A Multi-level Wavelet-based Time Series Approach for Network Traffic Prediction

Liu Si-yuan[1*], He Qian[1], Liu Peng[1], Zhang Fan[2]
[1]School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541004, China
{21031102009, heqian}@mail.guet.edu.cn, mofengliu9961@gmail.com
[2]School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, China
zfan2070@gmail.com

Abstract

Network traffic sequence is a time series with complex characteristics such as timeliness and nonlinearity. In order to extract potential features from network signals, this paper designs a neural network TSWNET based on wavelet packet transform to predict network traffic. This method utilizes the advantage of wavelet transform in analyzing signal features, extracting local and global features of the signal in both the time and frequency domains. These features are used to detect transients and singularities in network traffic sequences. This paper embeds wavelet transform into a neural network in a residual manner to form the basic unit FWDB(Flow Wavelet Dense Block), and finally constructs the neural network TSWNET(Traffic Series Wavelet Network). Finally, two actual network traffic datasets were used for experiments, and the experimental results showed that the proposed method has better performance in network traffic prediction compared to other similar methods.

Keywords: wavelet packet transform, network traffic, prediction, signal features

1 Introduction

With the advancement of technologies such as the Internet of Things, 5G, and deep learning, it has facilitated the creation and development of numerous intelligent applications. Intelligent applications necessitate substantial data transmission and computation [1]. The surge in network traffic places

---

The 7th International Conference on Mobile Internet Security (MobiSec'23), Dec. 19-21, 2023, Okinawa, Japan, Article No. 24

[*] Corresponding author: School of Computer Science and Information Security, Guilin University of Electronic Technology, 541004, China, 21031102009}@mail.guet.edu.cn

considerable strain on data centers [2]. For instance, intelligent traffic management requires real-time collection of traffic conditions at each observed road segment. After simple processing at the edge, rapid transmission to data centers for decision-making occurs, and the decision outcomes are promptly fed back to users. The quality of the network directly impacts the user experience of intelligent applications.Network traffic prediction plays an important role in network management and control, such as resource allocation, traffic engineering, and quality of service [3]. The prediction of network traffic is closely linked to the security of the mobile internet. Through the analysis of network traffic patterns, it becomes possible to detect abnormal behaviors and identify potential network threats, such as the propagation of malicious software and intrusion.

There are both linear and nonlinear relationships in network traffic data. Most prediction methods analyze the linear, nonlinear and stochastic components of the time series by decomposing them for the purpose of prediction. Some studies generate sequential data by using GAN models. C-RNN-GAN [4] uses LSTM networks as generators and discriminators for the whole model to generate music using continuous sequential data. Timegan [5] combines unsupervised paradigm with supervised training to improve the accuracy of time series prediction. There are also many research works using transformers for time series prediction, such as Informer [6], Reformer [7], etc. However, the complexity of the transformer is high and some of the original information of the sequence needs to be discarded when increasing the complexity of the transformer.

The wavelet transform is multi-resolution and allows a gradual observation of the signal from coarse to fine [8]. The wavelet packet transform has been widely used as an effective signal processing tool. Wavelet packets are not only suitable for extracting information from images [9], but can also analyze time series effectively [10]. Wavelet transform neural networks can handle nonlinear optimization problems [11], and it simplifies the training by replacing the activation functions of neurons with wavelet basis functions. Moreover, the wavelet transform is able to characterize the local features of the signal in both time and frequency domains, which is beneficial for detecting transients or singularities in traffic sequences.

In this study, we propose a novel neural network TSWNET for network traffic prediction. wavelet transform is used to extract features in traffic sequences and construct wavelet neural block FWDB to capture the dynamic changes of traffic sequences. Experiments show that our model can effectively predict the traffic variation. .

2 Problem formulation

Traffic Forecasting:Assuming that there is a univariate traffic time series $T=\{t_i,1:N\}$, and $t_i \in R$ is the value of the traffic time series at moment i. The ultimate goal of the trained model is to predict the value of the next $\tau \in N$ time step in a given time series, we aim to model the following formulation:

$$t_{t_n:t_n+\tau} = M(t_{1:n} | \Phi)(n<N) \tag{1}$$

In Equation 1, $t_{n:n+\tau}$ is the predicted value spanning a time step of τ. M is the prediction model. $\Phi \in R$ is the learnable parameter of the model for joint learning of T time series. For the traffic time series, the range we use for training is $\{t_i,1:n\}$, the time point t_{n+1} is the prediction start time, the time series $\{t_i,\ n:n+\tau\}$ is used as the target time series, and $\tau \in N$ is the step size of the predicted time.

3 Method

The basic principle of the wavelet transform study is the progressive refinement of the signal on multiple scales through scaling and translation operations. The wavelet transform can be automatically adapted to the requirements of time-frequency signal analysis by focusing on any detailed part of the sequence. The equations are as follows:

$$f(t) = \sum_{j} a_{i0,j} \Phi_{i0,j}(t) + \sum_{i>i0} \sum_{j} \omega_{i,j} 2^{i/2} \Psi(2^{i} t - j) \qquad (2)$$

Where i is the level index, j is the scaling index, Ψ is the mother transform, $\Phi_{i0,j}$ is the scaling function of the coarse scale factors and $a_{i0,j}$ & $\omega_{i0,j}$, j are the scaling functions of detail factors. $\Psi(2^{i} t - j)$ is orthogonal function. The data will be extracted into low frequency components and high frequency components.

Inspired by the fact that wavelet transform can handle time series, we designed a novel neural network TSWNET. this neural network incorporates the process of wavelet packet transform into the neural network to fully exploit the potential features in traffic time series. The overall architecture of our proposed TSWNET neural network is shown in Figure 3. the ultimate goal of TSWNET is to train traffic sequences $T_2\{t_i, 1:n\}$ and derive high-quality prediction results of traffic sequences $T_1\{t_i,\ n:n+\tau\}$.

3.1 Traffic forecasting model

3.1.1 Model Structure

As shown in Figure 1, TSWNET mainly consists of wavelet neural blocks. First, a traffic sequence with a training length of s is used as input and received with a fully connected layer. The traffic sequence enters the wavelet traffic analysis module, and the relevant features in the traffic sequence are extracted. Then a wavelet reconstruction operation is performed. If the length of the traffic sequence used for training is greater than the predicted length, we add a fully connected layer between the FWDB and the final output. This is done in order to prevent losing too much information at one time. Finally, the model is trained using the constructed loss function to generate high-quality traffic prediction sequences. The whole process can be described as Formula (3)

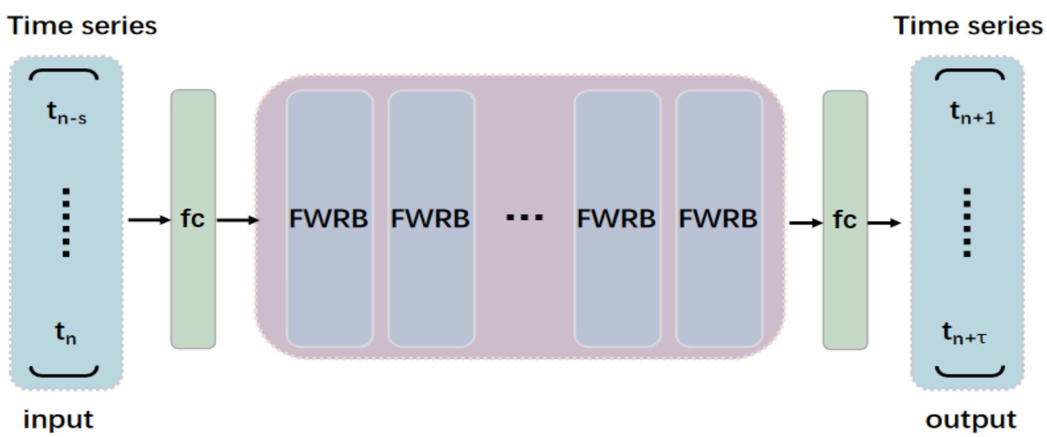

Fig 1: Framework of TSWNET

$$F_w = fc(FWDB(fc(T\{t_i, 1:n\}))) \tag{3}$$

To analyze the traffic sequence and wavelet coefficient information, the traffic wavelet residual block FWRB is designed.As shown in Figure 2, the sequence wavelet coefficient information is extracted by WPT and the wavelet coefficient residual information is learned by Denseblock to enhance the wavelet coefficient information features. After the wavelet coefficient information is passed through four Denseblocks, the wavelet coefficient information is reorganized by IWPT to obtain the feature-enhanced traffic sequence. To further enhance the output of the previous step and to ensure that the perceptual region is further improved without losing information, we add a Denseblock after IWPT. Let ω_l denote wavelet transform WPT, let ω_l^{-1} denote IWPT. Let R represent the process of residuals. This process can be expressed as follows:

$$W_{co} = R(\omega(T); Db(\omega(T))) \tag{4}$$

$$T_s = \omega_l^{-1}(W_{co}) \tag{5}$$

$$T_{W;D} = re(T_s; D(T_s)) \tag{6}$$

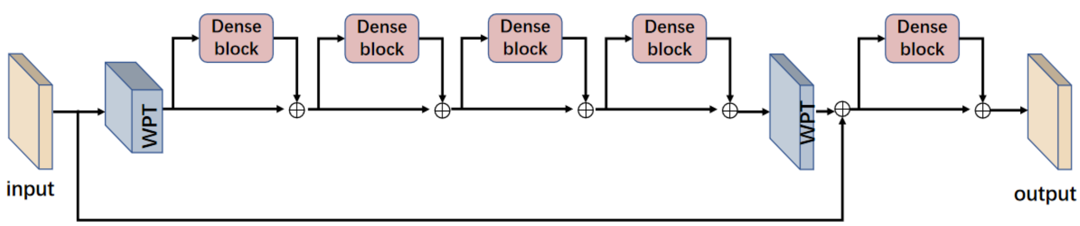

Fig 2: Architecture of the proposed FWRB

Where W_{co} represents the wavelet coefficients after wavelet packet transform and feature extraction. T_s represents the time series after feature extraction. $T_{W;D}$ is the output after one FWRB. And Db is the processing that represents the Denseblock neural block. The structure of Denseblock is shown in Figure 3.

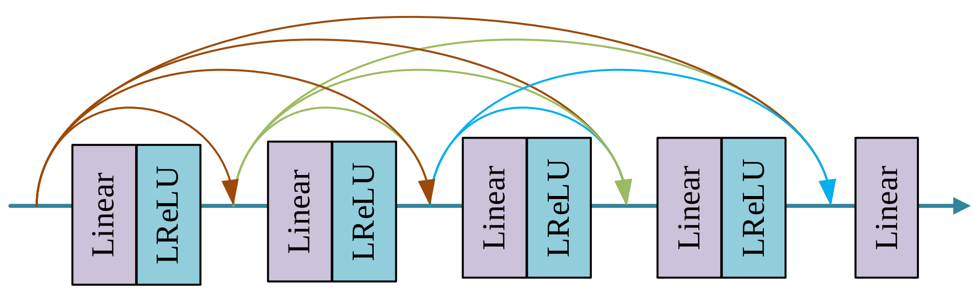

Fig 3: Architecture of Denseblock

This work modifies the structure of the Denseblock, which allows for a tighter network connection and larger capacity. Using such a structure helps to fully analyze the wavelet coefficients.

3.1.2 Loss functions

In this section we propose the loss function L_F for training TSWNET,In order to extract enough features from the traffic series, we formulate L_F as the sum of the wavelet loss $L_W\left(\theta_W\right)$ and the predicted value $L_V\left(\theta_V\right)$ loss. Then, $L_F\left(\theta_W;\theta_V\right)$ can be written as:

$$L_F\left(\theta_W;\theta_V\right)=\alpha L_W\left(\theta_W\right)+L_V\left(\theta_V\right) \tag{7}$$

where α is the coefficient that balance different loss terms, and $L_W\left(\theta_W\right)$ models how close the wavelet coefficient information of the predicted flow series is to the wavelet coefficient information of the real flow series. $L_W\left(\theta_W\right)$ is modified based on Charbonnier, as follows:

$$L_W\left(\theta_W\right)=\sqrt{|T_p-T|^2+\varepsilon_\omega^2} \tag{8}$$

Where T_p is the predicted sequence and T is the true sequence. ε is a scaling paraeter. $L_V\left(\theta_V\right)$ compares the proximity of the predicted series to the value of the true series at each time point. And we use MSE as a penalty function with the following equation.

$$MSE=\frac{1}{n}\sum_{i=1}^{n}|T_p-T|^2 \tag{9}$$

MSE can calculate how close the model's prediction T_p is to the true label T. And amplify the error between the predicted and true values, allowing the model to converge faster.

4 Simulations and Validation

4.1 experiment environment

This work analyzes three datasets. Dataset A belongs to a private ISP with hubs in 11 European cities. Dataset B is from UKERNA1 and represents aggregated traffic from the backbone of the UK academic network. In datasets A and B, we select traffic sequences with traffic collection time span of 5 minutes. The traffic values in the kaggle dataset are too small for the comparison of the final metrics, so these values are scaled up in the same proportions. We used LSTM, BILSTM, RNN and GRU as comparison models. In the evaluation process, we compare the predicted multi-step flow sequence values with the true values. The evaluation metrics used include the root mean square error (RMSE) and the mean absolute error (MAE):

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}|y-\hat{y}|^2} \tag{10}$$

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|y-\hat{y}| \tag{11}$$

where N is the number of predicted values in each group. The average of multiple sets of indicators are used as final result.

4.2 experiment result

Table 2 shows the comparison of TSWNET with other baseline models for multi-step prediction results of traffic sequences in the three datasets of the two datasets.

Model	Metric	A-5M	B-5M	kaggle_m7	kaggle_m14	kaggle_m21
LSTM	RMSE	4.4687	4.8129	15.9622	14.4293	15.9973
	MAE	7.8724	7.5481	24.2375	23.4024	24.7168
RNN	RMSE	5.0761	5.6729	17.9731	15.4767	13.6128
	MAE	9.2757	9.3211	20.2436	25.4134	23.7182
GRU	RMSE	4.9845	4.1126	15.9937	14.3826	15.5762
	MAE	9.1756	9.7887	24.2543	22.3946	23.7129
BILSTM	RMSE	4.0189	5.8092	13.9311	15.4671	12.5717
	MAE	7.0212	9.0692	21.2236	22.4107	20.7112
TSWNET	RMSE	1.1822	1.2835	2.2605	2.1798	2.2090
	MAE	0.3979	0.4185	0.7998	0.7706	0.78102

Table 1: Univariate time-series forecasting results on three datasets

The prediction length in each experiment was 8. From the results, it can be seen that TSWNET performs best on all datasets. The performance of the various baseline models is not very different, and BiLSTM performs better compared to the other baseline models when the predicted values are larger. The performance of LSTM and GRU is almost similar on each dataset. In datasets A and B, the predictions were relatively small data, and TSWNET and the other baseline models did not differ significantly in terms of the difference in metrics. In the three columns of data in kaggle, it can be seen that the error of various baseline models grows faster as the metrics increase. TSWNET, on the other hand, is more stable, and the error grows as the amount of data grows, but TSWNET performs better and is more stable compared to the error growth rate of the baseline model.

5 Conclusions

In this paper, A wavelet transform-based neural network TSWNET is proposed. The model uses the wavelet transform as the main structure of the neural network to extract the variation characteristics in the time series of network traffic. Compared with various previous methods, our proposed method is able to capture the time dependence of the network traffic sequences and obtains the best prediction results on the given data set.

In future work, we hope to improve the performance of traffic prediction and extend this traffic prediction method to other scenarios.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (62162018, 61967005); The Natural Science Foundation of Guangxi (2019GXNSFGA245004); Guangxi Collaborative Innovation Center of Cloud Computing and Big Data Found (YD1901); The Innovation Project of Guangxi Graduate Education (YCSW2022296).

References

[1] Mei J, Lei B, Wang X, et al. *Intelligent Universal Acceleration Framework and Verification for Edge Cloud Applications*[C]//2020 IEEE/CIC International Conference on Communications in China (ICCC Workshops). IEEE, 2020: 47-52.

[2] Varghese B, Wang N, Barbhuiya S, et al. *Challenges and opportunities in edge computing[C]//2016 IEEE international conference on smart cloud (SmartCloud)*. IEEE, 2016: 20-26.
[3] X. Cao, Y . Zhong, Y . Zhou, J. Wang, C. Zhu, and W. Zhang. *Interactive temporal recurrent convolution network for traffic prediction in data centers*[J]. IEEE Access, vol. 6, pp. 5276–5289, Feb. 2018.
[4] Mogren O. *C-RNN-GAN: Continuous recurrent neural networks with adversarial training*[J]. arXiv preprint arXiv:1611.09904, 2016.
[5] Srinivasan P, Knottenbelt W J. *Time-series transformer generative adversarial networks*[J]. arXiv preprint arXiv:2205.11164, 2022.
[6] Zhou H, Zhang S, Peng J, et al. Informer: *Beyond efficient transformer for long sequence time-series forecasting*[C]//Proceedings of the AAAI conference on artificial intelligence. 2021, 35(12): 11106-11115.
[7] Kitaev N, Kaiser Ł, Levskaya A. *Reformer: The efficient transformer*[J]. arXiv preprint arXiv:2001.04451, 2020.
[8] Abry P, Flandrin P. *On the initialization of the discrete wavelet transform algorithm*[J]. IEEE Signal Processing Letters, 1994, 1(2): 32-34.
[9] Wang J, Deng X, Xu M, et al. *Multi-level wavelet-based generative adversarial network for perceptual quality enhancement of compressed video*[C]//Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23 – 28, 2020, Proceedings, Part XIV 16. Springer International Publishing, 2020: 405-421.
[10] Yu F, Yue Q, Yunianta A, et al. *A novel hybrid deep correction approach for electrical load demand prediction*[J]. Sustainable Cities and Society, 2021, 74: 103161.
[11] Wang X, Yu K, Wu S, et al. Esrgan: *Enhanced super-resolution generative adversarial networks*[C]//Proceedings of the European conference on computer vision (ECCV) workshops. 2018: 0-0.