Deep Learning Network Traffic Prediction based on Bayesian Algorithm Optimization

Yuantao Lv *

School of North China Electric Power University, Beijing, China, 100096, China

* Corresponding author email: 120201090123@ncepu.edu.cn

Abstract. In order to solve the problem of low prediction accuracy caused by artificial determination of hyperparameters of network traffic prediction model, this paper applies Bayesian optimization algorithm (BOA) to the determination of hyperparameters of neural network, and proposes a CNN-LSTM network traffic prediction model based on Bayesian optimization algorithm. The local features and long-term development trends of the data are mined by combining the Convolutional Neural Network (CNN) and the Long Short-Term Memory Neural Network (LSTM), and the Bayesian optimization algorithm is used to calibrate the hyperparameters. This paper uses real network traffic data to test and compare with LSTM, GRU and other models. Experimental results demonstrate that the proposed BOA-CNN-LSTM model reduces the RMSE and MAE by 21.17% and 27.52%, compared to the traditional network traffic prediction model. The results presented in this paper show that the Bayesian optimization algorithm can more quickly and accurately calibrate the hyperparameters of the model, which provides a new method for predicting network traffic time series data in the future.

Keywords: 5G; Network Traffic Prediction; Bayesian Optimization Algorithm; Convolutional Neural Network; Long Short-Term Memory Neural Network.

1. Introduction

With the continuous breakthroughs in communication technology, wireless communication systems have advanced to the fifth generation. The service object of 5G has been oriented to the Internet of things. In September 2015, the ITU officially confirmed three 5G application scenarios, namely eMBB, uRLLC and mMTC [1]. 5G has ushered in the era of intelligent connectivity of everything. It serves three application scenarios and adopts the method of network slicing. 5G has technical advantages such as ultra-low latency, ultra-large bandwidth, and massive connectivity, which makes 5G the best in a wide range of industry scenarios. An important parameter that reflects the state of the network is the network traffic. Accurate prediction of 5G network traffic is of great significance for improving the utilization efficiency of network resources, reducing energy consumption and promoting the implementation of the "double carbon" plan.

As early as 2002, Sang demonstrated the predictability of network traffic when studying computer network traffic [2]. After that, many investigators have proposed many models in the field of network traffic forecast, such as linear time series models, nonlinear time series models, and hybrid models [3]. Zou *et al.* [4] proposed a network traffic forecast approach based on ARMA models. Wu *et al.* [5] built a BP neural network model for network traffic time series to forecast network traffic, but this approach has the disadvantage that the number of hidden layers is difficult to determine. For the past few years, deep learning has received a large amount of attention and research by scholars in the field of time series prediction. Zhang [6] proposed an urban-scale wireless traffic prediction model based on deep convolutional neural networks. CNN can extract local spatial features of data, but it has high computational complexity, large redundancy and no memory characteristics [7]. LSTM can learn long-short period information in time series, and has achieved good results in some prediction territory [8-9]. However, due to the high bandwidth peak and fast rising speed of 5G network traffic, traditional LSTM models cannot obtain information about burst traffic in advance during training, which makes the prediction accuracy poor [10]. To extract the long-term correlations of network traffic, Zhang Song proposed a cascaded bidirectional minimal gated recurrent unit model [11]. Li *et*

al. [12] proposed a network traffic anomaly prediction approach through deep learning. It combines the advantages of LSTM and CNN, and obtains high accuracy network security event prediction results.

In this paper, we propose a CNN-LSTM 5G traffic prediction model based on Bayesian algorithm optimization for the above series of cases. The following are the innovations of this article:

(1) BOA is used to obtain the optimized hyperparameters. Finding the optimal solution of the function by Bayesian algorithm can save time and lead to higher accuracy.

(2) Find the spatial correlation of data through CNN. In order to deeply explore the spatial features of 5G traffic data, this paper applies CNN to extract the local spatial features of data.

(3) The LSTM is applied to extract temporal features of the data. Because the LSTM can better deal with the influence of large time scale data, this paper applies LSTM to mine the time characteristics of 5G traffic data. Through the combination of CNN and LSTM, spatio-temporal feature extraction is achieved, which makes the prediction of 5G network traffic more accurate.

2. CNN-LSTM 5G Traffic Prediction Model based on Bayesian Optimization Algorithm

2.1 CNN

CNN is a feedforward multi-layer neural network, which is widely used in deep learning territory such as disease analysis and face recognition. Because the convolutional neural network has good feature extraction ability, and considering that 5G network traffic has the characteristics of high bandwidth and high frequency, this paper chooses to use CNN to extract the features of 5G network traffic, so as to achieve accurate prediction of network traffic bandwidth. The CNN graph is shown in Fig.1. The CNN consists mainly of input layers, convolutional layers, pooling layers, fully connected layers, and output layers. The input feature vector can be a multidimensional group of vectors, using local sensing and weight sharing [13]. Among them, the most important part of the convolution layer is the convolution kernel, also known as the feature filter, which enables the convolution layer to quickly and accurately extract features from the input data to form a feature map. The pooling layer can decrease network complexity, reduce training parameters, and prevent overfitting of training results. The fully connected layer merges the processed data and calculates the classification and regression results. Using the strong local feature extraction ability of CNN, the potential features in the input data can be deeply mined.

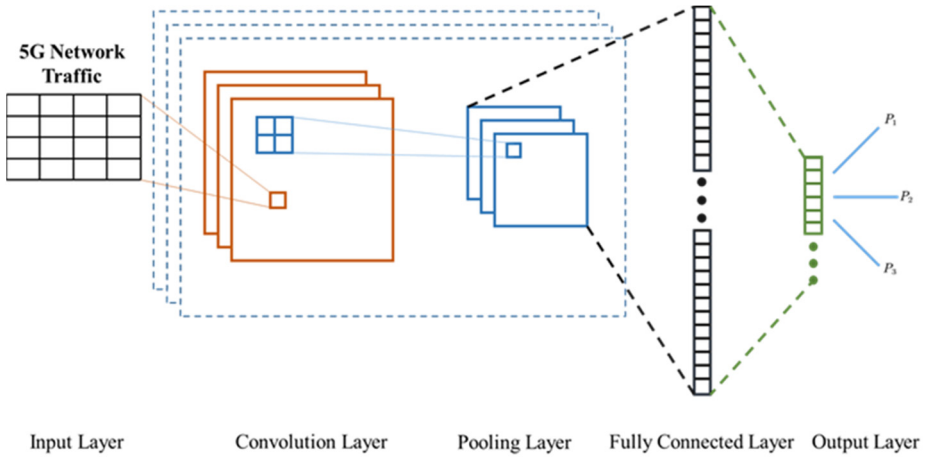


Fig 1. CNN

2.2 RNN

RNN is mainly used to deal with the prediction of continuous data. RNN has good time series learning ability. When predicting 5G traffic, it is necessary to consider both short-term data and long-

term historical data trends. Therefore, recurrent neural networks are used to predict 5G network traffic in this paper. There are three different RNN models for recurrent neural networks, which are the simple RNN model, the LSTM model and the GRU model.

2.2.1 Simple RNN Model

The structure of the simple RNN model is shown in Fig.2. The most significant feature of the RNN is that it focuses on the continuous growth and progress of each neuron in the hidden layer in the time dimension. For the RNN, its output can be expressed as follows:

$$S_t = f(W_{in}X + W_S S_{t-1} + b) \tag{1}$$

It can be seen that RNN has one more $W_S S_{t-1}$ output than the previous ordinary neuron model, which is an iterative relation between the hidden layers at $t-1$ and t. That is, neural networks have some memory capacity.

Another diagram of the RNN is shown in Fig.3. Here, x_t and y_t are the input and output at time t, respectively; S_t is the hidden layer cell at time t; W is the weight matrix, U is the input transformation matrix and V is the output transformation matrix.

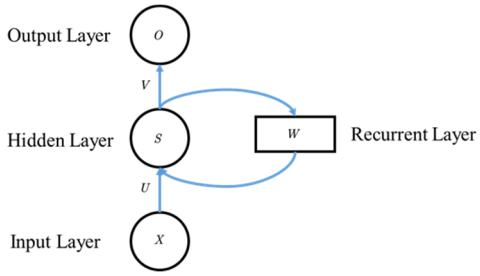


Fig 2. Network structure(RNN)

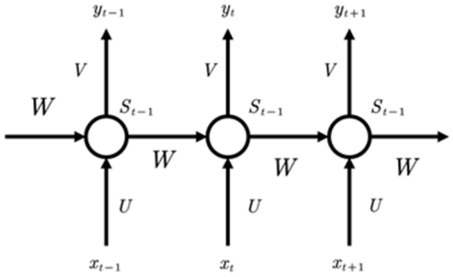


Fig 3. Simple RNN

2.2.2 LSTM Model

In dealing with data with long-term dependence on time, such as weather forecast, language recognition, the traditional recurrent neural network RNN is in straitened circumstances. LSTM can effectively overcome the gradient vanishing and gradient explosion problems in RNN.

The LSTM model is shown in Fig.4. There are three sigmoid functions and two tanh functions in the model. The sigmoid function is represented by $\sigma(x) = \frac{1}{1+e^{-x}}$, which can return the value of (0, 1) interval. The tanh function can return the value of (-1, 1) interval. The sigmoid function is a commonly used nonlinear activation function in machine learning, which is used as a gate to control the ratio of information outflow. The tanh function can control the direction of the information and lock the value in the (-1, 1) interval. In the LSTM model, three sigmoid functions manage three gates: the forget gate, the input gate and the output gate. Compared with the simple RNN model, LSTM adds the element of memory state C. The forget gate selects the previous memory; the input gate is used to obtain the current memory state value; the output gate acquires the current output value. The formulas are as follows.

$$C_t = C_{t-1} \cdot f_t + u_t \cdot i_t \tag{2}$$

$$H_t = o_t \cdot \tanh(c_t) \tag{3}$$

C_t and H_t will flow into the $t+1$ period, thus participating in the next calculation. See Reference [14] for details, no more details here.

2.2.3 GRU Model

The GRU model is based on the LSTM model and its structure is simplified as shown in Fig.5 [15]. Two sigmoid functions represent update gate and reset gate respectively. The GRU model synthesizes the input gate and the forgetting gate in the LSTM into one update gate. The GRU model considers

both the historical information part and the new information part when calculating the current state value H_t. The update gate determines the proportion of the previous state entering the current period, which is represented by z_t in the figure. When it is multiplied by the previous state value H_{t-1}, the part of the new state that belongs to the historical information is obtained. The reset gate determines the forgetting ratio of the previous information, which is denoted by r_t in the figure. It is multiplied by the previous state value H_{t-1} and added to the current input information part x_t, and then converted to the current information data by the tanh function. The current information data is multiplied by the corresponding ratio $(1-z_t)$, to obtain the fraction of new information in the new state. The historical information part and the new information part are added to obtain the final current state value, namely:

$$H_t = H_{t-1} \cdot z_t + (1 - z_t) \cdot \tanh(w_x \cdot x_t + w_h \cdot H_{t-1} \cdot r_t) \tag{4}$$

GRU is computationally more efficient than LSTM. GRU has fewer parameters, so it only needs less data to generalize. However, it is worth noting that the powerful expressive power of LSTM may yield better results when the GRU and LSTM models are in front of sufficient data.

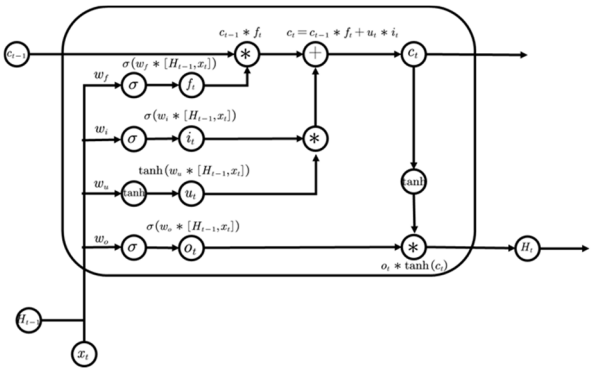


Fig 4. LSTM model unit structure

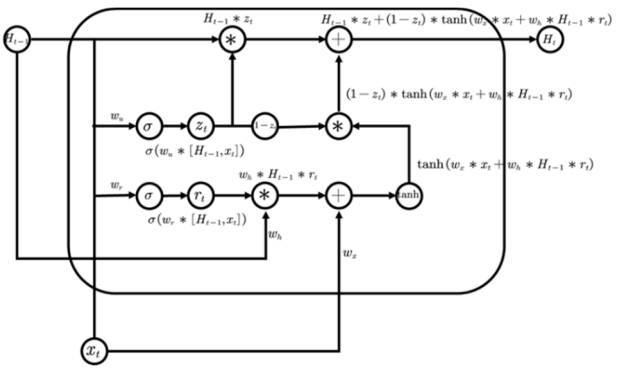


Fig 5. GRU model unit structure

2.3 Bayesian Optimization Algorithm

Bayesian statistics has a wide range of far-reaching applications in many territory, such as parameter estimation, model prediction, and posterior probability [16-17]. When applying the LSTM neural network, there are such a batch of parameters that cannot be obtained through the training data and need to be preset. Such parameters are called hyperparameters. Usually, the hyperparameters need to be set artificially and optimized by the final training results. When the LSTM model is data-driven for learning, the appropriate number of hidden layers, learning rate, and other hyperparameters can be selected to solve the problem of overfitting in the model. This paper uses Bayesian optimization algorithm to select hyperparameters with good results.

The random variable H represents the parameters of the probability model, and the random variable D represents the data set. Then the Bayesian theorem can be expressed as:

$$P(H \mid D) = \frac{P(D \mid H)P(H)}{P(D)} = \frac{P(D \mid H)P(H)}{\int_H P(D \mid H')P(H')dH'} = \frac{P(D,H)}{\int_H P(D,H')dH'} \tag{5}$$

For the neural network, H can be the trained neural network parameter Θ, and D can be the training data. With this definition, the notion of Bayes can be generalized:

$$P(\Theta \mid D) = \frac{P(D_y \mid D_x, \Theta)P(\Theta)}{\int_\Theta P(D_y \mid D_x, \Theta')P(\Theta')d\Theta'} \propto P(D_y \mid D_x, \Theta)P(\Theta) \tag{6}$$

The Bayesian method can predict future data. In the case of a given training data set D, the prediction of future data x is obtained by Bayesian method [16]:

$$P(x \mid D) = \int_{\Theta} P(x, \Theta \mid D) = P(x \mid \Theta, D) P(\Theta \mid D) \tag{7}$$

In the case of a given model, $P(x \mid \Theta, D)$ is usually reduced to $P(x \mid \Theta)$, since the data is sampled from independent and identically distributed samples.

Bayesian optimization includes statistical models from hyperparameter values to objective functions evaluated on the validation set. In Bayesian optimization, observations are collected to display the machine learning model as few times as possible while displaying as much information about the function as possible, especially the best position. The core of Bayesian optimization algorithms are probabilistic surrogate models and acquisition functions. The probabilistic surrogate model is used to represent the unknown objective function. Based on the known data, starting from the hypothesis prior, the more accurate surrogate model is eventually obtained by increasing the amount of information and modifying the prior through successive iterations. The acquisition function is constructed based on the posterior probability distribution, and the position of the next evaluation point is selected by maximizing the acquisition function. In addition, the effective acquisition function can ensure the selection of the evaluation point sequence to minimize the total loss.

2.4 Model Framework

In order to solve the unpredictable characteristics of large bandwidth and high peak values of 5G network traffic, this paper uses the CNN-LSTM network model consisting of two neural networks to design the prediction. This model combines the advantages of CNN and LSTM, which can not only mine its potential features in time series data, but also have certain memory ability. Also, BOA is used to predict the probability distribution of the function value at any point by using the Gaussian process. The core of BOA is to use the acquisition function and the probabilistic proxy model to obtain the optimal solution by using the approximate Bayes theorem.

Initially, the call order data is preprocessed. Then, the preprocessed data is trained by the CNN-LSTM neural network, and some hyperparameters are selected by Bayesian optimization algorithm. Finally, the output data is denormalized to complete the network traffic prediction. Fig.6 is the flowchart of this paper.

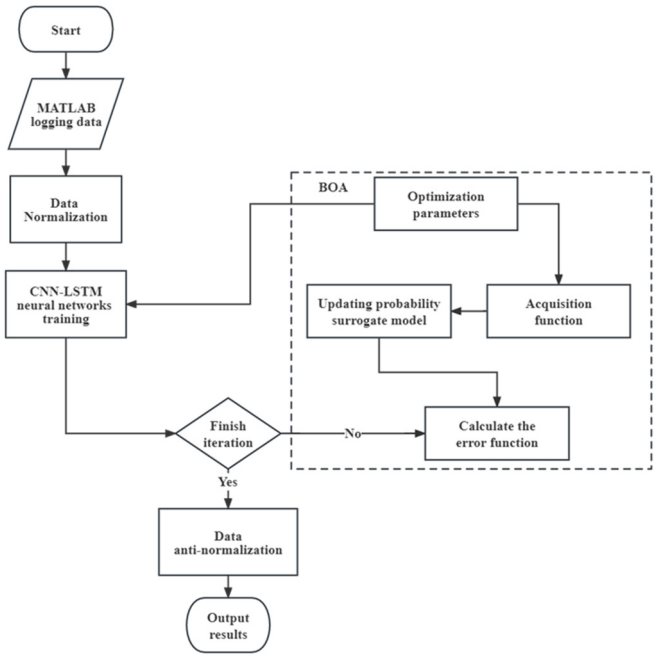


Fig 6. Model of CNN-LSTM network based on Bayesian optimization algorithm

3. Model Hypothesis and Example Application

3.1 Data Introduction

The data used in this paper is from the open-source dataset provided in the 'Big Data Challenge' by Telecom Italia. The dataset describes data on telecommunications, weather, and social networks in the city of Milan and the province of Trentino. The data in are the number of CDRs generated in a given time interval. The data consists of 62 days of base station traffic data in Milan. For the sake of security, Telecom Italia desensitized the data, and encrypted the data to a certain extent under different data distribution and statistical characteristics. The specific treatment is as follows:

$$D'_i(t) = D_i(t) \cdot k \tag{8}$$

Among them, $D'_i(t)$ means the public data set, $D_i(t)$ means the collected data, and k means a constant, which is defined by Telecom Italia. This method hides the real situation reflected by the data.

3.2 Data Pre-processing

When applying the CNN-LSTM model, the input data need to be preprocessed. In this paper, the normalization method is adopted, and the calculation formula is as follows:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{9}$$

Among them, X denotes the raw data of the sample, X' means the normalized data, $X_{\max}$ is the maximum value of the sample, and $X_{\min}$ is the minimum value of the sample. After normalization, the values of the original data are within [0, 1].

3.3 Evaluation Index of Model

In order to better analyze and evaluate the prediction results and performance of the model, this paper uses mean absolute error (*MAE*), root mean square error (*RMSE*), and R^2(*R Squared*) as evaluation indicators [18-19]. The mathematical expression of the evaluation index is as follows.

MAE is in the range of $[0, +\infty)$ .The larger the error, the greater the *MAE*. Here is the formula for the *MAE* calculation.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{10}$$

RMSE is in the range of $[0, +\infty)$. It can reflect the degree of dispersion of the trained model.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{11}$$

R^2(*R* Squared), R^2 mainly reduces the regression results to between 0 and 1, allowing scholars to compare the prediction results of different problems. When the fitting degree of the model is ideal, the value of R^2 will be close to 1. The mathematical expression of R^2 is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{12}$$

where n means the number of data; y_i represents the actual value of the data; $\hat{y}_i$ means the predicted value of the model.

3.4 Example Analysis based on Model

3.4.1 The Hardware Facilities of this Paper

The hardware configuration of the experimental environment is: Intel(R) Core (TM) i7-10510U CPU @ 1.80GHz 2.30GHz, 512G memory, 64-bit operating system. The software configuration is: MATLAB R2021a.

3.4.2 The Initial Parameters and Parameter Adjustment Process of this Paper

The neural network learning and training process is set to a dataset ratio of 0.8 and the MaxEpochs parameter is set to 10. The convolution layers of CNN in the network model are 2 layers. The number of convolution kernels of convolutional layer 1 and convolutional layer 2 is set as [8 64], and the optimal parameters are selected by Bayesian optimization algorithm. The convolution kernel size of convolution layer 1 and convolution layer 2 is set to [3 16], that is, the optimal parameters are selected in the range of 3×3 to 16×16. The time lag order range is [3 9], and the value range of learning rate optimization is $[1\times10^{-5}, 1\times10^{-1}]$.

3.4.3 Prediction Result Analysis

The proposed CNN-LSTM model based on Bayesian optimization algorithm is used to extract features and make predictions on the training dataset. The algorithm iteration process of the hyperparameter optimization part is as follows.

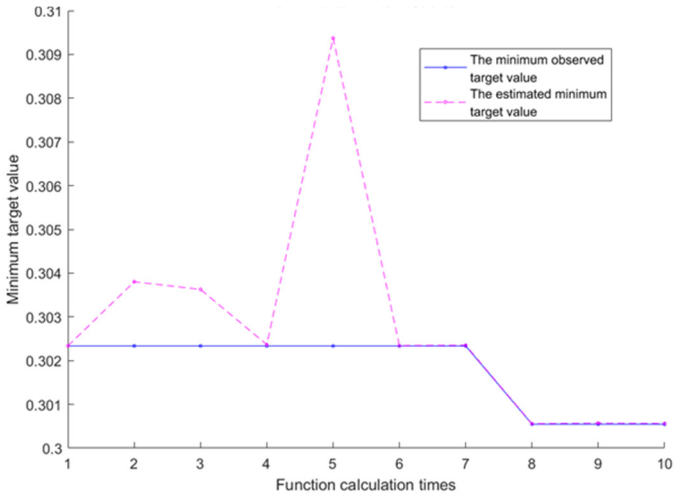


Fig 7. Algorithm iteration process

It can be seen from Fig.7 that Bayesian optimization balances 'exploration' and 'utilization' by the acquisition function. One is to explore the location region as much as possible, and the other is to excavate points around known points to find local optimal solutions. The Bayesian optimization algorithm in this paper explores the region around the minimum objective value in the first six evaluations of the function. In the last four calculations, the observed minimum target value is roughly the same as the estimated minimum target value, reaching an acceptable area. At this point, the Bayesian optimizer exits the iteration and returns the results there.

Table 1. Optimal hyperparameters in the model

Hyperparameter name	Hyperparameter value
NoFilter1	34
NoFilter2	11
FilterSize1	16×16
FilterSize2	12×12
Lag	3
Learningrate	1.34×10^{-4}

To further investigate the training situation and results of the proposed model, the hyperparameter values of the model are recorded and the prediction results are presented in the form of graphs. The optimal hyperparameters of the coupled model are shown in Table 1, and the local display of the network traffic predicted by the model is shown in Fig. 8.

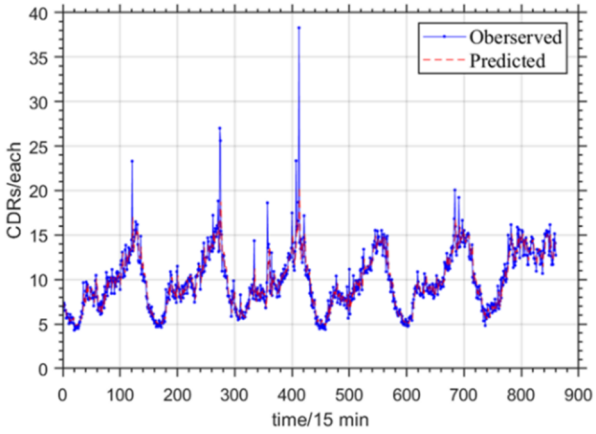


Fig 8. The model predicts the local display of network traffic

It can be concluded that the value of the symmetric mean absolute percentage error of the model is 2.3789%. The accuracy obtained by using Bayesian optimization algorithm to determine the hyperparameters is relatively high, and it has better global convergence and optimization ability.

The relative error diagram of the model test is shown in Fig. 9, and the scatter diagram of the prediction model is shown in Fig. 10. The value of *MAE* in the test is 0.9794, the value of *RMSE* in the test is 1.6290 and the value of R^2 is 0.88085.

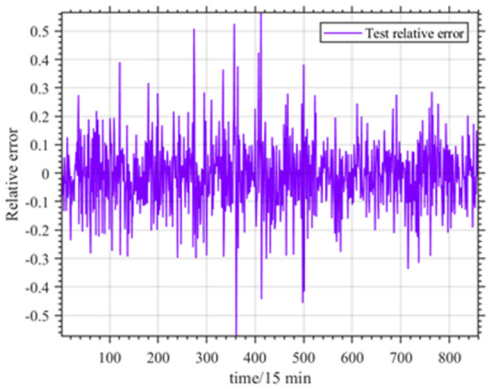


Fig 9. Model test relative error

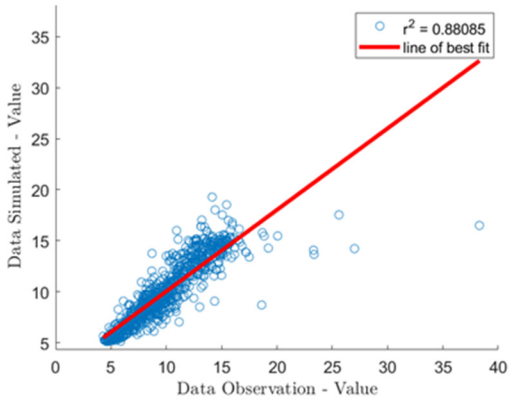


Fig 10. Scatter plot of prediction model

From the results, we can see that the proposed CNN-LSTM model based on the Bayesian optimization algorithm provides a good fit. Although it is not perfect in the peak prediction of CDRs data, the overall trend prediction of data traffic is ideal, and the index values of the prediction performance are in line with expectations. The model in this paper is shown to perform well in predicting 5G network traffic data.

3.5 This Article Compares Model Settings and Results

The training dataset is predicted by the GRU model, the CNN-LSTM model and the CNN-LSTM model based on Bayesian optimization algorithm. The algorithm of parameter update iteration in the GRU network model is Adam, the maximum number of iterations MaxEpochs is set to 100, the learning rate is set to 0.01, and there are 100 neurons in each layer of the GRU network. The experimental results are shown in Table 2, and the comparison of the various prediction models is shown in Fig.11.

Table 2. Evaluation indexes of different prediction models

Model	*RMSE*	*MAE*	*sMAPE*	*R* Squared
LSTM	2.0664	1.3513	6.9267%	0.69021
GRU	2.0174	1.2994	6.7354%	0.67879
CNN-LSTM	1.7416	1.0241	4.7213%	0.82431
BOA-CNN-LSTM	**1.6290**	**0.9794**	**2.3789%**	**0.88085**

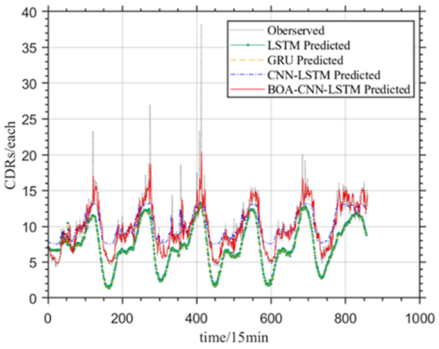


Fig 11. Comparison of CDRs network traffic prediction of different models

The experimental results show that the training results obtained by LSTM and GRU are close, and it can be approximately considered that the training effects of the two are equivalent. Compared with the LSTM, the training results of the CNN-LSTM and the BOA-CNN-LSTM have been significantly improved in accuracy. The *RMSE* and *MAE* of the CNN-LSTM model were reduced by 15.72% and 24.21%, and the *RMSE* and *MAE* of the BOA-CNN-LSTM model were reduced by 21.17% and 27.52%, indicating that the CNN can extract the spatial correlation of 5G network traffic, so as to achieve more accurate prediction of 5G network traffic.

Comparing the CNN-LSTM model with the CNN-LSTM model based on Bayesian optimization algorithm, it can be concluded that the accuracy of using Bayesian optimization algorithm to determine hyperparameters is higher than that of artificially determining hyperparameters, and the selection gradient is smaller. The optimal hyperparameters selected by the optimization algorithm have better global convergence ability. The indicators of the BOA-CNN-LSTM model are better than the CNN-LSTM model, indicating that the Bayesian optimization algorithm has more efficient optimization capabilities. Meanwhile, Fig. 8 shows that the BOA algorithm can quickly update the fitness values and approach the optimal solution faster.

Finally, Fig. 11 shows that the predictions of the GRU and LSTM models are not ideal. Compared with the LSTM model and the GRU model, the predicted values of the CNN-LSTM model and the BOA-CNN-LSTM model are more suitable for the actual network traffic data. Among them, the BOA-CNN-LSTM model is better than the CNN-LSTM model in predicting the peak and valley values of network traffic. The hyperparameters determined by the Bayesian optimization algorithm enable the model to better approximate the maximum and minimum values of the actual data. At the same time, the trend of network traffic can also be perfectly fitted. In summary, among the four models, the proposed CNN-LSTM model based on the Bayesian optimization algorithm has the best prediction accuracy and overall performance. Hyperparameter optimization improves the predictive power of neural networks significantly.

4. Conclusion

In this paper, CNN is combined with LSTM, and BOA is applied to the hyperparameter optimization of CNN-LSTM model. The CNN-LSTM network traffic prediction model based on the Bayesian optimization algorithm is proposed, which significantly promotes the prediction accuracy and performance of the model. In contrast to artificially determining hyperparameters, the Bayesian optimization algorithms address to some extent the problem of low prediction accuracy of artificially determining hyperparameters. Better hyperparameters improve the depth of the deep learning

network and make the model more robust. It is worth noting that the network traffic prediction model proposed in this paper provides a new mentality for predicting time series data. In the future, we can consider combining factors such as delay to predict 5G network traffic, and further improve the prediction ability of the prediction model to 5G network traffic.

Acknowledgments

The author would like to appreciate the anonymous referees for their invaluable insights.

References

[1] ITU-RM 2083-0.IMT vision, framework and overall objectives of the future development of IMT for 2020 and beyond. ITU-R, Document 5/199-E, 2015.

[2] SANG A, LI S. A predictability analysis of network traffic. Computer Networks,2002,39(4): 329-345.

[3] Xiong Hao, Liu Jiayong, Wang Junfeng. Network traffic prediction based on neural network and autoregressive model. Computer Applications, 2021, 41 (S1): 180-184.

[4] Zou Baixian, Liu Qiang. Network traffic prediction based on ARMA model. Computer Research and Development, 2002 (12): 1645-1652.

[5] Wu Wenqing, Zhao Liming, Liu Jiakun. Network traffic BP neural network prediction based on chaos theory. Computer Engineering and Application, 2005 (17): 3-6.

[6] Zhang Chuanting. Urban-scale wireless traffic prediction based on deep learning. Jinan: Shandong University, 2019.

[7] Kang Mengxuan, Song Junping, Fan Pengfei, Gao Bowen, Zhou Xu, Li Zhuo. Research on network traffic prediction based on deep learning. Computer Engineering and Applications, 2021, 57 (10): 1-9.

[8] Lu Haipeng, Yang Fan. A network traffic prediction model based on wavelet transformation and LSTM network // Proc of the 9th IEEE International Conference on Software Engineering and Service Science. Piscataway, NJ: IEEE Press,2018:1131-1134.

[9] Liang Sisheng, Nguyen L, Jin Fang. A multi-variable stacked long-short term memory network for wind speed forecasting/ /Proc of International Conference on Big Data. Piscataway, NJ: IEEE Press,2018:4561-4564.

[10] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation,1997,9(8):1735-1780.

[11] Zhang Song. Research on network traffic prediction based on improved gated recurrent neural network. Changsha: Hunan University, 2019.

[12] Li Jiayue, Zhao Bo, Li Xiang, Liu Hui, Liu Yifan, Zou Jianwen. Network traffic anomaly prediction method based on deep learning. Computer Engineering and Application, 2020, 56 (06): 39-50.

[13] Zhou Feiyan, Jin Linpeng, Dong Jun. Review of convolutional neural network research. Journal of Computer Science, 2017, 40 (06): 1229-1251.

[14] Huang Lin, Wang Diangang, Liu Xiao and so on. LSTM-based network traffic prediction method. Computer Application Research, 2020, 37 (S1): 264-265 + 272.

[15] Guo Niangrong, Gan Haoyu, Chen Lifeng, Lai Yue. A 5G traffic prediction model based on CNN-GRU network. Modern Computer, 2022, 28 (04): 76-81.

[16] Zhu Jun, Hu Wenbo. Review of Bayesian machine learning. Computer Research and Development, 2015, 52 (01): 16-26.

[17] Gelman A, Carlin J, Stern H, et al. Bayesian Data Analysis. Boca Raton: CRC Press, 2013.

[18] Pan Shaowei, Li Hui, Wang Yaling, et al. LSTM neural network crude oil price forecast. Computer technology and development, 2021, 31 (05): 180-185.

[19] Xue Yang, Wang Lin, Wang Shu, et al. An ultra-short-term wind power prediction model combining CNN and GRU network. Renewable energy, 2019, 37 (03): 456-462.