Available online www.jsaer.com

Journal of Scientific and Engineering Research, 2022, 9(8):22-32

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Traffic Flow Prediction for Different Flow Directions at Intersections

Xiaoqing Wang, Pengsheng Zhao, Yuan Fu, Shulong Dong, Yingcui Du, Benxing Liu

School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo, Shandong, China

Abstract As the number of motor vehicles increases year by year, urban traffic problems are becoming more and more prominent, and the traffic efficiency at intersections is low. In this paper, BP neural network and LSTM neural network are used to predict the traffic in different flow directions of the intersection, and the corresponding signal timing optimization is carried out according to the predicted value. Then use SUMO software to simulate the signal timing scheme before and after optimization. Finally, the LSTM neural network prediction result is the best, and the average delay of the optimized intersection is reduced by 19.04%, 11.30%, and 4.83% compared with the actual intersection running state, BP neural network LM algorithm, and BP neural network Bayesian regularization algorithm, respectively. The results show that the traffic flow prediction of different flow directions at the intersection studied in this paper can provide an effective basis for intelligent traffic management and control.

Keywords BP Neural Network; LSTM; Different Flow Directions; SUMO; Traffic Flow Prediction

Introduction

With the accelerating urbanization process and the increasing consumption level of residents, the number of domestic motor vehicles has increased dramatically. By the end of March 2022, the number of motor vehicles in China has reached 402 million, which makes China face serious traffic congestion.

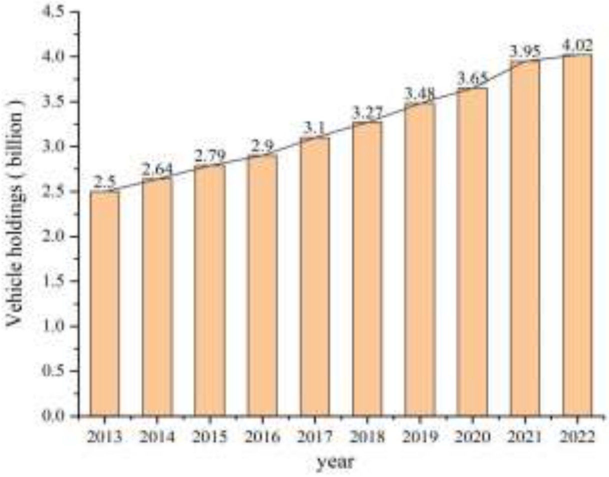

Figure 1: Trend of motor vehicle ownership in China

The solutions include the construction of roads and infrastructure, the optimization of road signal control system, and the construction of diversion to improve traffic supply. By optimizing the road signal control system, the traffic efficiency of ground traffic can be improved by 10 % -20 % [1]. The traditional intersection has 4 inlets and 12 different directions. This paper studies the traffic flow of each direction of the intersection and predicts it, and then optimizes the signal timing of the intersection according to the predicted value.

Compared with other scholars' prediction of intersection traffic, this paper can predict the intersection operation state more accurately, and make the signal timing more in line with the intersection operation characteristics, so that the vehicle passes the intersection with less delay, shorter queue length and less parking times. In summary, it makes the traffic flow more continuous through the intersection and improves the traffic efficiency.

Methods Traffic flow data processing

Intersection traffic flow prediction is inseparable from big data, and the accuracy of prediction results is closely related to big data. Nanjing Xincun intersection is selected as the intersection of the study, the selected data is 2021.11.29-2021.12.05 week data, including time points, the traffic flow of each steering, etc. The selected time window is 5 minutes, Nanjing Xincun intersection of all time window data has 24192 data. There are many problems in the original data, which will affect the prediction accuracy. Therefore, the data should be processed before predicting the intersection flow [2].

The method for processing missing data is as follows:

(1) For smaller data sources, such as only one hour of traffic flow information, it can be discarded. The discarded data are no longer obtained by the same method, but are filtered through multiple layers of data, and the filtered data are plotted by drawing points. Then, the missing data are found through image observation, and the discarded data are supplemented or completely discarded. After completely discarded, only the average value of other traffic flow is used to fill the missing data [3].

(2) For data sources with large amount of data, the activity model is established according to the data, and then the partial missing data is predicted according to the activity model to obtain more complete traffic flow information [4].

(3) The traffic flow data is time-segmented after processing in different periods, and the information collected at the intersection is spatial. At this time, the average method, the expected value method or the incremental method can be used to supplement the missing data.

The method for processing redundant data is as follows:

(1) Completely repeated data. First, one data is retained and then its data are deleted.

(2) Multiple similar data, take the average or median value and leave a short record.

(3) Grey Wolf Simhash redundant data monitoring algorithm can be used when dealing with large amount of data and prone to produce redundant data [5].

The method for processing abnormal values is as follows:

(1) Short-time outliers can be deleted directly.

(2) There are many but discontinuous abnormal values, and the average value of the before and after periods can be used to replace the abnormal value.

(3) There are many abnormal values and continuous, which can be replaced by the average value of the traffic flow data before and after several days.

Traffic flow prediction based on neural network BP neural network

Back propagation (BP) neural network is a concept proposed by scientists led by Rumelhart and McClelland in 1986. It is a multi-layer feedforward neural network trained according to the error back propagation algorithm, and is one of the most widely used neural network models.

The artificial neural network does not need to determine the mathematical equation of the mapping relationship between input and output in advance. It only learns some rules through its own training, and obtains the results closest to the expected output value when the input value is given. As an intelligent information processing system, the core of artificial neural network to achieve its function is algorithm. BP neural network is a kind of multi-layer feed-forward network trained by error back propagation (referred to as error back propagation). Its basic idea is gradient descent method, and gradient search technology is used to minimize the error mean square deviation of the actual output value and the expected output value of the network.

The basic BP algorithm includes two processes: signal forward propagation and error back propagation. That is to say, the calculation error output is carried out in the direction from input to output, and the adjustment of weights and thresholds is carried out in the direction from output to input. In forward propagation, the input signal acts on the output node through the hidden layer, and the output signal is generated by nonlinear transformation. If the actual output is not consistent with the expected output, the error is transferred to the reverse propagation process. Error back-propagation is to transmit the output error layer by layer from the hidden layer to the input layer, and the error is distributed to all units of each layer. The error signal obtained from each layer is used as the basis for adjusting the weight of each unit. By adjusting the connection strength between the input node and the hidden layer node and the connection strength and threshold between the hidden layer node and the output node, the error decreases along the gradient direction. After repeated learning and training, the network parameters (weights and thresholds) corresponding to the minimum error are determined, and the training is stopped. At this time, the trained neural network can self-process the input information of similar samples with the smallest output error through nonlinear transformation.

BP neural network consists of input layer, hidden layer and output layer. The specific structure is shown in Figure 2:

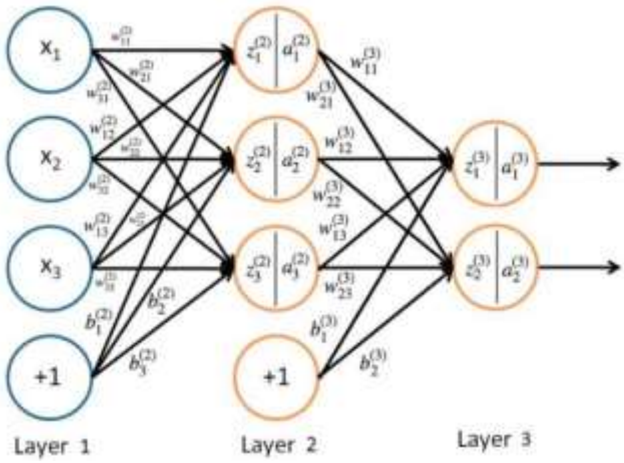

Figure 2: BP neural network structure

X1, X2 and X3 are the input of the first neuron, w ij is the weight of the input layer to the hidden layer, and bij is the bias. Weights and biases are randomly initialized at the beginning, with uncertainty, and back propagation is designed to get model parameters close to optimal values. Reverse propagation is based on gradient, similar to finding the fastest rising direction around climbing. The neural network output after reaching the optimal solution.

The flowchart of the algorithm implementation is shown in Figure 3:

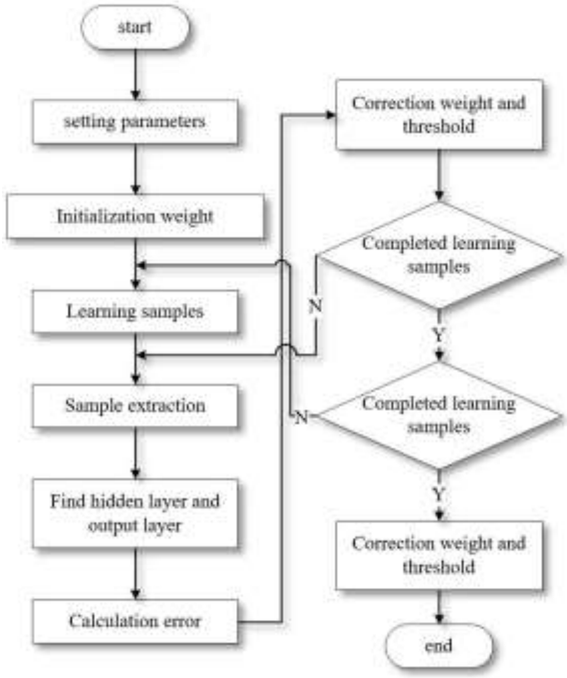

Figure 3: Algorithm flowchart

Levenberg-Marquardt algorithm of BP neural network

Completed in MATLAB, first of all, the processed data as input data into the BP neural network LM algorithm, after constant parameter training to get the best BP neural network. Output training set, validation set, test set correlation coefficient diagram.

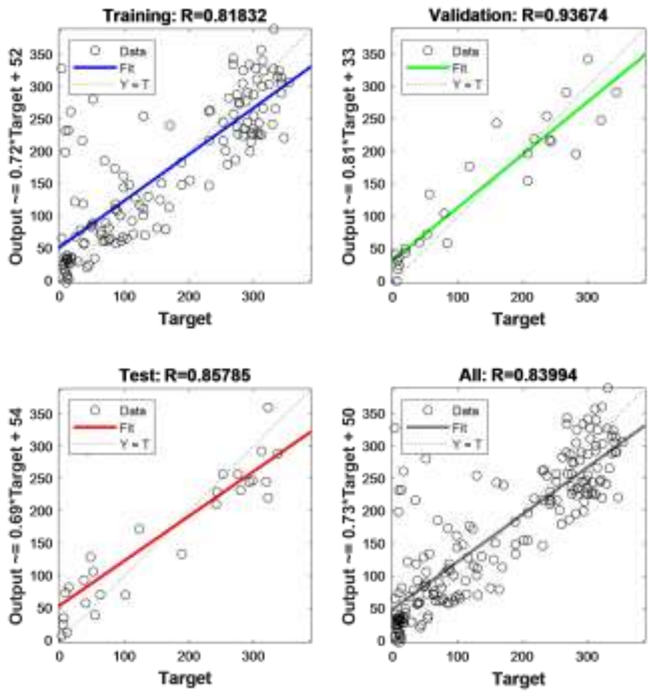

Figure 4: Correlation coefficient of LM algorithm training set, validation set and test set

The algorithm is fast but has low accuracy of prediction results. Next, Bayesian regularization in BP neural network is used to improve it.

Bayesian-regularization algorithm of BP neural network

Compared with LM algorithm, Bayesian regularization algorithm can improve the generalization ability of BP neural network, and can fit a sinusoidal sample data with white noise. The accuracy rate is higher than that of LM algorithm, and it can prevent the over-fitting phenomenon in processing large amounts of data. However, these advantages are at the expense of time.

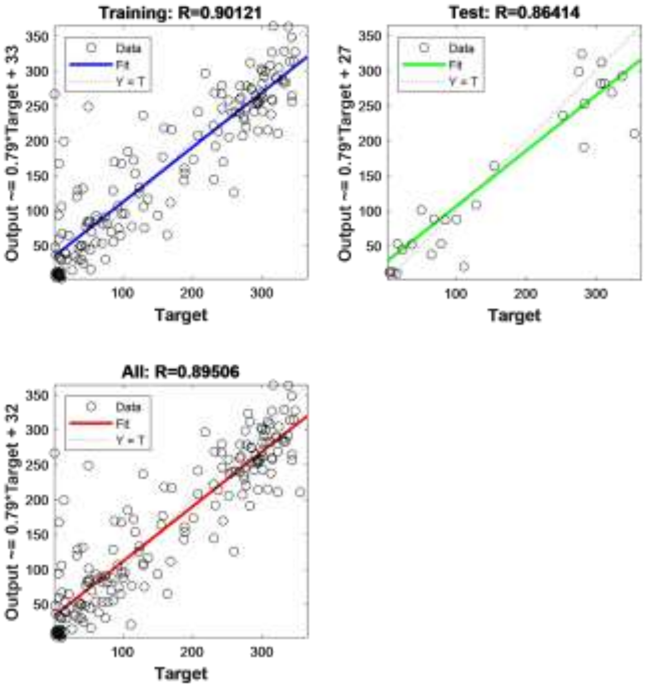

Figure 5: Correlation coefficient of Bayesian regularization algorithm training set, validation set and test set

Lstm Neural Network

Long Short-Term Memory (LSTM) is a kind of time recurrent neural network, which is specially designed to solve the long-term dependence problem of general RNN (recurrent neural network). All RNNs have a chain form of repetitive neural network modules [6]. In the standard RNN, this repeated structural module has only one very simple structure, while the LSTM model has multiple structures, as shown in Figs. 6 and 7.

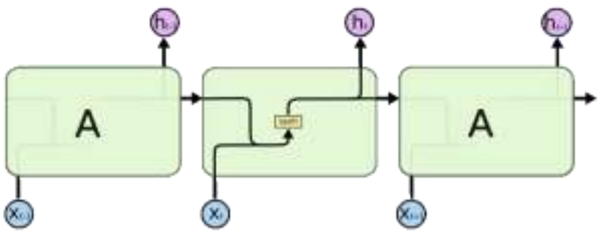

Figure 6: Repeated Chain Structure in Standard RNN

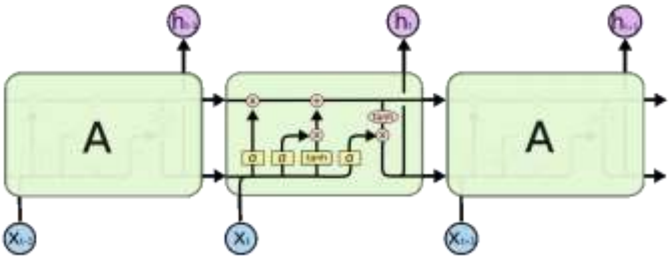

Figure 7: Repeated Chain Structure in Long Short Term Memory Network (LSTM)

As a kind of deep learning, LSTM [7] not only considers the nonlinear mapping relationship between multivariate variables, but also solves the long-term dependence problem of time series that cannot be solved by traditional artificial neural network [8].

It can also solve the problem of gradient disappearance. For RNN, for each time point, the information in the memory will be covered. However, unlike in LSTM, it multiplyes the value in the original memory by a value added with input into the unit, and its memory and input are added. Therefore, unlike RNN, it will be covered at each time point. As long as the information at the previous moment is formatted, the influence will disappear. However, the influence in LSTM will always exist, unless the forgetting door cleans the information in the memory.

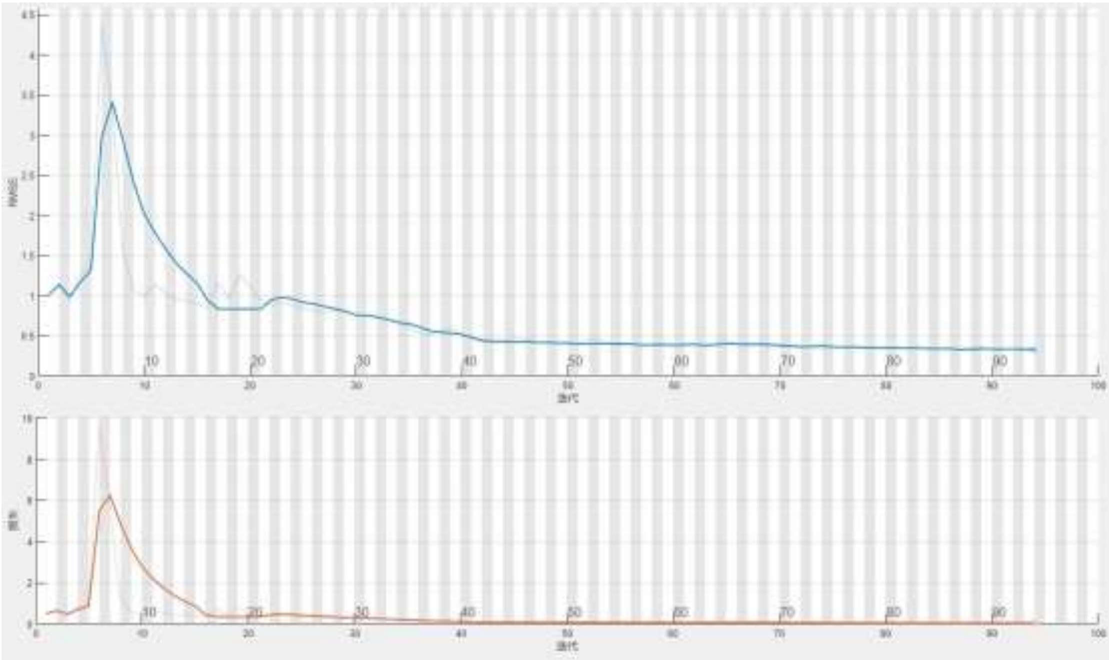

Figure 8: Iterative process of LSTM algorithm

In the implementation of LSTM algorithm, through the iteration in the process, we can observe the variation curves of root mean square error and loss function. In 60 iterations, the root mean square error and loss function curve change little, so the 100 iterations can meet the accuracy requirements.

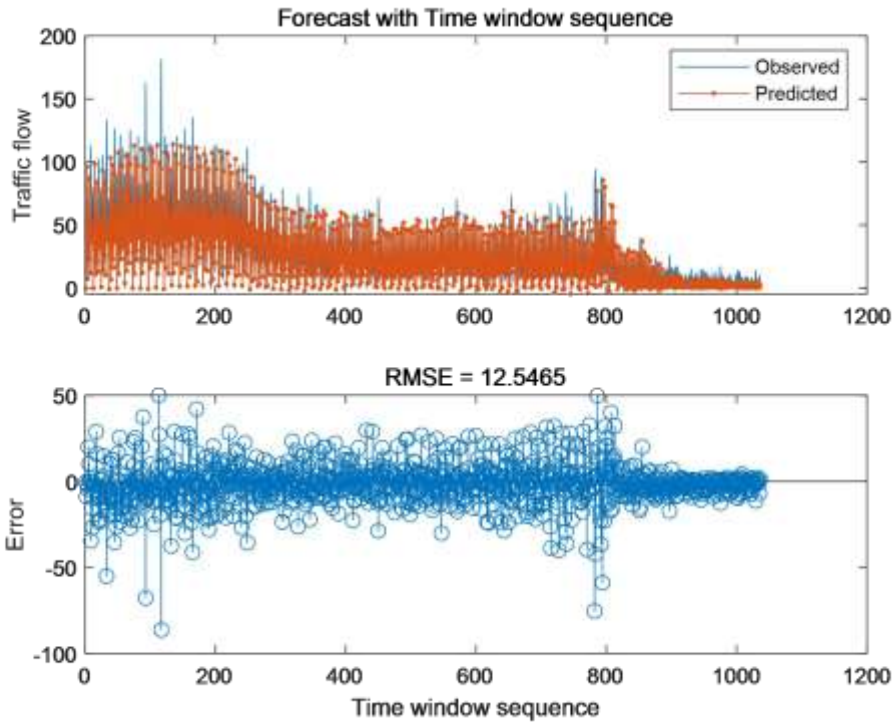

Figure 9: Prediction of traffic flow and root mean square error

The predicted different traffic flow data are as follows:

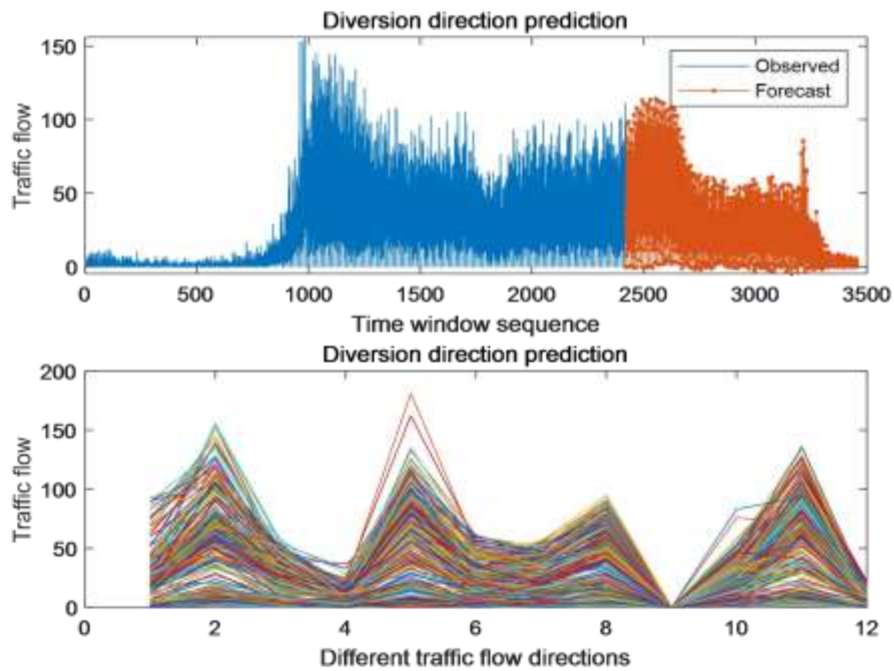

Figure 10: Traffic flow prediction results of different flow directions

Simulation Verification

Traffic simulation refers to the use of simulation technology to study traffic behavior. It is a technology to track and describe the changes of traffic movement with time and space. It contains random characteristics, which can be microscopic or macroscopic, and involves a mathematical model describing the real-time movement of the transportation system during a certain period.

The role of traffic simulation is to reproduce or grasp the traffic operation status of the existing system or the future system in advance, so as to explain, analyze and identify the crux of the problem for complex traffic phenomena, and finally optimize the traffic system studied. Compared with the traditional traffic analysis technology, the advantages of traffic simulation technology are: more flexible model mechanism, more accurate model description, more open traffic analysis, and more powerful dynamic traffic state description of road network.

Intersection Construction

SUMO [9] is an open source, microscopic, multi-mode traffic simulation software with powerful functions of generating intersections, vehicles, pedestrians and so on. The software has its own Python interface, which can be used for secondary development of Python, and the overall function is flexible and powerful. Therefore, this paper uses SUMO software to simulate the intersection [10].

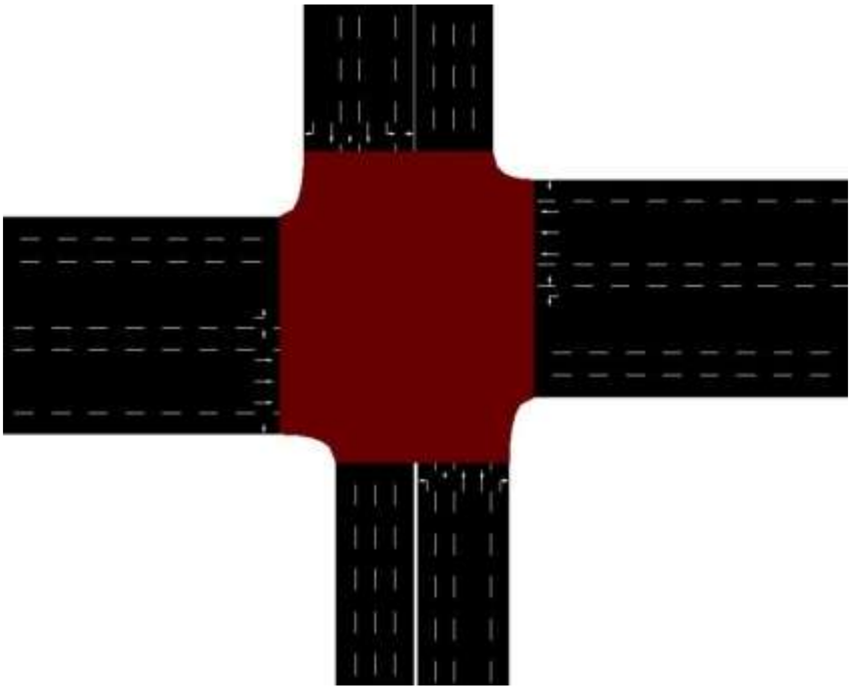

Figure 11: Intersection Construction of SUMO Simulation Platform

Edit intersection vehicle input information:

```
<!-- 1 -->
<flow id="f_0" begin="0.00" from="xijinkou" to="beichukou" end="20700" number="86"/>
<flow id="f_1" begin="0.00" from="xijinkou" to="dongchukou" end="20700" number="310"/>
<flow id="f_2" begin="0.00" from="xijinkou" to="nanchukou" end="20700" number="52"/>
<flow id="f_10" begin="0.00" from="beijinkou" to="nanchukou" end="20700" number="392"/>
<flow id="f_11" begin="0.00" from="beijinkou" to="xichukou" end="20700" number="100"/>
<flow id="f_9" begin="0.00" from="beijinkou" to="dongchukou" end="20700" number="59"/>
<flow id="f_3" begin="0.00" from="dongjinkou" to="nanchukou" end="20700" number="54"/>
<flow id="f_4" begin="0.00" from="dongjinkou" to="xichukou" end="20700" number="398"/>
<flow id="f_5" begin="0.00" from="dongjinkou" to="beichukou" end="20700" number="52"/>
<flow id="f_6" begin="0.00" from="nanjinkou" to="xichukou" end="20700" number="74"/>
<flow id="f_7" begin="0.00" from="nanjinkou" to="beichukou" end="20700" number="166"/>
<flow id="f_8" begin="0.00" from="nanjinkou" to="dongchukou" end="20700" number="64"/>
<!-- 2 -->
<flow id="f_12" begin="20700" from="xijinkou" to="beichukou" end="22500" number="53"/>
<flow id="f_13" begin="20700" from="xijinkou" to="dongchukou" end="22500" number="115"/>
<flow id="f_14" begin="20700" from="xijinkou" to="nanchukou" end="22500" number="28"/>
<flow id="f_15" begin="20700" from="beijinkou" to="nanchukou" end="22500" number="131"/>
<flow id="f_16" begin="20700" from="beijinkou" to="xichukou" end="22500" number="24"/>
<flow id="f_17" begin="20700" from="beijinkou" to="dongchukou" end="22500" number="30"/>
<flow id="f_18" begin="20700" from="dongjinkou" to="nanchukou" end="22500" number="7"/>
<flow id="f_19" begin="20700" from="dongjinkou" to="xichukou" end="22500" number="94"/>
<flow id="f_20" begin="20700" from="dongjinkou" to="beichukou" end="22500" number="21"/>
<flow id="f_21" begin="20700" from="nanjinkou" to="xichukou" end="22500" number="52"/>
<flow id="f_22" begin="20700" from="nanjinkou" to="beichukou" end="22500" number="88"/>
<flow id="f_23" begin="20700" from="nanjinkou" to="dongchukou" end="22500" number="47"/>
```

Figure 12: Edit traffic input of intersection

Signal Timing Optimization

After building the road network, the intersection signal timing and flow data of different directions are imported into SUMO software, and the SUMO config file is used to analyze the intersection delay. For the predicted values obtained by the above method, the real-time optimized signal timing scheme is used to meet the intersection change, so as to reduce the delay. In order to make the signal timing scheme more in line with the variation characteristics of traffic flow, the whole day is divided into different periods, and different signal timing schemes are used in different periods [11,12].

After signal optimization, add a full-day signal timing scheme using the add-in file in SUMO :

```
<additional-files value="PAM-SOM.add.xml"/>
```

Figure 13: Adding all-day signal timing scheme

Edit signal timing in each period :

```
<tlLogic id="J0" type="static" programID="a1" offset="0">
    <phase duration="15" state="GrrrrrGGGGrrGrrrrGGGGrr"/>
    <phase duration="3" state="GrrrrrGYYYrrGrrrrGYYYrr"/>
    <phase duration="7" state="GrrrrGrrrGGGrrrrGrrrGG"/>
    <phase duration="3" state="GrrrrGrrrYYGrrrrGrrrYY"/>
    <phase duration="13" state="GGGGrrGrrrrrGGGGrGrrrrr"/>
    <phase duration="3" state="GYYYrrGrrrrrGYYYrGrrrrr"/>
    <phase duration="17" state="GrrrGGGrrrrrGrrrGGrrrrr"/>
    <phase duration="3" state="GrrrYYGrrrrrGrrrYGrrrrr"/>
</tlLogic>
```

Figure 14: Signal timing design Editing signal timing switching logic in different periods:

```xml
<WAUT startProg="0" refTime="0" id="YUANSHISHIDUANHUAFEN">
    <wautSwitch to="a1" time="20700"></wautSwitch>
    <wautSwitch to="b2" time="22500"></wautSwitch>
    <wautSwitch to="d4" time="27900"></wautSwitch>
    <wautSwitch to="e5" time="31500"></wautSwitch>
    <wautSwitch to="c3" time="34200"></wautSwitch>
    <wautSwitch to="f6" time="36000"></wautSwitch>
    <wautSwitch to="c3" time="37800"></wautSwitch>
    <wautSwitch to="b2" time="42300"></wautSwitch>
    <wautSwitch to="a1" time="45000"></wautSwitch>
    <wautSwitch to="c3" time="47700"></wautSwitch>
    <wautSwitch to="f6" time="49500"></wautSwitch>
    <wautSwitch to="b2" time="53100"></wautSwitch>
    <wautSwitch to="g7" time="57600"></wautSwitch>
    <wautSwitch to="h8" time="61200"></wautSwitch>
    <wautSwitch to="d4" time="69300"></wautSwitch>
    <wautSwitch to="c3" time="72000"></wautSwitch>
    <wautSwitch to="h8" time="74700"></wautSwitch>
    <wautSwitch to="c3" time="78300"></wautSwitch>
    <wautSwitch to="h8" time="81000"></wautSwitch>
    <wautSwitch to="g7" time="83700"></wautSwitch>
    <wautSwitch to="a1" time="86700"></wautSwitch>
</WAUT>
```


Figure 15: Switching Logic Design of Signal Timing Scheme

After completing the signal timing and flow input, run all files and output the simulation results.

Results

According to the predicted value, the signal timing optimization is compared with the actual survey data simulation, and the evaluation index is the delay of the vehicle passing through the intersection. In order to reduce the impact of contingency on simulation experiments and improve the accuracy of simulation experiments, this paper will take three experiments to take the average method.

Table 1: Output of simulation results

Experimental Scheme		Delay		
		First experiment	Second experiment	Third experiment
Actual intersection operation state		77.669	68.785	69.663
BP neural network	LM algorithm	68.056	67.732	69.901
	Bayesian regularization	70.313	62.371	59.022
LSTM neural network		56.196	63.088	55.674

The intersection delay of the signal timing scheme obtained by the traffic flow predicted by the LSTM neural network is the smallest, which is 19.04 % less than that under the actual intersection operation state, 11.30 % less than that obtained by the BP neural network LM algorithm, and 4.83 % less than that obtained by the BP neural network Bayesian regularization algorithm. The LSTM proposed in this paper predicts different flow directions at intersections, which is of great significance for improving the operation state of intersections.

Conclusion

In this paper, the traffic flow at the intersection is predicted. Considering that the prediction of the total flow at the intersection is not a good solution to its congestion problem, the traffic flow at different directions is predicted, and it is predicted by different network models. The prediction performance of LSTM model is better than that of BP neural network model, which can provide new ideas and basis for intelligent traffic management and control. In the future, we will continue to learn this knowledge, optimize the model, and realize more intelligent and more accurate traffic flow prediction methods.

References

[1]. Ren Bohan, Ding Xuemei. Intersection signal control time division method based on traffic big data (2021). *Strait technology and industry*, 34(08): 51–53.

[2]. Zhang R, Sun F, Song Z. Short-Term Traffic Flow Forecasting Model Based on GA-TCN (2021). *Journal of Advanced Transportation*, *Hindawi*, 2021: e1338607.

[3]. Wu, Huang, You, et al. Progress in taxi trajectory data mining (2019). *Journal of Surveying and Mapping*., 48(11): 1341–1356.

[4]. Wang, Li, and Mao. The data cleaning method of electric police and the application of data in signal control (2019). Papers on quality transportation and collaborative governance *Annual Conference of Urban Traffic Planning in China. Chengdu, Sichuan, China: Academic Committee of Urban Traffic Planning, China Society of Urban Planning*, 2019: 3526–3535.

[5]. Ruan Jiakun, Cai Yanguang, Cai Hao, etc.Simhash redundant data detection algorithm based on grey wolf algorithm (2020). *Journal of Dongguan University of Technology*, 27(05): 38–43.

[6]. Application of LSTM network in short-term power load forecasting under the framework of deep learning (2017). *Power information and communication technology*, 15(05): 8–11.

[7]. Graves A. Long Short-Term Memory (2012). Supervised Sequence Labelling with Recurrent Neural Networks, *Springer, Berlin, Heidelberg*: 37–45.

[8]. Shen Miaosheng, Gao Gengjun. Short-term highway traffic volume prediction based on LSTM. *Control engineering*.: 1–8.

[9]. J. Rummel, P. Wagner, and E. Wiessner. Microscopic Traffic Simulation Using SUMO. 2018.

[10]. Cheng Yuyang, Zhou Bingtao, Shi Chengxi. Traffic signal timing optimization based on long-term and short-term memory artificial neural network and SUMO simulation (2021). *Science and technology innovation*, (26): 67–70.

[11]. Chen Dong. Application and Research of Intelligent Algorithm in Intelligent Traffic Signal Control (2019). *Shanghai University of Applied Technology*.

[12]. Li Yaman. Single intersection signal timing optimization based on multi - time division (2021). *Chang 'an University*.