2020 International Conference on Information, Cybernetics,
and Computational Social Systems (ICCSS)

Prediction accuracy improvement of passive optical network traffic by a LSTM model with a new activation function

1st Xiang Zhu[1,2]
1. School of Information Engineering
Guangdong University of Technology
Guangzhou, China
2. College of Electrical and Power Engineering
Taiyuan University of Technology
Taiyuan, China
zhuxiang0294@link.tyut.edu.cn

2nd Ou Xu[1,2]*
1.School of Information Engineering
Guangdong University of Technology
Guangzhou, China
2. *Guangdong Provincial Key Laboratory of Information Photonics Technology*
xuou@gdut.edu.cn

3rd Guoyong Li
College of Electrical and Power Engineering
Taiyuan University of Technology
Taiyuan, China

2020 7th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS) | 978-1-7281-6246-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICCSS52145.2020.9336907

***Abstract*—With the commercialization of the fifth-generation mobile communication technology, the global network traffic is exploding. Network traffic prediction is studied in order to schedule network resources more efficiently and ensure quality of service. However, the traditional neural network traffic prediction model has poor prediction effect on time series, and cannot accurately predict the high speed and rapid change of network traffic. In this paper, a neural network prediction model is proposed with prediction accuracy improvement of passive optical network traffic, based on Long Short-Term Memory (LSTM) with a new activation function--Swish activation function. According to the characteristics of actual passive optical network traffic, the data simulated by simulation software are tested. Compared with BP neural network and recurrent neural network (RNN), the proposed model is proved to display high prediction accuracy.**

Keywords—Passive optical network; network traffic predict; Long Short-Term Memory neural network; Swish

I. INTRODUCTION

Passive Optical Networks (PON) is a widely used broadband access network technology. It is a point-to-multipoint network structure, which has great advantages in quality of service and transmission speed. EPON is a widely used PON structure that supports downlink/uplink transmission speeds of 2.5Gbps/1Gbps. With the rapid development of network, the size of network traffic rate increases sharply. In order to meet current development needs and ensure network service quality, IEEE proposed a new standard IEEE 802.3ca in 2016, aiming to develop a high-speed PON network of 50G to 100G [1]. Because of the structure of PON as point-to-multipoint [2], multiple terminals share one bandwidth when uploading data. In order to avoid data conflicts, ensure the rate and accuracy of data transmission, and ensure the quality of service, we need to find an appropriate bandwidth allocation algorithm.

With the continuous development of computer network, the prediction of network traffic is getting more and more attention. Accurate prediction of network traffic is conducive to reasonable scheduling and arrangement of network resources by network operators and other and to improve the quality of network service.

As for network traffic prediction, many scholars have studied it. Firstly, some scholars proposed a linear prediction model to model network traffic, such as Piecewise Autoregressive Moving Average(ARMA). In [3], Li used the ARMA theory to describe the network traffic for scheduling of optical fiber cores. However, this linear model is suitable for more stable traffic, and not suitable for the high-speed changing network. Later, other scholars developed nonlinear prediction models, such as support vector machine(SVM), Gaussian process(GP) and other models. In [4], Sun used SVM to predict and classify the network congestion. In [5], Xu establish a wireless traffic prediction model by applying the Gaussian Process method. Compared with the linear model, these methods have a good performance for the nonlinear characteristics of current network traffic. SVM requires a small amount of data, but the model parameters are hard to determine, the model reliability is poor, and the GP model has a high requirement for data. In recent years, neural networks and deep learning have been widely concerned by scholars, and they also have very good effect in solving nonlinear problems, including traditional feedback neural network, BP neural network, radial basis neural network RBF and so on. In [6], Li presents a hybrid model based on the combination of wavelet denoising and BP neural network to predict network traffic flow. In [7], Wei takes neural networks and chaotic time series theory as the foundation to propose a prediction model based on RBF neural network optimized by improved gravitation search algorithm. Although such neural networks have achieved good results in network traffic prediction, their disadvantages include: they only consider the current input information in each step of the prediction, and they cannot well extract the characteristics of data for the network traffic, which is closely related to the time series before and after. At the same time, for the network traffic that changes rapidly within a short period of time, these methods cannot follow the network changes well.

For high-speed PON traffic prediction, there are two difficulties at present: 1. Network traffic speed is very fast, with the peak reaching 10Gb/s or even 50Gb/s; 2. Bandwidth allocation needs to be completed in a very short period of time, and the amount of traffic changes rapidly within a short period of time. When calculating the output of each moment, the recurrent neural network(RNN) takes into account the previous input and output information, so it has a good predictive effect for time series such as network traffic [8]. But RNN is prone to gradient disappearance and explosion risk. In this paper, a variant LSTM neural network of RNN is adopted, which overcomes the shortcomings of traditional RNN in gradient disappearance and explosion and achieves good

The research work was supported by the National Key R&D Program of China (2019YFB1803803), the National Natural Science Foundation of China (Grant No.61941107) and The Construction Project of Guangdong Provincial Key Laboratory (2020B121201011)

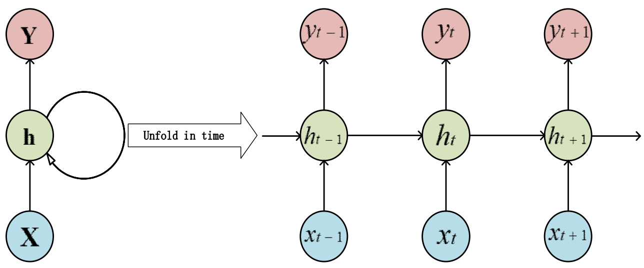


Fig. 1 Structure of RNN Neural Network

results in natural statement recognition and time series prediction [9, 10].

The rest of this paper is organized as follows. Section II introduces the basic concepts of LSTM, and a new activation function Swish is introduced, which can greatly improve the prediction accuracy. Section III, the feasibility of the proposed method is verified by the flow data obtained through simulation. Finally, we conclude this paper.

II. Methods

A. Long Short-Term Memory

The structure of RNN is shown in Fig. 1, where $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{h}$ represent input, output and hidden state. It can be seen from the Fig.1 that the output value of each moment of the RNN neural network is affected by the value of the previous moment, which fully considers the characteristics of the whole time series. However, as the calculation goes on, a large amount of useless information produced long ago is stored in the hidden layer of RNN. At the same time, a serious problem of gradient explosion and gradient disappearance will occur in RNN. In order to solve the defects of RNN, Hochreiter and Schmidhuber improved the hidden layer nodes of RNN in 1997 [11], developing the LSTM. It filters information by introducing forget gates, input gates, output gates and a hidden cell structure, retaining important information and forgetting unimportant information. The structure of LSTM is shown in Fig. 2. Where, h_t represents the output of the node, C_t represents the hidden cell body, which contains the information retained throughout the calculation, and X_t represents the input value. The calculation process of the whole node is as follows:

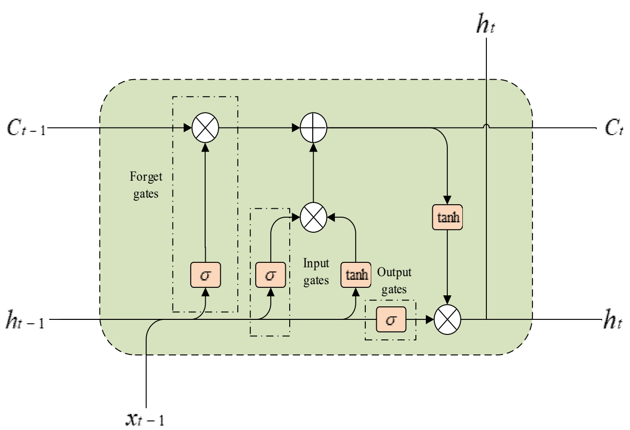


Fig. 2 Structure of LSTM Neural Network

- The forgetting gate is responsible for selecting the information to be forgotten. After h_{t-1} is combined with x_t, a weight matrix f_t is obtained through *sigmoid* function calculation. The size of f_t is between [0,1], where 0 represents forget all information and 1 means keeps.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

- The input gate is responsible for deciding what information to retain.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

- The output gate is responsible for calculating the information to be output.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{3}$$

- As for the most important components of LSTM structure, C_t is generated mainly in the following ways: First, the calculation result of forgetting gate f_t is multiplied by the cell body information C_{t-1} at the previous moment, which selectively discard some old and useless information; Then, the value of h_{t-1}combined with X_t is calculated by *tanh* function and multiplied by i_t to obtain a short-time cell body with temporary information. The cell body information at that time was obtained by adding the two parts.

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{4}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{5}$$

- The cell body information is calculated by *tanh* and multiplied by the output gate result o_t to obtain the final output value of the node.

$$h_t = o_t \cdot \tanh(C_t) \tag{6}$$

Compared with RNN, LSTM reasonably processed with previous information, thus it has greatly improved the experimental simulation effect of the RNN problem of gradient explosion and disappearance. LSTM is not only widely used in the prediction and modeling of time series, but also has good effects for various sequence problems.

B. Swish

As an important part of the neural network, activation function is responsible for mapping the input value to the specified range and space. Perceptron, the predecessor of neural network, adopts linear activation function, but its disadvantage is that it can only solve linear problems. With the introduction of nonlinear activation function, the neural network can fit arbitrary nonlinear problems, which greatly improves the application range of neural network. Common nonlinear activation functions include *sigmoid*, *tanh* and so on. At present, the most commonly used and successful activation is the ReLU function, whose expression is:

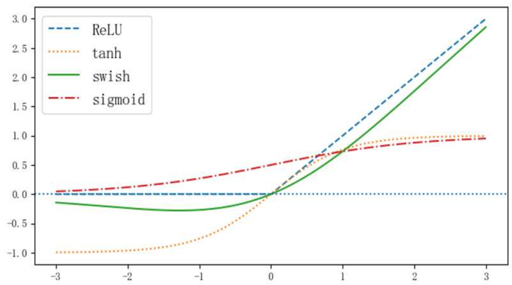


Fig. 3 Comparison of different activation

$$f(x) = \max(x, 0) \quad (7)$$

As can be seen from the expression, when the input value x is less than 0, the output of the activation function is 0, which leads to the necrosis of some neurons in the neural network, so the value cannot be output, and the weight cannot be updated in the subsequent calculation. To solve this problem, the researchers proposed new activation functions such as Leaky ReLU, ELU, SeLU, etc.

In 2017, Google brain team adopted the methods of reinforcement learning and machine learning in [12], searched and tested in the specified search domain, and found a new activation function Swish. The expression of Swish function is as follows:

$$f(x) = x \cdot \sigma(\beta x) \quad (8)$$

Where:

$$\sigma = \frac{1}{1+e^{-x}} \quad (9)$$

When β is zero, the activation function becomes a ReLU function. When the β is infinite, the activation function becomes a linear activation function. It can also be seen from the expression that Swish activation function is easy to differentiate, which will not increase too much computation in the network training process. In [12], the author proved the superiority of Swish activation function over ReLU through the simulation effect on CIFAR data set.

As we can see from Fig. 3, both *tanh* and *sigmoid* functions have upper and lower dead zones. When the training data enters the dead zone, the weights will update slowly or even pause, which is not conducive to network training. The ReLU and Swish functions are linear outputs when the input is greater than 0, which can well retain information. When the ReLU is less than 0, the ReLU output is 0, which may cause a lot of neuronal necrosis and the weight cannot be updated. However, Swish function can retain part of the information and improve the prediction effect.

III. Simulation and Discussion

A. Data sources and processing

In 1993, Leland, Taqqu et al. analyzed the network traffic and found that network traffic had similar properties on different time scales, which is called self-similarity of network traffic [13]. On this basis, researchers propose a number of methods to simulate network traffic. One of the most widely used is the ON/OFF model. The ON/OFF model generates traffic during ON and no traffic during OFF. The duration of ON and OFF needs to meet the heavy-tailed distribution. Through such continuous alternations of ON and OFF, the generated data has self-similarity similar to the actual network traffic. After many studies, it is found that the data sources generated by ON/OFF model have the same characteristics as the traffic data of the access network in reality. In this paper, 32 ON/OFF data sources are superimposed, and the duration of ON and OFF satisfy Pareto(1, 1.4) and Pareto (0.01, 1.2) respectively. In the software, the data source sends random data packets according to the random size, while in the bandwidth allocation, the bandwidth should be allocated according to the size of the traffic data in the current queue. Therefore, in the simulation, it is necessary to count the total amount of traffic in the queue within a certain time. In this paper, the sampling time of 2ms is adopted, and the sum of all data generated within 2ms is taken as the primary traffic data. After 20 seconds of simulation, 10,000 sets of data were obtained. Partial data distribution is shown in Fig. 4. It can be seen from the data graph that the variation range of traffic data within a short period of time is very large, which leads to the difficulty of current traffic prediction. And the average value of the traffic size in the whole data set is around 8Gb/s

After that, the data is normalized and changed to [0,1], which can accelerate the computing speed of the neural network, improving the accuracy and reduce the error. In order to ensure the accuracy of prediction and prevent the occurrence of overfitting, the first 85% data of the data set is taken as the training set and the remaining 15% as the test set. After the training set is used to train the model, the test set is used to verify the accuracy of the model.

B. Evaluation criteria

In order to select a better prediction model, it is necessary to adopt a set of appropriate error expression methods as the criterion of the model. In this paper, three error criteria are selected: MAE, RMSE and 5% accuracy. The expressions of RMSE and MAE are shown in (10) and (11). The 5% accuracy rate refers to: when the error between the predicted value and the actual value is below 5%, the default prediction is correct; The percentage of the total data set that predicts the correct number is 5% accurate.

$$MAE = \frac{1}{M}\sum_{m=1}^{M}\left|y_m - \hat{y}_m\right| \quad (10)$$

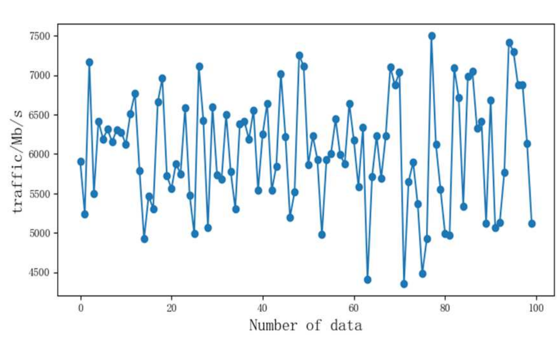
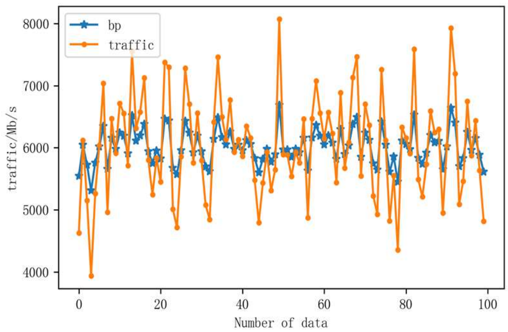
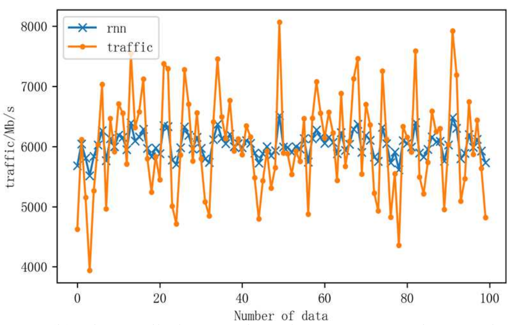


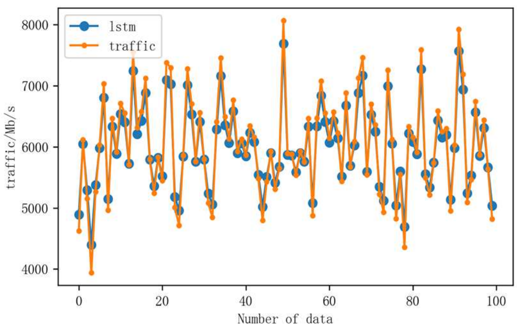
Fig. 4 Network traffic data



(a) The prediction results using BP neural network



(b) The prediction results using RNN neural network



(c) The prediction results using LSTM neural network

Fig. 5 Comparison of prediction results of different models

$$RMSE = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(y_m - \hat{y}_m\right)^2} \quad (11)$$

Where, y is the actual value, $\hat{y}$ is the predictive value.

C. Stimulation results and discussion

First, the classic BP neural network is used, and the traditional RNN neural network is compared with the LSTM method used in this paper, as shown in Fig. 5.

As we can see from the Fig.5, although the prediction effect of BP neural network and traditional RNN network can follow the change trend of the actual data, it cannot track the rapid and large changes of the actual network traffic, while the prediction result of LSTM is very close to the actual traffic value. The prediction error is show in Table Ⅰ.

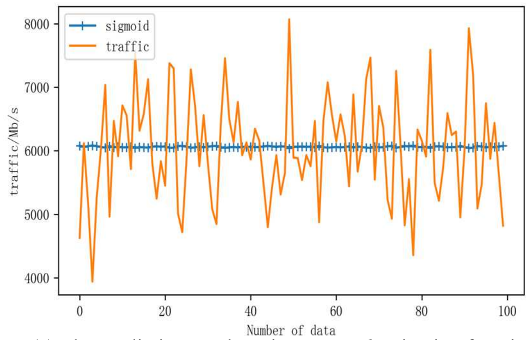
The active function is an important part of the neural network, and the performance of the neural network is different when different activation functions are used. Fig.6

TABLE I. COMPARISON OF DIFFERENT MODEL ERRORS

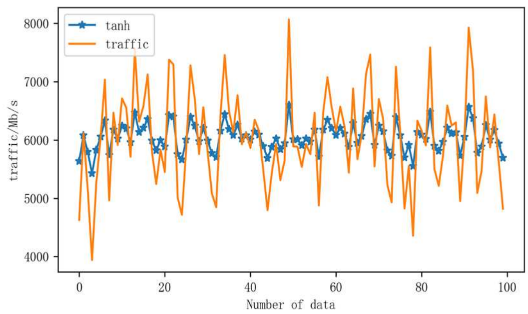
	RMSE	**MAE**	**5% accuracy**
BP	0.092	0.081	19%
RNN	0.088	0.071	21%
LSTM	0.013	0.013	96%

compares the network traffic prediction errors of relatively common activation functions *sigmoid*, *tanh*, ReLU and Swish.

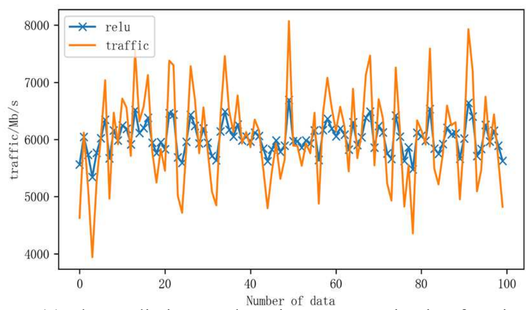
As we can be seen from Fig. 6, the *sigmoid* function changes the prediction results within a small range due to the



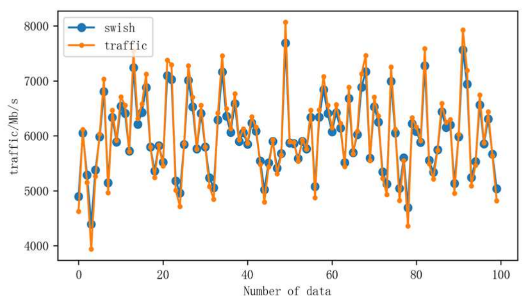
(a) The prediction results using *sigmoid* activation function



(b) The prediction results using *tanh* activation function



(c) The prediction results using ReLU activation function



(d) The prediction results using Swish activation function

Fig. 6 Comparison of results with different activation functions

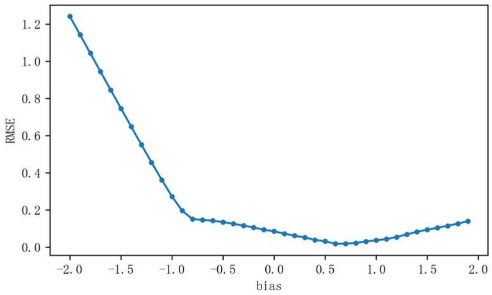


Fig. 7 The influence of bias on prediction error

existence of upper and lower dead zones. The predicted results of *tanh* and ReLU functions can well follow the changes of the actual flow, but the range of changes is still relatively small. The prediction result of Swish activation function is very close to the real traffic value, and the prediction effect is very good, which proves the superiority of Swish function. Table Ⅱ shows the final prediction errors of the four activation functions.

TABLE Ⅱ. COMPARISON OF DIFFERENT ACTIVATION FUNCTIONS ERRORS

	RMSE	**MAE**	**5% accuracy**
sigmoid	0.507	0.489	0%
tanh	0.135	0.107	15%
ReLU	0.098	0.079	19%
Swish	0.013	0.013	96%

In this simulation, after many tests, it was found that adding a bias value after activation function could improve the prediction accuracy and reduce the prediction error. Fig. 7 shows the variation of prediction error with the change of bias. It can be seen from the figure that with the continuous increase of bias, the prediction error will decrease first and then increase. When the bias is 0.7, the prediction error is the smallest and the prediction effect is the best.

IV. CONCLUSION

In this paper, LSTM neural network model is used to predict the traffic data of access network simulated by software. It can be seen from the simulation results that the LSTM neural network prediction model with Swish function as the activation function used in this paper solves the difficulties in short-term traffic prediction: the high speed traffic and rapid change. The traffic at the next moment can be quickly and accurately predicted at the current moment. From the final experiment, it can be seen that the prediction accuracy of the method used in this paper is very high, and it can play a very good role in the later bandwidth allocation system, which needs rapid prediction and response.

REFERENCES

[1] IEEE. 50G-EPON Task Force, Physical layer specifications and management parameters for 25 Gb/s and 50 Gb/s passive optical networks:IEEE P802.3ca: 2018[S/OL]. [2019-08-06]. http://www.ieee802.org/3/ca/

[2] Ragheb A, Fathallah H. Candidate modulation schemes for next generation-passive optical networks (NG-PONs)[C]// High Capacity Optical Networks and Emerging/Enabling Technologies. IEEE, 2013.

[3] H. Li, F Meng, J Kong, G Wang, B Lu, Y Liu, et al., "An ARMA-Based Traffic Abnormal Identification Approach for Remote Smart Scheduling of Optical Fiber Cores," 2019 IEEE International Conference on Industrial Internet (ICII), Orlando, FL, USA, 2019, pp. 126-131, doi: 10.1109/ICII.2019.00034.

[4] Sun Q, Lan H, Yang X. Prediction of Optical Network Congestion Based on SVM[C]// 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT). IEEE, 2019.

[5] Xu Y, Xu W, Yin F, Lin J, Cui S. High-Accuracy Wireless Traffic Prediction: A GP-Based Machine Learning Approach[C]// GLOBECOM 2017 - 2017 IEEE Global Communications Conference. IEEE, 2017.

[6] Li Y, Huang J, Chen H. Time Series Prediction of Wireless Network Traffic Flow Based on Wavelet Analysis and BP Neural Network[J]. Journal of Physics Conference Series, 2020, 1533:032098

[7] Wei, D. Network traffic prediction based on RBF neural network optimized by improved gravitation search algorithm. Neural Comput & Applic 28, 2303–2312 (2017). https://doi.org/10.1007/s00521-016-2193-z

[8] Wen Y, Zhu G. Prediction for Non-Gaussian Self-Similar Traffic with Neural Network[C]// World Congress on Intelligent Control & Automation. IEEE, 2006.

[9] Chen Q, Zhu X, Ling Z, Ling Z, Wei S, Jiang H, Inkpen D. Enhanced LSTM for Natural Language Inference[J]. 2016.

[10] Sagheer A, Kotb M. Time series forecasting of petroleum production using deep LSTM recurrent networks[J]. Neurocomputing, 2019, 323(JAN.5):203-213.

[11] Hochreiter S, Schmidhuber J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8):1735-1780.

[12] Prajit Ramachandran, Barret Zoph, and Quoc V Le. 2017. Searching for activation functions. arXiv preprint arXiv:1710.05941 (2017).

[13] Leland W E, Taqqu M S, Willinger W, Wilson D V. On the Self-Similar Nature of Ethernet Traffic[C]// ACM, 1995: p.203-213.