cities network backbone. However, the prediction model must have the ability to adapt to new traffic patterns which may change frequently with varying applications. Also, the model should be able to tackle the burstiness of network traffic which may generate negative incentives during the learning process of the model and affect the prediction accuracy of the model.

Complicated Characteristics of Network Traffic

The network traffic changing trend of 100 samples with a sample interval of eight hours and five minutes is depicted in Figs. 1a and 1b. The original data comes from the European cities' network backbone. It can be seen clearly in Fig. 1a that the network traffic data has a cyclical character

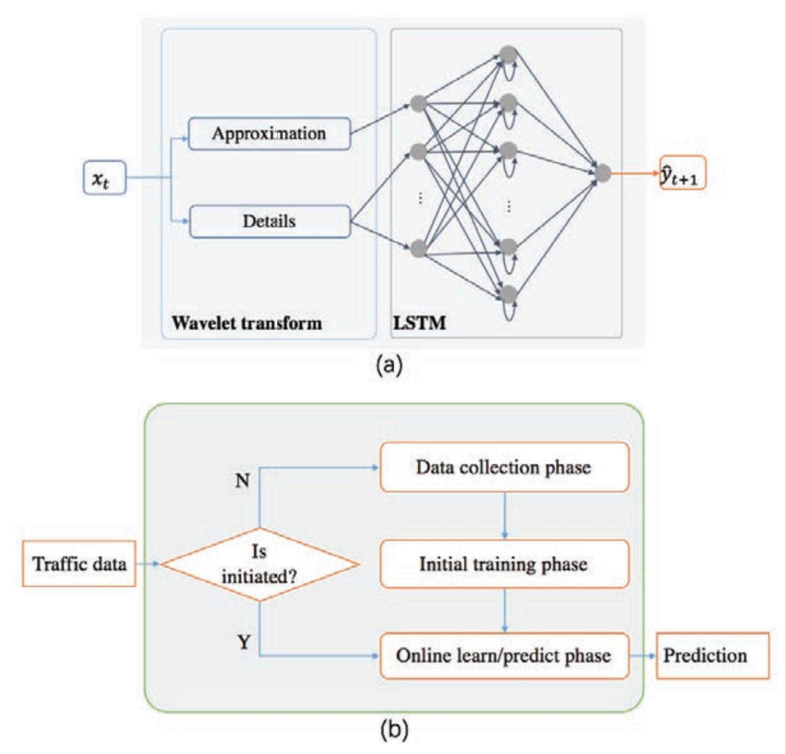

FIGURE 2. The LNTP model: a) the spatial structure of the LNTP model; b) the logic division of the LNTP model.

from the macroscopic view, due to people's living habits. Figure 1b shows the characteristics of frequent burstiness of network traffic data from the micro-level. The primary problem that researchers face is how to make the model capture the various characteristics contained in the network traffic data. Wavelet transform is a multi-resolution analysis method. The discrete wavelet transform processes discrete-time signals and it can be used to represent a signal as the sum of multiple wavelet functions that are local in time and space. After the network traffic is processed by wavelet transform, it is decomposed into multiple components with different frequencies. The components of different frequencies will carry different trends of time granularity so that the neural network can more accurately learn the changing rules of network traffic. In this article, the proposed LNTP model can be divided into two parts, including wavelet transform and improved LSTM. The original network traffic data is first decomposed by wavelet transform, and then the components obtained by wavelet transform decomposition are synchronously input into LSTM. Finally, through the training by LSTM, the model cannot only learn the internal laws of each component but also capture the connections between the various components, and thus it can learn more intrinsic features of network traffic data.

Dynamics of Traffic Patterns Caused by Different Network Applications

The changing pattern of network traffic varies with different kinds of Internet applications which develop rapidly. Figure 1c shows the trend of network traffic with a total of 100 samples whose sampling interval is 12 hours. The original data comes from the United Kingdom's academic network backbone. From Fig. 1c, notable changes in the traffic pattern have taken place from the 70th sample. Thus, if the model does not keep learning, the accuracy of the prediction model which is trained based on the historic records may decrease. In this article, LNTP is logically divided into three stages, namely, data acquisition, initialization training, and online learning and prediction. The LNTP defines the calculation process during the online learning phase. This enables the model to adapt to new network traffic changes in real-time, and maintains high accuracy in long-term prediction.

Burstiness of Network Traffic

Online learning is defined as an algorithm for generating a series of models on a given training data stream [12]. The model for the next moment only depends on the model at the current time and limited data streams. As the number of training samples is limited, the weight fluctuation caused by network traffic burstiness during the online learning process will be particularly obvious. Figure 1d shows the change of one weight of a certain neuron in LSTM with the stochastic gradient descent (SGD) optimization algorithm during online learning. The weight oscillation issue is quite serious, which will inevitably lead to the inaccuracy of the prediction model. We also test the performance of two other adaptive weight optimization algorithms, namely, Momentum [13] and Adam [14] during online learning. By analyzing the results, we find that the weight is more stable with Momentum and Adam. However, the prediction error is larger as the enhancement of the optimization algorithms compared to the SGD lies in the improvement of the convergence speed and the ability to escape from the saddle point.

How to solve the weight fluctuation caused by traffic burstiness in the online learning process has become a new challenge. In this article, we propose a novel neural network weight optimization algorithm named SWGD. The idea of the SWGD algorithm is to establish a sliding window in the backpropagation process for each weight. We record the error partial derivative of each backpropagation in the window and the window will move to the latest partial derivative. We recalculate the update step only when the partial derivative value in the window is in the same direction. This method can effectively solve the problem of the weight oscillation caused by the network traffic burstiness, and improve the accuracy of the model during the online learning/prediction process.

Model and Solution

The proposed LNTP model is depicted in Fig. 2. Figure 2a shows the spatial structure of the model, which is an end-to-end frame structure and mainly contains two parts, namely wavelet transform and improved LSTM. Figure 2b shows the logical division of the model in temporal. We divide the model into three phases, that is, data acquisition, initial training, and online learning and prediction.

The LNTP Model

The randomness and burstiness of network traffic usually generate a lot of noise. The noise will mislead the training of the model, which might cause a certain degree of interference to the final prediction result. To address this problem, we first analyze the characteristics of network traffic data and then adopt the *Symlets* wavelet function to decompose the original traffic data and get a group of approximate data and multiple sets of detailed data, which are collectively called *Component* in this article. As Fig. 2a shows, the components at time t are the input of LSTM. After multiple iterations of training, LSTM derives the intrinsic relationship between the components. The output of LSTM is the prediction value $\hat{y}_{t+1}$ of next time $t + 1$. When the traffic statistics under the real data y_{t+1} arrive, we will use them as the labeled data of this training to calculate the error, and propagate it back. Therefore, the model is updated through the online learning of new data. Then, y_{t+1} is set as an input x_{t+1} to calculate the prediction value $\hat{y}_{t+2}$ of the next moment.

The neural network model used in the LNTP model consists of three layers: the input layer, hidden layer, and output layer. In [15], extensive experiments are given to compare the performance of the neural network models with different numbers of hidden layers under different RNN structures when solving the network traffic prediction problem. Since the input of the neural network is four sets of components obtained by wavelet transform decomposing the original network traffic data and the output is the network traffic prediction value of the next moment, so the neural network adopts a four-dimensional input node and a one-dimensional output node. The hidden layer adopts LSTM neurons as network nodes. The four input layer nodes respectively receive a set of approximate data and three sets of detailed data, and the output values of the output layer nodes represent the traffic prediction values at the next moment. Based on the method, we construct labeled training data as $X = \{A_{t-1}, D1_{t-1}, D2_{t-1}, D3_{t-1}\}$, and $Label = \{R_t\}$. The loss function during LSTM training uses the mean square error (MSE) to describe.

Online Learning

The learning/prediction process of the LNTP model is divided into three phases: data acquisition, initial learning, and online learning and prediction. As shown in Fig. 2b, in the data acquisition phase, the necessary initial training data is first prepared for the initial learning phase. From Fig. 1a, we can find that the network traffic has a significant periodicity and the period is one week. Thus, data for more than one week is needed in the learning phase. Then, the online learning phase is entered whose process is described as follows. When the real traffic statistics of the next moment y_{t+1} arrives, LNTP sets y_{t+1} as the training label of the current time, calculates the error, and corrects it inversely. Then, LNTP sets y_{t+1} as x_{t+1} for the calculation of predicted value $\hat{y}_{t+2}$ in the next time. The online learning of new data enables real-time updating of the model, which can maintain high prediction accuracy. The most important part of the online learning/prediction phase is the update of weights. We demonstrate the update calculation process of the weight W_0 in the online learning process in the LNTP model and the update process of other weights is similar. The updated amount of a single weight is the partial derivative of the error of the currently predicted result relative to the current weight multiplied by the learning rate μ. It is specifically divided into two steps:

- The partial derivative value of the relative result error is obtained according to the chain-derivation rule.
- The update step size is obtained by the optimization algorithm for the partial derivative value.

In the online learning phase, we only make one iteration for each new sample, to avoid the model overfitting the new data.

According to the backpropagation rule of back propagation through time (BPTT) algorithm and the reverse derivation formula of LSTM, we can calculate the computational complexity of updating the weight of neural network every time in the online learning phase of the model: $O(N_h^2 + N_h(N_i + N_o))$, where N_h, N_i, and N_o represent the node numbers of the hidden layer, the input layer, and the output layer, respectively. The computational complexity of a single update model in the online learning phase is within the acceptable range, which can guarantee that the model updates online in real-time to adapt to the new traffic pattern so that the model always maintains high prediction accuracy under long-term operation.

The SWGD Algorithm

To suppress the weight fluctuation of the model due to traffic burstiness in the online learning phase, we propose a novel weight optimization algorithm, called SWGD. The supervised neural network training process can be summarized into two steps: result forward calculation and error backpropagation. Each weight in the neural network corrects itself according to the update step calculated by the partial derivative of its relative result error to make the output error smaller in the next round. The core idea of SWGD is to create a sliding window. When the weight gains a new partial derivative during backpropagation, the window will move one step with time. Only when the direction of all partial derivatives in the window is the same, the partial derivative is averaged and then multiplied by the learning rate, and the obtained result is set as the step size of this update. When the partial derivative in the window is updated, the current window is destroyed and no longer moves. Then, a new window will be re-established at the next moment to avoid the impact of the old partial derivatives on the direction of future corrections. There are two parameters in the SWGD algorithm which are N and μ. N indicates the size of the window, that is, the number of partial derivatives that are simultaneously viewed, and μ is the learning rate. Figure 3a shows a running scenario of SWGD and Fig. 3b shows the change of the neural network weight mentioned in Fig. 1d when using SWGD as the weight optimization algorithm. From Fig. 3, it is obvious that the SWGD algorithm can avoid weight fluctuation effectively. It is to note that the updates of all weights are not synchronous, and thus, the overall trend of the weight in Fig. 1c and Fig. 3b is inconsistent.

The extra loop is not generated in the SWGD algorithm, so the complexity of a single training will not increase in magnitude. We can consid-

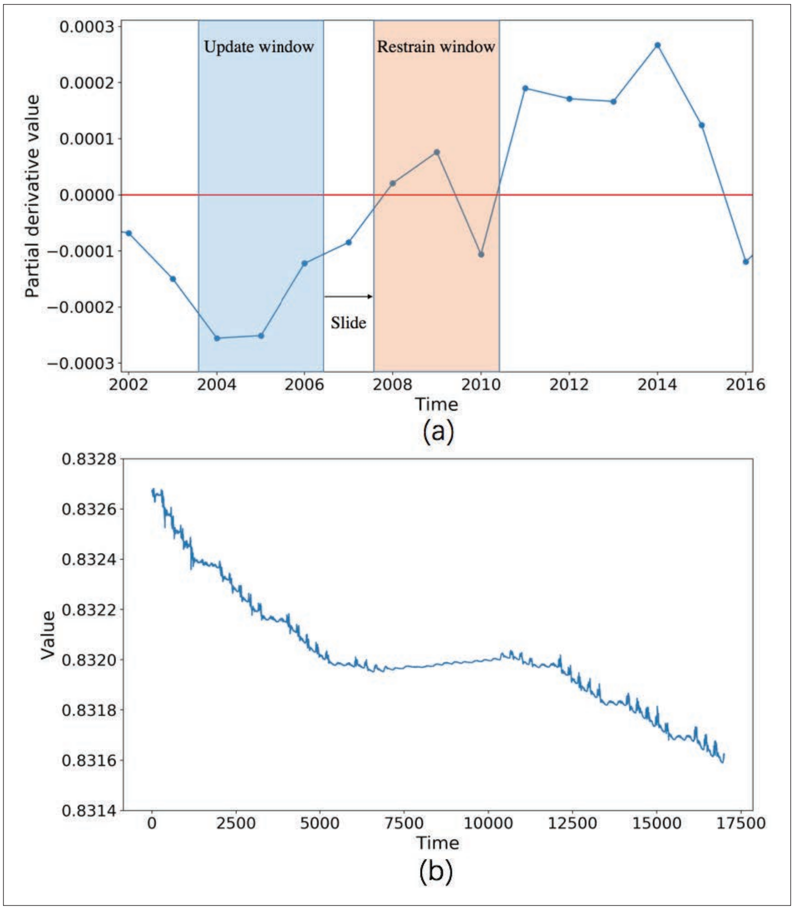

FIGURE 3. SWGD algorithm: a) Running scenario of the algorithm; b) the change of weight value of a neuron in the online learning process of LSTM model using SWGD as the optimization algorithm.

er the SWGD algorithm as a filter, which masks the updates that may generate negative incentives for weights. When the direction of the partial derivative values in the window is not the same, the update of this time will be given up. Suppose the number of data nodes is W. If the SGD algorithm is used as the weight optimization algorithm of the neural network model, all weights will be updated once when a new data node arrives during the online learning stage. That is, the update times of a single weight is W with SGD. However, when SWGD is adopted as the weight optimization algorithm, the weights will be updated only when the partial derivative value in the sliding window with the length of N is in the same direction. If there is a weight updated, the sliding window needs to be recreated. Otherwise, the sliding window will be moved forward by one step. Therefore, for a single weight, the upper bound of cumulative times of update will be $W\%N$ with SWGD.

Above all, the wavelet transform of the proposed LNTP model is a data preprocessing module, which decomposes the original network traffic data into multiple components. Then all the components are input into the LSTM neural network of the proposed LNTP model to capture the characteristics of each component and relationship between the components. The minimal data volume requirement in the data acquisition phase is determined according to the periodicity of the network traffic. This is to meet the first challenge analyzed above. Once the requirement is satisfied, LNTP moves on to the initialization learning phase. In this stage, historical data is iteratively learned many times to initialize the model to tackle the second challenge. Then, in the online learning process, we propose a weight optimization algorithm SWGD that is more suitable for network traffic prediction. The SWGD can effectively suppress the weight fluctuation caused by network traffic burstiness corresponding to the third challenge, which makes the model always maintaining good accuracy in the online learning/prediction phase.

Evaluation

Datasets

The real network traffic of private ISPs provided by the United Kingdom academic network backbone and European cities network backbone is used as our experimental datasets, called *uk_data_set* and *ec_data_set*, respectively. Both datasets are sampled at intervals of five minutes. *uk_data_set* has 19888 sets of sampled data, and *ec_data_set* has a total of 14772 sets of sampled data. The overall traffic trend of the two datasets is shown in Fig. 4, where Fig. 4a is *uk_data_set* and Fig. 4b is *ec_data_set*.

Evaluation Metrics

The experiments use root mean square error (RMSE) and mean absolute percent error (MAPE) as evaluation metrics, the lower values of the two metrics, the smaller error between the predicted value and the true value, and the higher accuracy of the model. The value of the evaluation metric is affected to some extent by the dataset. Therefore, the evaluation metrics of different datasets should not be directly compared, and can only be used as a relative reference for the results of different models of each dataset.

Experiment Setting

The datasets are divided into multiple intervals in time order and all the predicted data are also divided into multiple intervals by intervals of one week in the experiment. The prediction result of each interval is evaluated separately, to evaluate the performance of the LNTP model from multiple dimensions.

First, a series of experiments were done and the predicted error is calculated with the different number of neurons in the hidden layer. The experimental results show that when the number of neurons is greater than or equal to 6, the error tends to be stable. When the number of neurons is equal to 10, the lowest error is obtained, and increasing the number of neurons directly increases the amount of computation of the neural network. Therefore, the number of neurons in the hidden layer is set to 10.

In the initialization phase, Adam, an adaptive learning rate optimization algorithm, is used to update neural network weights for its characteristics to fast convergence and effectively escaping the saddle point [15]. The recommended values of the parameters in Adam are utilized. We compare four different models in the experiments, in which instance 1 is the DBNG model, instance 2 is the LNTP model without using online learning algorithms, instance 3 is the LNTP model with the SGD as the online learning weight optimization algorithm, and instance 4 is an LNTP model with the SWGD algorithm as the online learning weight optimization algorithm. Based on the extensive experiments mentioned above, when the value of μ is fixed at 0.02, we can find that the system performs better with a lower error rate when N ranges from 3 to 5. However, when N becomes larger than 5, the error rate of both RMSE and MAPE shows an obvious increasing tendency. When N is 3 and μ is in the range of 0.01 and 0.06, the system performs best with the lowest error. When μ is too small (i.e., ≤ 0.005) or too large (i.e., ≥ 0.06), the error of RMSE and MAPE becomes larger. Without loss of generality, the parameters used in the SWGD algorithm are finally determined to be $N = 3$ and $\mu = 0.02$.

Experimental Results

Based on the *uk_data_set* and *ec_data_set*, we conduct experiments to compare the performance of the four instances mentioned above. The size of the training set and test set are 2888 and 17000, respectively, in the *uk_data_set*. In the *ec_data_set*, the values are 2772 and 12000, respectively. During the simulation process, the training dataset is visible in the initialization phase. However, in the online learning/prediction process, the real data of the next moment in the test dataset can be visible only when the model completes the prediction of the next moment based on the historical data. Although instances 1 and 2 do not use online learning, how the data is provided is consistent with instance 3 and 4. The difference lies in that the model does not update its weights based on new data. It is to note that the validity of the wavelet transform is verified before evaluating the performance of different models, the experimental results show that the model uses the wavelet transform to decompose the original data can reduce the prediction error by 15 percent compared to the ones without using it when other parameters are identical.

Prediction results are divided by a time interval of one week, and each interval contains 2016 data nodes. Two metrics of the prediction error, namely, RMSE and MAPE, are calculated for each interval. Figure 5 shows in detail the changes of RMSE and MAPE for four instances on two datasets over multiple time intervals. It can be seen that on two different datasets, the proposed LNTP model has the lowest prediction error in each time interval. It is obvious that in the fourth and fifth week of the *uk_data_set* prediction interval in Figs. 5a and 5b, the prediction errors of instance 1 and instance 2 are significantly higher than instance 3 and instance 4, because the traffic change pattern has changed greatly in this interval. From Fig. 4a, instance 3 and instance 4 benefit from online learning so that the prediction results always maintain good accuracy and can adjust quickly. Instance 4 can effectively suppress the random burst of network traffic which results in weight fluctuation during neural network training as it uses the proposed SWGD weight optimization algorithm. As a result, the stability of the model is further improved and the prediction error is further reduced. At the same time, it is worth noting that for *uk_data_set*, the MAPE metric of instance 4 shows a steady downward trend. This is because the traffic data for neural network learning continues to accumulate, and the model continuously optimizes itself so that the error continues to decrease.

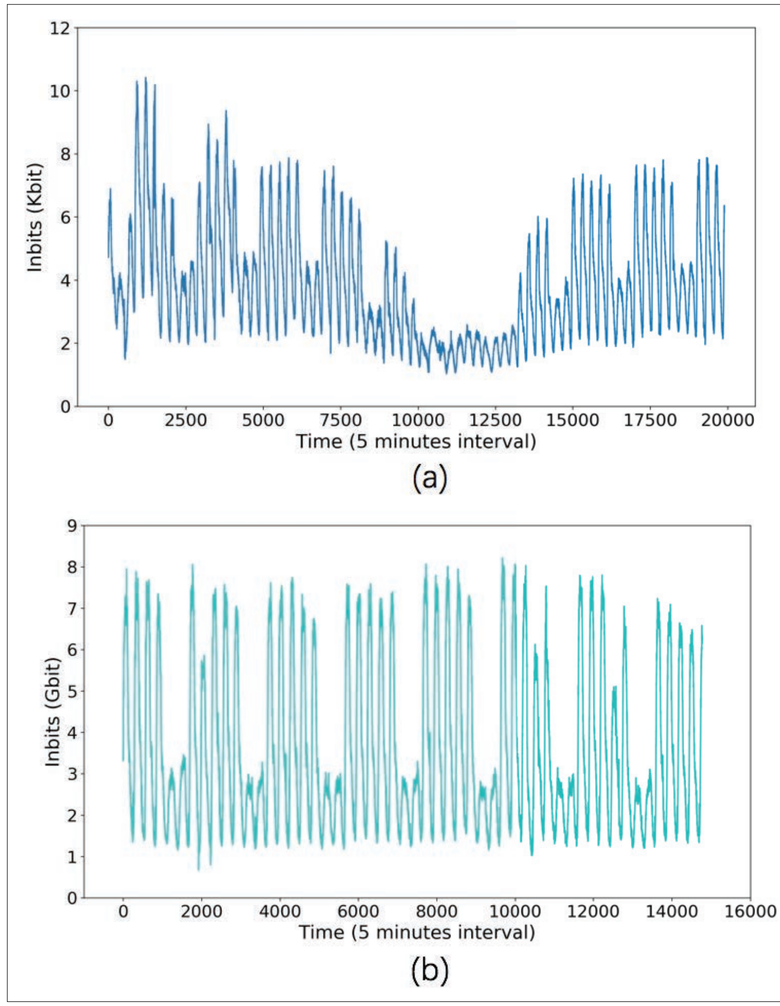

FIGURE 4. Datasets: a) the traffic trend of uk_data_set; b) the traffic trend of ec_data_set.

The overall prediction results for all instances are also compared based on the two datasets. Figure 6 shows the overall prediction error of four instances, where Fig. 6a uses the RMSE metric and Fig. 6b uses the MAPE metric. As can be seen from Fig. 6, the errors of instance 3 and instance 4 using the *uk_data_set* are much smaller than those of instance 1 and instance 2, but this phenomenon is not obvious on the *ec_data_set*. This is because the traffic of the *uk_data_set* is greatly reduced in the middle for half a month, which can be seen by observing Fig. 4a. Instance 3 and instance 4 can perform online learning and will update its weight in real-time to adapt to the new data model. For the *ec_data_set*, network traffic data does not change over time, and it seems to follow a similar traffic pattern.

Above all, the experimental results reveal that the overall prediction error of the proposed LNTP model without the proposed SWGD algorithm has an average decrease of 19.59 percent and 19.35 percent compared with the DBNG model with RMSE and MAPE, respectively, while the prediction error of the model with the proposed SWGD algorithm is reduced by 31.76 percent

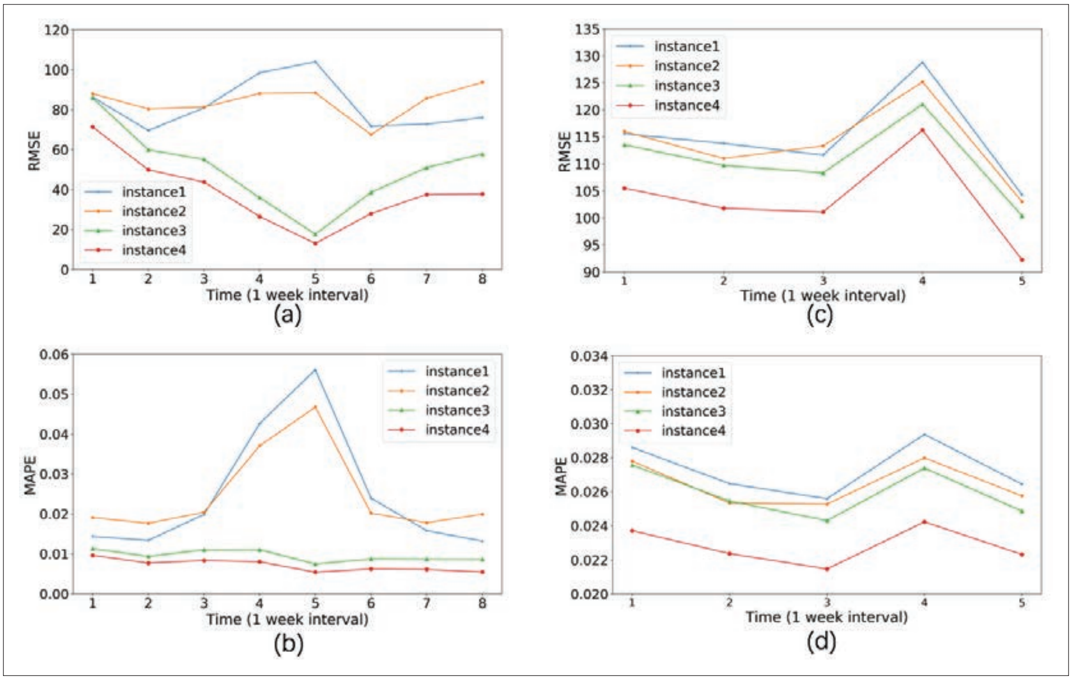

FIGURE 5. Comparison of various metrics in time interval: a) RMSE metrics on uk_data_set; b) MAPE metrics on uk_data_set; c) RMSE metrics on ec_data_set; d) MAPE metrics on ec_data_set.

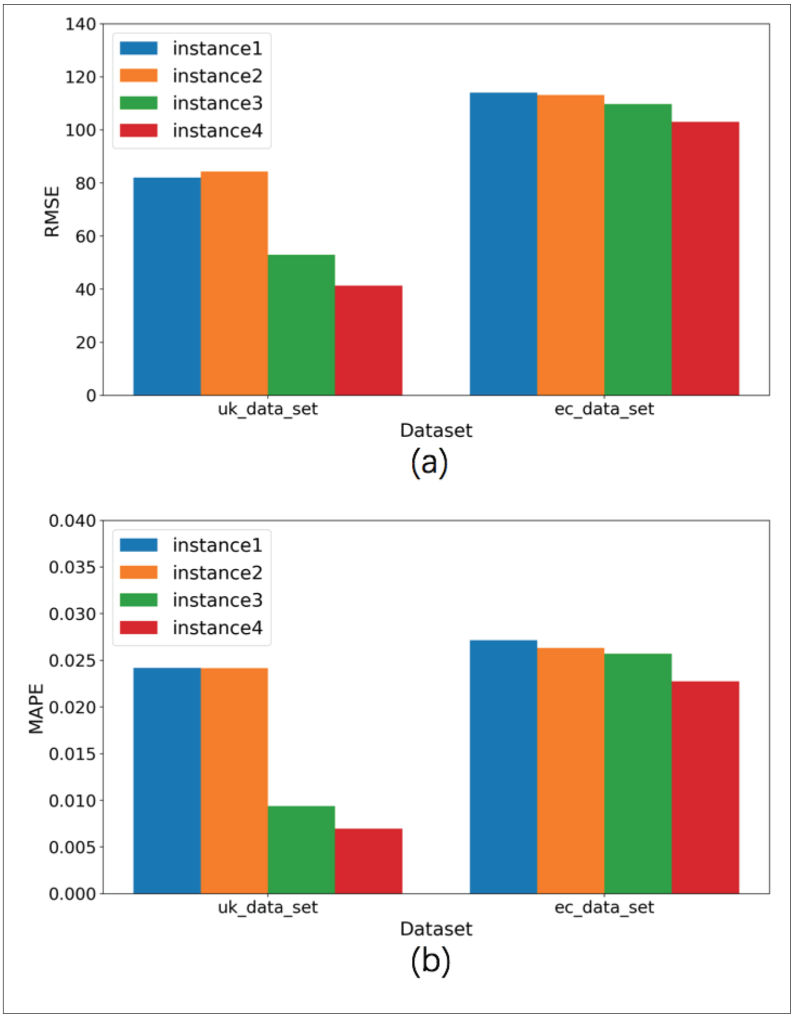

FIGURE 6. Complete comparison of various metrics: a) RMSE metric; b) MAPE metric.

and 29.67 percent. At the same time, on the *uk_data_set*, the overall prediction error of instance 4 compared with instance 3 has a decrease of 22.1 percent and 25.64 percent with RMSE and MAPE, respectively, while the reduction of the two metrics with the *ec_data_set* data set is 6.15 percent and 11.67 percent. This verifies the advantages of the LNTP model and the SWGD optimization algorithm in dealing with network traffic prediction problems.

Conclusions

To improve the efficiency of network management, we have researched network traffic prediction and proposed three innovations for the three major challenges currently faced by current network traffic prediction.

- An end-to-end online traffic prediction model named LNTP is proposed. The model first uses the wavelet transform to decompose the original traffic data, and uses the decomposed component as the input of the LSTM neural network to capture the relationship between the components of the original traffic data to cope with the difficulty of capturing network traffic data features.
- This model is divided into three stages, namely, data collection, initial learning, online learning, and prediction, so that the model can adjust itself in low complexity and high efficiency during the operation of the network to adapt the current network traffic mode all the time and maintain good predictive accuracy.
- A weight optimization algorithm named SWGD is proposed to suppress the gradient turbulence caused by network traffic burstiness in the weight update of the model in the online learning process, and thus to further improve the stability and accuracy of online learning of the model.

Finally, two different real datasets are used to validate the existing DBNG model and several versions of the LNTP model proposed in the article under different conditions. The test dataset is innovatively divided into several intervals, and the experimental results of each interval are evaluated, which shows that the proposed LNTP model and the proposed SWGD optimization algorithm can reduce the prediction error by 29 percent in the network traffic prediction problem. For the next step, we will consider paying attention to the traffic characteristics of multiple network nodes simultaneously, and introduce a traffic matrix to analyze the features, looking forward to further improvement of the prediction accuracy.

Acknowledgments

This work was supported in part by the Natural Science Foundation of China (Nos. 61572191, 61602171 and 62072175) and the Natural Science Foundation of Hunan Province, China (No. 2020JJ4058).

References

[1] Internet Live Stats; http://www.internetlivestats.com, 2019, accessed 30 June 2019.
[2] H. Kalbkhani *et al.*, "Adaptive LSTAR Model for Long-Range Variable Bit Rate Video Traffic Prediction," *IEEE Trans. Multimedia*, vol. 19, no. 5, May 2017, pp. 999–1014.
[3] G. Han *et al.*, "A Dynamic Multipath Scheme for Protecting Source-Location Privacy Using Multiple Sinks in WSNs Intended for IIoT," *IEEE Trans. Ind. Inform.*, vol. 16. no. 8, Aug. 2020, pp. 5527–38.
[4] Y. Shu *et al.*, "Wireless Traffic Modeling and Prediction Using Seasonal ARIMA Models," *IEICE Trans. Commun.*, vol. E88B, no. 10, Oct. 2005, pp. 3992–99.
[5] Y. Zhang *et al.*, "Component GARCH Models to Account for Seasonal Patterns and Uncertainties in Travel-Time Prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, Apr. 2015, pp. 719–29.
[6] J. Feng *et al.*, "DeepTP: An End-to-End Neural Network for Mobile Cellular Traffic Prediction," *IEEE Netw.*, vol. 32, no. 6, Nov. 2018, pp. 108–15.
[7] H. Liu *et al.*, "Multivariate Multi-Order Markov Multi-Modal Prediction with Its Applications in Network Traffic Management," *IEEE Trans. Netw. Serv. Manag.*, vol. 16, no. 3, Sept. 2019, pp. 828–41.
[8] L. Dai *et al.*, "EMD-Based Multi-Model Prediction for Network Traffic in Software-Defined Networks," *Proc. 11th IEEE Int'l. Conf. Mobile Ad Hoc and Sensor Systems*, Oct. 2014, pp. 539–44.
[9] L. Nie *et al.*, "Network Traffic Prediction Based on Deep Belief Network and Spatiotemporal Compressive Sensing in Wireless Mesh Backbone Networks," *Wirel. Commun. Mob. Com.*, vol. 2018, Jan. 2018, Article 1260860.
[10] W. Huang *et al.*, "Backbone Network Traffic Prediction Based on Modified EEMD and Quantum Neural Network" *Wirel. Pers. Commun.*, vol. 99, no. 4, Apr. 2018, pp. 1569–88.
[11] W. Chen *et al.*, "A Novel Hybrid Network Traffic Prediction Approach Based on Support Vector Machines," *J. Computer Networks and Commun.*, vol. 2019, Feb. 2019, article 2182803.
[12] S. Bahraini *et al.*, "An Online Learning Algorithm for Demand Response in Smart Grid," *IEEE Trans. Smart Grid*, vol. 9, no. 5, Sept. 2018, pp. 4712–25.
[13] N. Qian, "On the Momentum Term in Gradient Descent Learning Algorithms," *Neural Netw.*, vol. 12, no. 1, Jan. 1999, pp. 145–51.
[14] D. Kingma *et al.*, "Adam: A Method for Stochastic Optimization," *Proc. 3rd Int'l. Conf. Learning Representations*, 2015, pp. 1–15.
[15] S. Ruder, "An Overview of Gradient Descent Optimization Algorithms," https://arxiv.org/abs/1609.04747, 2019, accessed 30 June 2019.

Biographies

Lianming Zhang received his Ph.D. degree from the School of Information Science and Engineering of Central South University, China, and his M.S. and B.S. degrees from the Department of Physics of Hunan Normal University, China. He is currently a professor in the College of Information Science and Engineering, Hunan Normal University, China. His research interests include computer networks, software-defined networking and edge computing.

Huan Zhang received his B.S. degree in computer science from the Hunan Vocational College of Commerce, China. He is currently pursuing the master's degree with Hunan Normal University, China. His research interests include network traffic prediction and software-defined networking.

Qian Tang received her B.S. degree in education technology from the Hunan First Normal University, China. She is currently pursuing the master's degree with Hunan Normal University, China. Her research interests include data center network and software-defined networking.

Pingping Dong received her B.S., M.S. and Ph.D. degrees from the School of Information Science and Engineering at Central South University, China. Currently she is an associate professor in the College of Information Science and Engineering, Hunan Normal University, China. Her research interests include protocol optimization and protocol design in wide area networks and wireless local area networks.

Zhen Zhao received his B.S. degree in software engineering from the Tiangong University, China. He is currently pursuing the master's degree at Hunan Normal University, China. His research interests include edge computing.

Yehua Wei received his Ph.D. degree in computer science and engineering from Hunan University, China. He is currently an associate professor with the College of Information Science and Engineering, Hunan Normal University, China. His research interests include embedded systems, cyber-physical systems, and the Internet of Things.

Jing Mei received her Ph.D. in computer science from Hunan University, China. She is currently an assistant professor in the College of Information Science and Engineering at Hunan Normal University, China. Her research interests include parallel and distributed computing, cloud computing, etc.

Kaiping Xue (M'09-SM'15) received the B.S. degree from the Department of Information Security, University of Science and Technology of China (USTC), and the Ph.D. degree from the Department of Electronic Engineering and Information Science (EEIS), USTC. From 2012 to 2013, he was a postdoctoral researcher with the Department of Electrical and Computer Engineering, University of Florida. He is currently an associate professor with the Department of Information Security and the Department of EEIS, USTC. His research interests include next-generation Internet, distributed networks, and network security.