Satellite Traffic Forecast Based on Multi-dimensional Periodic Features

Weidong Zhou[1], Yana Qian[2], Kanglian Zhao[1], Wenfeng Li[1](✉), and Fa Chen[1]

[1] Nanjing University, Nanjing, China
zhaokanglian@nju.edu.cn, leewf_cn@hotmail.com
[2] Shanghai Astronautics Electronic Co., Ltd., Shanghai, China

Abstract. With the development of satellite networks and the increase in business requirements, it is a challenge to better provide high-quality service to users. The accurate prediction of end-to-end traffic can contribute to the realization of congestion control, resource allocation and anomaly detection. Accurate flow forecasts need to consider the short-term and long-term features of the flow. Our paper proposed a model based on deep learning named Multi-Dimensional Temporal Feature Neural Network (MTFNN) to capture both short-term dependencies and long-term dependencies for traffic prediction. MTFNN mainly contains two components: 1) Short-Term Temporal Dependencies which based on Long-Short Term Memory network (LSTM), used to predict the basic trend of traffic. 2) Long-Term Temporal Dynamic Similarity which based on LSTM and Attention mechanisms, used to improve the model's sensitivity to fluctuations and peak prediction. Experiments performed on real-world public traffic datasets show our proposed model has a smaller prediction error and more accurate peak prediction.

Keywords: Traffic flow prediction · LSTM network · Attention mechanism

1 Introduction

The satellite information network is characterized by large coverage and long communication distance, and is an important part of the integration of the sky-ground network. With the construction of satellite information networks, satellite traffic services continue to increase. However, the bandwidth and other resources of satellite nodes are relatively limited. Therefore, further management and allocation methods are needed to avoid the degradation of service quality due to increased services. Traffic distribution can intuitively reflect the current network status, and accurately predict the future network traffic distribution can be used for congestion control, resource allocation, and intelligent routing.

High non-linearity, multi-scalar, self-similarity and long-term dependence are the main characteristics of satellite networks [1]. These statistical characteristics determine the predictability of traffic. Previous work had made traffic predictions by analyzing these characteristics of traffic, and had achieved good research results.

Q. Guo et al. (Eds.): WiSATS 2021, LNICST 410, pp. 267–277, 2022.
https://doi.org/10.1007/978-3-030-93398-2_27

Previous research work has proposed many forecasting models, which can be summarized into two categories: linear algorithms and nonlinear algorithms. The Auto Regressing Moving Average (ARMA) [2], the Auto Regressing Integrated Moving Average (ARIMA) [3, 4] and the Holt-Winter algorithm [5] are most general algorithms applied to network traffic prediction problems. However, these algorithms are based on mathematical theoretical assumptions (such as linear, moving average), and in large-scale networks, these methods have insufficient ability to deal with nonlinear traffic characteristics, which will lead to a decrease in prediction accuracy. Nonlinear prediction methods commonly involve neural network (NN) [6]. Dingde Jiang [7] analyzed the Recurrent Multilayer Perceptron (RMP) to forecast traffic matrices in large scale network. Bermolen and Rossi [8, 9] found that the SVM model has good performance in link load prediction problems. Recurrent Neural Network (RNN) can store the state of each time interval in the network, thus having the characteristics of time memory. RNN is widely used in time series prediction and classification tasks, and has excellent performance in sequence modeling tasks Ramakrishnan [10] compared the RNN and its variants (such as LSTM, Gated Recurrent Unit Network (GRU)) for traffic prediction to linear forecasting models.

However, the above approaches only consider the short-term time dependence of the flow sequence, but ignore the long-term time dependence. Considering short-term time characteristics to predict future flow can achieve good prediction accuracy, but there are deficiencies in the prediction of severe fluctuation sequences and peak flow. Traffic data has strong daily, weekly or even monthly periods, which can be used for future traffic forecasting. Satellites will serve different regions with the orbital period. Switching between regions will produce fluctuations in traffic, and the day-to-night changes in the region itself will also produce fluctuations in traffic. These features have long-term characteristics, and only considering short-term time characteristics will obviously ignore these characteristics which can help the forecasting model to improve the sensitivity to traffic fluctuations. LSTM is a powerful RNN architecture that can be used for time series forecasting. But in traffic prediction, LSTM can only consider the flow sequence of the past several minutes or a few hours (which is determined by time interval), so some long-term temporal features will not be fully used.

In this case, we propose a novel deep-learning mode named Multi-Dimensional Temporal Feature Neural Network (MTFNN), which mainly contains two components: 1) Short-Term Temporal Dependencies employs the Long-Short Term Memory network (LSTM) to capture the short-term dependence, 2) Long-Term Temporal Dynamic Similarity uses attention-based recurrent structures to model the long-term temporal dependencies of each flow, and then predicts flow volume according short-term dependencies and long-term dependencies.

The rest of the paper is organized as follows. Section 2 presents the MTFNN model to make predictions and the dynamic periodicity of traffic. Experiment and results are presented in Sect. 3. The paper is concluded by Sect. 4.

2 Prediction Model

This section describes the details for our proposed model, Multi-Dimensional Temporal Feature Neural Network (MTFNN). Figure 1 shows the architecture of our model, which can be divided into three parts: short-term time dependence, long-term time dependence and final prediction. The short-term time dependence part is based on the historical flow sequence near the forecast time to make the prediction and is used to fit the basic trend of the flow sequence, while the long-term time-dependence part is composed of a traffic input sequence strongly correlated with the forecast time in a longer time interval to capture the periodic characteristics and peak flow and is used to strengthen the model's accuracy of volatility and peak forecasting, and the final forecasting part is mainly to perform the final mapping of the predicted values of the first two parts to obtain The final predicted value can dynamically adjust the impact weight of the short-term and long-term predicted results on the final predicted value, which makes the model more robust and can adapt to different traffic sequences.

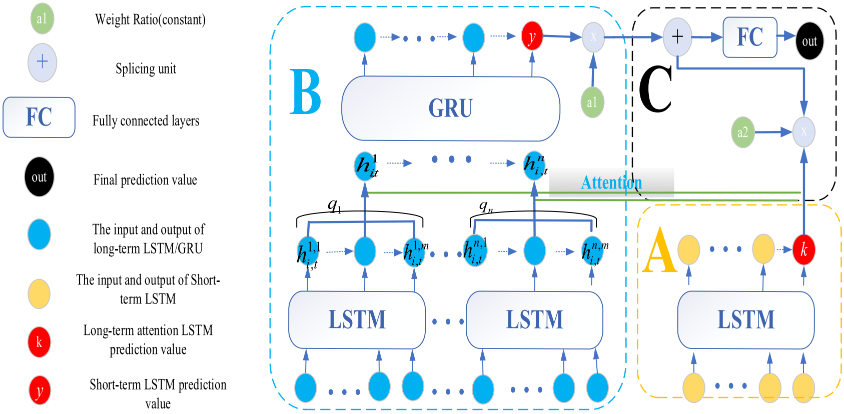


Fig. 1. The Architecture of MTFNN. A: Short-term temporal dependency. B: Long-term temporal network. C: Final prediction.

2.1 Short-Term Temporal Dependency

The short-term dependency part means that we obtain short-term characteristics of traffic and make traffic forecasts based on short-term historical data. Long short-term memory (LSTM) is a special model used to solve the problem of gradient disappearance and explosion during RNN training. The three forget gate structures can fully deal with the time characteristics in the sequence, and show excellent performance in natural language processing and time sequence prediction tasks.

LSTM adds three memory gate structures on the basis of RNN (Fig. 2). The input of the forget gate is the previous unit output h_{t-1} and the current time input data x_t. The output f_t of the forget gate determines what kind of information can pass through the unit. The output i_t of the input gate is a control signal, which determines which new input information can be input to the unit. The output gate determines what information is output [11].

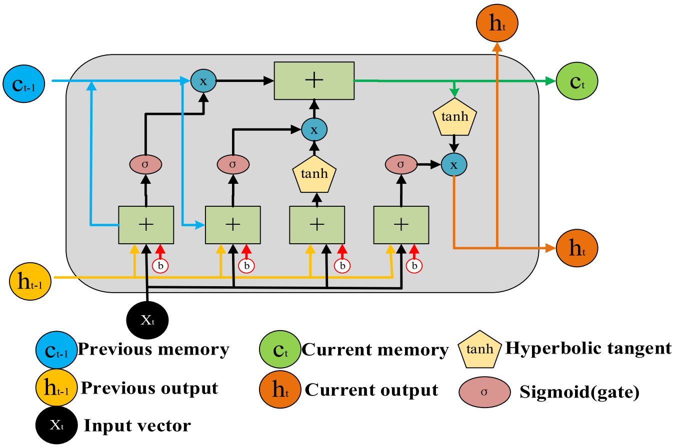


Fig. 2. LSTM architecture.

The specific working process of the LSTM unit is as follows: At time t, the LSTM input x_t includes the input at the current time, the last hidden state h_{t-1} and the last unit state c_{t-1} at the previous time, that is, retain the useful information in the traffic sequence at the last moment. x_t and h_t get the current input unit state c_t through activation function *Tanh*. The current unit state c_t is the sum of the last unit state c_{t-1} passing through the forget gate f_t and the current input unit state c_t passing through the input gate i_t. Finally, LSTM output h is calculated by the current unit state c_t through the activation function and output gate o_t. The calculation formula for each variable is as Eq. (1) to Eq. (5).

$$i_t = \sigma_g(W_{xi} * x_t + W_{hi} * h_{t-1} + W_{ci} * c_{t-1} + b_i) \tag{1}$$

$$f_t = \sigma_g\left(W_{xf} * x_t + W_{hf} * h_{t-1} + W_{ci} * c_{t-1} + b_f\right) \tag{2}$$

$$C_t = f_t \otimes c_{t-1} + i_t \otimes \sigma_c(W_{xc} * x_t + W_{hc} * h_{t-1} + c) \tag{3}$$

$$o_t = \sigma_g(W_{xo} * x_t + W_{ho} * h_{t-1} + W_{co} * c_t + b_0) \tag{4}$$

$$h_t = o_t \otimes \sigma_g(c_t) \tag{5}$$

where: W_{xf}, W_{xi}, W_{xo}, W_{xc} are the training matrices related to the input x_t while W_{hc}, W_{hi}, W_{ho}, W_{hf} are the training matrices related to the last hidden state h_{t-1}, W_{ci}, W_{cf}, W_{co} are diagonal matrices connected to the current unit state a c_t and gate function, b_i, b_f, b_c, b_o are offset factors. σ_g is the activation function and $\otimes$ denotes the Hadamard product.

2.2 Long-Term Temporal Dependency

The long-term dependency part means that we obtain long-term characteristics of traffic and make traffic forecasts based on long-term historical data. Although compared with the traditional time series prediction model, LSTM has a stronger ability to process long-term dependence, but as the time sequence rises, LSTM will face the problem of disappearing gradient and inconspicuous periodicity. In order to avoid the above problems, we need construct a new time series that reflects long-term characteristics, rather than just increase the length of the time series. Usually we use the relative time interval of a clear prediction moment, for example (the same moment yesterday, the same moment last week). A reasonable and clear relative time interval can reflect more time characteristic information at the current moment. Through the analysis of the real data set, we found that the traffic data is not strictly periodic, and there is a certain offset. A week's summary traffic situation of a certain node is selected from the Abilene dataset, as shown in Fig. 3, the interval is one day and the peak traffic appeared between 20:40–23:00 (Although the peak traffic on Thursday appeared at 16:40, the traffic volume at 21:40 is still a local maximum). In satellite network, the relative period interval is set according to the satellite orbit period, which can predict the traffic situation of a certain area in the past period of time, and reduce the abnormal traffic fluctuation caused by the change of satellite coverage area.

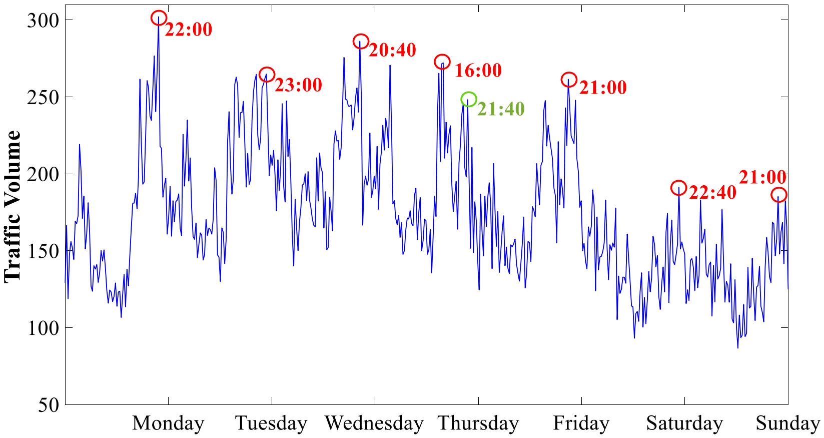


Fig. 3. Long-term periodicity and shifting of traffic

In order to better capture the long-term time characteristics of the time series, we need to set the relative time interval reasonably and consider the deviation of the flow. As shown in Fig. 1, we set the relative time interval to one day, where N represents the past the number of days, and M represents the number of traffic nodes we select in each day, for example, when the next moment we want to predict is 21:00 and M = 5, we select the past 20:50–21:10 a total of 5 moments of traffic matrix. Then we use LSTM to extract the temporal features of the past daily sequence.

Inspired by the Attention paper [12], in the last step, we adopted the idea of attention mechanism to solve the problem of traffic offset, and obtained the weighted result by calculating the correlation between the daily traffic in the past and the traffic at the current forecast moment. First, after LSTM, output the short-term forecasts of each day in the past, and get the weighted summation result $h_{i,t}^{n}$ of m time interval every day according to the attention. Specifically defined as:

$$h_{i,t}^{n} = \sum_{q \in Q} \alpha_t h_{i,t}^{n,m} \tag{6}$$

where weight α_t measures the similarity of the time interval m in day $n \in N$. Formally, the weight α_t is defined as:

$$\alpha_t = algin(q, k) = \frac{\exp(score(q, k))}{\sum_{q \in Q} \exp(score(q, k))} \tag{7}$$

where k means the output of short-term LSTM, q means the sequence of m time intervals in day $n \in N$.

In this paper, the alignment function is defined as:

$$\text{score} = \text{sigmoid}\left(v_a^T \tanh(W_a q + Uk + b)\right) \tag{8}$$

where ${v_a}^T$, W_a, U, b are learned parameters. ${v_a}^T$ denotes the transpose of v_a. For each pervious day p, we get the weighted sum presentation $h_{i,t}^{p}$. Then we use a GRU to preserve the sequential information by using these periodic representations as input.

$$\text{y} = \text{GRU}\left(h_{i,t}^{p}, h_{i,t}^{p-1}\right) \tag{9}$$

where we regard the output of the last time interval y as the representation of long-term temporal dependency.

2.3 Final Prediction

After getting long-term attention LSTM prediction value y and short-term LSTM prediction value k, we splice y and k through a fully connected layer to achieve the final result. And the out contains spatial and temporal information. However, the long-term characteristics of different traffic sequences have different strengths. Simply splicing y and k give the same weight to y and k (the weight of the influence on the forecast result) and could not get the best prediction result. For different time series, the weight distribution of y and k need to be adjusted flexibly to achieve a better prediction effect.

3 Experiment Settings

3.1 Dataset

Real satellite network traffic data sets are difficult to obtain, and simulated traffic data sets are obtained through simulation formulas and parameters, which lack the variability and complexity of real data. The uncertainty and complexity of real traffic will lead to a decrease in the accuracy of traffic prediction. Therefore, this paper selects ground traffic data sets to conduct preliminary research on satellite network traffic predictions.

According to the topology diagram of the Abilene network (Fig. 4), it can be seen that the nodes in the Abilene network are distributed in various major states in the United States. The distance between the nodes is relatively long and there are direct and indirect links between the nodes, which similar to the satellite topology. We use the ground node as the ground base station, and the node sending traffic is used as the traffic data of the ground base station accessing the satellite, and the traffic received by the node is used as the traffic accessing the ground base station after transmission through the satellite. Therefore, this paper uses the Abilene [13] dataset to evaluate our traffic prediction model. The Abilene network consists of 12 nodes (the two nodes of Atlanta are summarized into one node in the Fig. 4) and 15 links (the connection between the two nodes of Atlanta is also omitted), containing traffic information at 48383 moments. The time interval of the Abilene datasets is 5 min.

Fig. 4. Topology of Abilene network

We summarized the inflow and outflow of each nodes and spliced them into a matrix. We have performed matrix prediction and node aggregate traffic prediction respectively. Our experiment uses about 30,000 (100 days) of traffic data, of which 80% is used for model training, and the remaining 20% is used to test model performance. Finally, we normalize the data by max-min normalization method.

3.2 Evaluation Metric and Baselines

In order to intuitively reflect the prediction performance of our proposed model, we use the mean square error (MSE) as the evaluation factor

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{y}_i - y_i\right)^2 \tag{10}$$

where y_i is the actual value, $\overset{\wedge}{\mathrm{y}}_i$ is the predicted value and n represents the number of predicted sequences.

We compare our model with some widely used time series regression models, including (1) Naive Method (NM) (2) ARIMA (3) Support Vector Machine (SVM) (4) LSTM [14] (the state-of-the-art method in traffic prediction). For model (2), (3), We use the latest implementation in the scikit-learn library.

3.3 Hyper-parameter Settings

We adjust the hyper parameters based on the performance of the validation set. In the temporal part, the length of short-term LSTM sets to 15 (considering the previous 55 min of traffic), for long-term periodic information, P is set to 7 (considering the previous 7 days), and Q is set to 7 (considering 15 min before and after of relative predicted time), the number of hidden elements of LSTM is 64, the entire network adopts the Pytorch architecture and is optimized by Adam. The batch size of the data set is 64, the learning rate is 0.0001. To avoid over-fitting, we introduced the Dropout layer and early-stop mechanism. The training process runs for 500 epochs.

4 Results

We tested our model on the Abilene dataset and took a subset of 30000 traffic volume at each node and similarly split this data set into training and test datasets. For the Abilene data set, these 30000 data points corresponded to volume measurements from May 29, 2004 at 20:00 to September 10, 2004 at 23:55.

This paper compared the performance between our proposed model and other comparison methods in matrix prediction. The matrix is composed of the inflow and outflow of 12 nodes. In addition, ARIMA and SVM output one-dimensional results, we use ARIMA and SVM to forecast each aggregate traffic and the obtained MSE is averaged as the matrix predicted MSE.

As show in Fig. 5, we can see that our proposed model achieves the best performance, and the performance of the methods using the neural network are better than the traditional network prediction method, which shows that the traditional network prediction method has shortcomings in dealing with nonlinearity and complex dynamics.

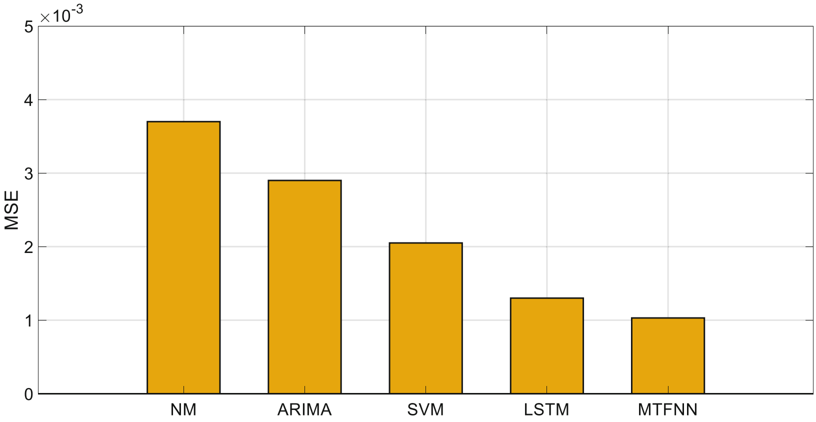


Fig. 5. Comparison of prediction errors for next matrix.

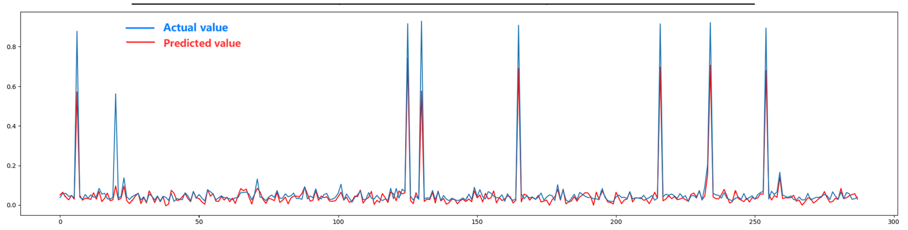
This paper also performed summary traffic prediction for each single node, and in this part, we focus on comparing the performance of LSTM and MTFNN. Compared with traffic matrix forecasting, the cost and overhead of single-node traffic forecasting are smaller. In matrix forecasting, other traffic sequences will have an impact on the current sequence, and these effects may be positive or negative.

The experiment randomly selected four nodes from 12 nodes to predict the inflow and outflow traffic. As shown in Table 1, compared with LSTM, MTFNN achieved a maximum reduction of 25.3% and a minimum of 10.3% in terms of MSE on Abilene. These results well justify the effectiveness of long-term temporal dependence on the prediction quality.

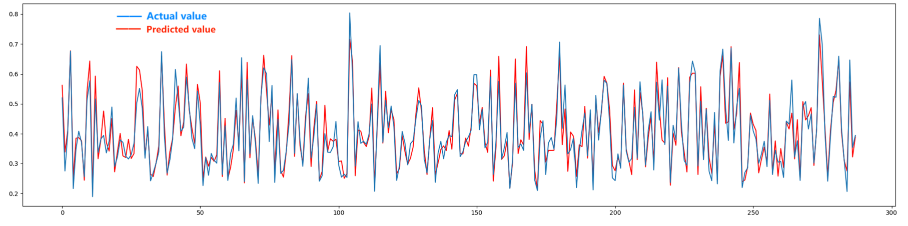
Table 1. Prediction results and comparision of end-to-end traffic

Aggregate traffic	MSE	
	LSTM	MTFNN
C-in	0.000784	0.000659
C-out	0.000244	0.000212
D-in	0.000240	0.000211
D-out	0.001217	0.001087
H-in	0.000554	0.000497
H-out	0.000838	0.000742
L-in	0.000224	0.000200
L-out	0.000677	0.000506

At the same time, the experiment randomly selected a backbone traffic sequence and the prediction situation of the edge traffic sequence. Figure 6 shows the comparison result of the specific predicted value and the true value. In part (a) (we can think of the traffic of satellites passing through low-demand areas), MTFNN can achieve low-value stable fitting and accurate peak calibration (it can be effective by setting the threshold later. Early warning peak traffic). In part (b), the predicted value can better fit the true value. Some of the predicted values are greater than the actual value. This may be due to the large peak in the long-period flow of the predicted node in the past. It is caused by a large time characteristic value, but it does not constitute an obvious misjudgment.



(a) Low demand flow result.



(b) High demand flow result.

Fig. 6. Comparison of predicted value and true value.

5 Conclusion

In this paper, we propose a novel traffic matrix prediction method named Multi-Dimensional Temporal Feature Neural Network (MTFNN). This model considers the multiple dimensions of the flow to predict the future flow. Firstly, we use short-term LSTM and long-term attention LSTM to extract the short-term and long-term temporal dependency of the traffic. The evaluation on the Abilene dataset show that the proposed model has better performance than the previous model. This paper also provides new ideas for traffic forecasting from multiple dimensions. In the space network, in terms of the communication requirements of the satellite orbit coverage area, this paper uses the ground real traffic data set to simulate the satellite traffic, and has achieved good experimental results. Compared with the LSTM model, it is 10.3%–25.3%. The promotion. However, the frequent changes in the topology of the space network and the frequent interruption of the link cannot be simulated by the ground data set. Therefore, the follow-up will use the ground real data set and the simulation data set to make further research work on the satellite traffic prediction.

References

1. Leland, W., Taqqu, M., Willinger, W., Wilson, D.: On the self-similar nature of Ethernet traffic. In: Proceedings of SIGCOMM 1993, pp. 183–193 (1993)
2. Yang, C., Jia, S., Zhang, H., Liu, X.: The prediction model for WSN traffic based on ARMA. Eur. J. Oper. Res. **219**(3), 738–750 (2018)
3. Maequez, P., Pinos, D., Juan, I.O.: Performance comparison in network traffic prediction for polynomial regerssion to P1P versus ARIMA and MWM. In: 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EUConRus). IEEE (2018)
4. Yu, Y., Wang, J., Song, M., Song, J.: Network traffic prediction and result analysis based on seasonal ARIMA and correlation coefficient. In: Proceedings of International Conference on Intelligent Systems Design and Engineering Applications (ISDEA), vol. 1, pp. 980–983, October 2010
5. Brutlag, J.D.: Aberrant behavior detection in time series for network monitoring. In: LISA, vol. 14, pp. 139–146 (2000)
6. Dharmadhikari, V.B., Gavade, J.D.: An NN approach for MPEG video traffic prediction. In: 2nd International Conference on Software Technology and Engineering, San Juan, USA, pp. V1-57–V1-61 (2010)
7. Jiang, D., Hu, G.: Large-scale ip traffic matrix estimation based on the recurrent multilayer perceptron network. In: IEEE International Conference on Communications, ICC 2008, pp. 366–370. IEEE (2008)
8. Bermolen, P., Rossi, D.: Support vector regression for link load prediction. Comput. Netw. **53**(2), 191–201 (2009)
9. Katris, C., Daskalaki, S.: Prediction of Internet traffic using time series and neural networks. In: Proceedings of International Work-Conference on Time Series Analysis (ITISE), vol. 1, pp. 594–605 (2014)
10. Ramakrishnan, N., Soni, T.: Network traffic prediction using recurrent neural networks. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, pp. 187–193 (2018). https://doi.org/10.1109/ICMLA.2018.00035
11. Gers, F.A., Schraudolph, N.N., Schmidhuber, J.: Learning precise timing with LSTM recurrent networks. J. Mach. Learn. Res. **3**, 115143 (2003)
12. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, pp. 5998–6008 (2017)
13. Zhang, Y.: Abilene dataset. http://www.cs.utexas.edu/yzhang/research/AbileneTM
14. Azzouni, A., Pujolle, G.: A long short-term memory recurrent neural network framework for network traffic matrix prediction. CoRR, abs/1705.05690 (2017)