

Scientific Research and Reviews (ISSN:2638-3500)



Predicting the Time Required to Pass Congested Road Based on Neural Network Algorithm

Zhong Zheng^{1*}, Yu Cao², Hairui Zhang², Tianlong Wang³, Yunxiao Wu¹

¹College of Electrical Engineering & New Energy, China Three Gorges University, Yichang, 443002, China. ²College of Science, China Three Gorges University, Yichang, 443002, China. ³College of Civil Engineering & Architecture, China Three Gorges University, Yichang, 443002, China.

ABSTRACT

In order to predict the duration of traffic congestion, this paper established a traffic congestion evaluation model based on cumulative ratio Logistic regression and a traffic congestion time prediction model based on BP neural network. Combining Pearson test, numerical combination, standard deviation method and other methods to solve the problem. Based on the measured data of Jinshui Road in Zhengzhou, the average error is 0.019m/s and the prediction error rate is 0.15%, both within a reasonable range. The model can improve the accuracy of congestion time prediction and provide some help to real life.

Keywords: Cumulative ratio Logistic regression; BP neural network; Pearson test; Standard deviation method

*Correspondence to Author:

Zhong Zheng

College of Electrical Engineering & New Energy, China Three Gorges University, Yichang, 443002, China.

How to cite this article:

Zhong Zheng, Yu Cao, Hairui Zhang, Tianlong Wang, Yunxiao Wu. Predicting the Time Required to Pass Congested Road Based on Neural Network Algorithm. Scientific Research and Reviews, 2020; 13:112

 **eSciPub**
eSciPub LLC, Houston, TX USA.
Website: <http://escipub.com/>

1. Introduction

Nowadays, people's life is getting better and better, private cars have become the standard of every family, which makes it more convenient for people to travel. But at the same time, as the number of vehicles increases, the roads become more and more crowded. It also raises concerns about the amount of time spent on the road. In today's navigation software, real-time GPS data is used to determine current road conditions and give the owner an estimated time to jam. However, when the road is extremely crowded and the speed is extremely slow, the estimation of the speed is very inaccurate. The result is poor accuracy in estimating the duration of traffic jams. The actual time required is sometimes several to ten times less than predicted. This can be particularly troublesome for people with particularly high time accuracy requirements, so it makes practical sense to build more accurate predictive models. In this paper, a model is built to predict the time of traffic congestion in order to improve the accuracy of the prediction.

2. Model assumptions

Hypothesis 1: Think that the weather has very little effect on traffic status;

Hypothesis 2: Traffic congestion is not caused by tailgating, highway maintenance and other emergencies;

Hypothesis 3: GPS to know the distance of the vehicle from the congestion center.

3. Problem analysis

The research of traffic jam prediction model is an information science of pattern recognition problem. According to practical problems, the real-time GPS data is combined with the previous driving data to consider as many influencing factors as possible, and the BP neural network prediction model is established to enable navigation to accurately predict the duration of traffic congestion. In addition, in this paper, traffic congestion is regarded as a prediction factor of BP neural network, so a Logistic regression evaluation model is established to discriminate traffic congestion. The idea diagram to solve the problem is as follows:

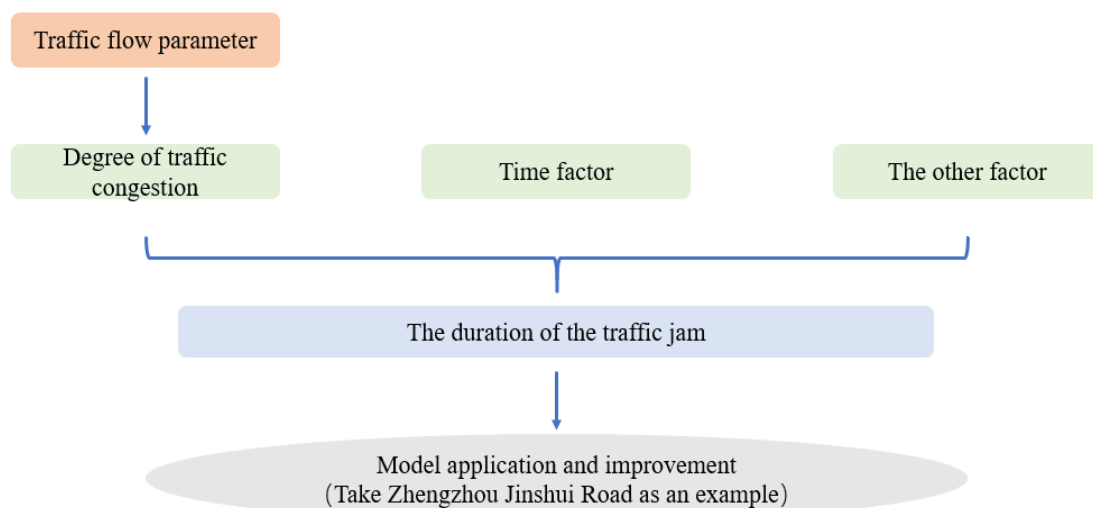


Fig.1: Idea diagram

Firstly, a model is established based on the traffic flow congestion degree judgment parameter for discriminating, and then a model is established to predict the congestion time by combining the time factor, other factors and congestion degree. Finally, the actual road data of Jinshui

Road in Zhengzhou is taken as an example to apply and test the established model, and the improvement is proposed.

4. Model building

4.1 Traffic congestion evaluation model based on cumulative ratio logistic regression

4.1.1 Model preparation

A. Explain variables and how to divide them

In this paper, multiple traffic flow parameters such as "average driving distance", "average speed" and "average traffic flow" are selected to evaluate the degree of traffic congestion. In combination with the established evaluation model, these variables need to be quantified and assigned.

a. Traffic flow parameters: traffic flow parameters are the parameters that most intuitively express the traffic state 错误!未定义书签。. By referring to relevant literatures and combining the actual significance of data, this paper selects three

parameters, namely "average driving distance (x_1)", "average speed (x_2)" and "average driving flow (x_3)", as the traffic flow parameters. Where, the average traffic flow is the number of vehicles passing on a certain highway in a unit time. Therefore, the more congested the traffic is, the smaller the average traffic flow will be.

b. Traffic congestion degree: based on the actual situation of most urban sections, this paper divides the traffic congestion degree into three different levels, namely "unimpeded", "congestion" and "congestion", and quantifies this parameter with "traffic density Y ", $Y=1$, "unblocked", $Y=2$, "congestion", $Y=3$, "clog".

Table 1: Table of influencing factors

Influencing factors	Quantitative assignment
Mean distance between vehicles (x_1)	$>120\text{m}=1$ $90 \sim 120\text{m}=3$ $60 \sim 90\text{m}=5$ $30 \sim 60\text{m}=7$ $5 \sim 30\text{m}=9$
Average speed (x_2)	$1 \sim 20\text{km/h}=9$ $20 \sim 40\text{km/h}=7$ $40 \sim 60\text{km/h}=5$ $60 \sim 80\text{km/h}=3$ $>80\text{km/h}=1$
Average traffic flow (x_3)	$<800\text{veh/h}=9$ $800 \sim 2000\text{veh/h}=7$ $2000 \sim 3000\text{veh/h}=5$ $3000 \sim 4000\text{veh/h}=3$ $>4000\text{veh/h}=1$
Density of driving (Y)	$0 \sim 30\text{veh/km}=1$ $30 \sim 60\text{veh/km}=2$ $>60\text{veh/km}=3$

B. The theory basis:

Cumulative Ratio Logistic Regression: A Statistical Analysis Method for Ordered Multi-Classification Reaction Variables 18, This method can be used to classify things indirectly, by collecting statistical parameters of each state, and then calculate the probability in different categories, select the maximum probability category as the

final category of things.

4.1.2 Modeling

Establish an analytical model based on the cumulative ratio Logistic regression, with the following steps:

Step1: Quantitative assignment of data: In the process of assignment, because the cumulative ratio Logistic is regression as a statistical

method of the reaction variable slot tinted in order and multiple classification, it is necessary to divide the grade weight according to the actual situation.

Step2: Define the cumulative ratio Logistic regression model:

Set the driving density of the result variable (Y)

$$\text{logit } p_i = \text{logit } p(y > i) = -\beta_i + \sum_{j=1}^3 \alpha_j x_j \quad i=1,2$$

In the formula:

i is the rank of ordered dependent variable;

$p(y > i)$ represents the probability that the rank is greater than i ;

β_i is the parameter corresponding to the rank of ordered dependent variable i ;

as the ordered dependent variable y of three grades, respectively, $Y=1$, "unimpeded", $Y=2$, "congestion";, $Y=3$, "jam"; Three variables, average driving distance (x_1), average speed (x_2) and average driving flow (x_3) are taken as independent variables, namely, $x^T = (x_1, x_2, x_3)$, get that:

α_j is the parameter of the corresponding

independent variable x_j ;

j is the serial number of independent variables, which can be taken as 1,2,3 in this paper;

The above equation can also be equivalent to:

$$p(y \leq i) = \frac{1}{1 + \exp(-\beta_i + \sum_{j=1}^3 \alpha_j x_j)} \quad i=1,2$$

In the formula: $p(y \leq i)$ represents the probability that the rating is less than or equal to i ; i is the rank of ordered dependent variable;

Step3: Parameter estimation: after data processing with MATLAB, calculate in SPSS, and get the corresponding sum β_i and α_j ;

Step4: Test the goodness of fit of the model:

After the establishment of the model, the validity and accuracy of the model are tested. The commonly used test methods are test $\text{pearson}\chi^2$ and deviation statistic test. This paper uses test $\text{pearson}\chi^2$:

a. With the probability obtained, the total number of samples corresponding to each category is calculated;

b. Actual statistics of the number of samples corresponding to each category;

c. According to the category of the population sample (there are 3 categories here), determine the test rejection domain:

$$\{\chi^2 \geq \chi_{1-\alpha}^2(5)\}$$

Make that $\alpha=0.95$, if so, refer to the table to get that: $\chi_{0.95}^2(5)=11.07$

d. Select 3 categories to calculate χ^2 respectively: $\chi^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}$

In that, n_i is the observed (actual) frequency, np_i is the theoretical frequency, and n is the total number of samples. The meaning of this equation ($n_i - np_i$) is the difference between the observed (actual) frequency and the theoretical frequency. It can also be obtained directly by SPSS software.

4.2 Based on BP neural network prediction model

4.2.1 Model preparation

A. Explanatory variables and their partitioning methods:

In the analysis of traffic congestion state, the influencing factors of traffic congestion are considered from the aspects of traffic congestion state, external time factors, special circumstances and so on, and then the duration of traffic congestion is predicted by combining BP neural network. In addition to the factors listed in 4.1.1(A), there are the following factors:

a. Time factor: traffic conditions are different in different time periods, especially in the peak period of commuting. Therefore, for the quantification of time factor T , 9 is used to represent the corresponding time from 7:00 to 9:00

in the morning and 17:00 to 20:00 in the evening. Daytime travel time from 9:00 to 17:00 is expressed as 7; It is expressed as 3 from 6:00 to 7:00 before morning peak and from 20:00 to 22:00 after evening peak. The rest of the time period is denoted by 1.

b. Other factors: The influence of special events E , mainly involving municipal engineering, road maintenance, traffic accidents, sports events, etc. When there is no special event on the road, it is denoted by 1. When the above

special events exist, they can be expressed by 3, 5, 7 and 9 successively according to their influence on the traffic state. This variable is not considered in detail in the model of this paper, but is set as 1.

c. Description of the duration of traffic congestion: Described by the average speed (x_2). On a certain road section, when the length of the road section is certain, the lower the average speed is, the more serious the congestion is.

Table 2: table of influencing factors

Impact factors	Quantitative assignment
Time factor T	7:00 ~ 9:00 and 17:00 ~ 20:00=4 9:00 ~ 17:00=3 6:00 ~ 7:00 and 20:00 ~ 22:00=2 Other time=1
Special events E	Including solutions in this paper=1

B. Theory basis:

1. BP neural network: It is a multi-layer feedforward neural network trained according to the error backward propagation algorithm. Generally, 70% of the sample is used to train the model, and then 30% of the sample is used to test the accuracy of the model, so as to improve the model.

2. Survival analysis: The general survival analysis mainly refers to the duration of a phenomenon of concern, while the duration of road congestion refers to the duration from the beginning of traffic congestion and until the end of the congestion, so it can be combined with the relevant knowledge of survival analysis to analyze the problem. The schematic diagram is as follows:

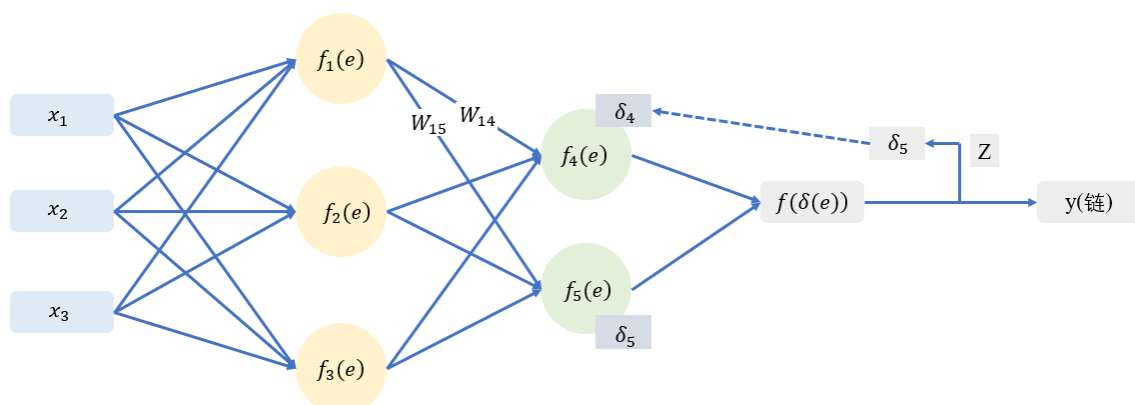


Fig.2: BP Neural Network Prediction Schema

4.2.2 Modeling

Time prediction model based on BP neural network in this paper, in order to train operation density Y , average driving distance (x_1),

average traffic flow (x_3), time factor T , special events E and average speed (x_2) as the input values, at an average speed of the output value and 70% of the sample as the training

sample of neural network training, with 30% of the samples for training result accuracy test. The time prediction model based on BP neural network is established, with the following steps:

Step1: Selection of sample values: As far as possible to select a large number of sample values, and reasonable allocation of the use of sample values, with 70% of the sample for training the training of the neural network, with 30% of the sample for the accuracy test of the training results;

Step2: The determination of input and output values: according to this article needs to solve problems and data collection, based on the traffic density Y , average driving distance

$$x'_{mn} = \frac{x_{mn} - \bar{x}_n}{s_n} \quad \bar{x}_n = \frac{1}{a} \sum_{m=1}^a x_{mn} \quad s_n = \left[\frac{1}{a} \sum_{m=1}^a (x_{mn} - \bar{x}_n)^2 \right]^{1/2}$$

In the formula:

x'_{mn} represents the standardized value of the number in row m and column n ;

x_{mn} represents the original value of the number in row m and column n ;

\bar{x}_n represents the average of column n ;

a represents the total number of rows of all the original data;

Step4: Establish BP neural network and train and precision test network;

Step5: Applying the model established;

5. Model application -- a case study of Jinshui Road, Zhengzhou, China

(x_1) , average traffic flow (x_3) , time factor T , special events E , and average speed (x_2) as the input values, at an average speed (x_2) of the output value, at an average speed to describe traffic congestion of the link length to determine the duration;

Step3: Data preprocessing: for the parameters of qualitative categories, the quantitative specification is eliminated directly in the process of quantization, and for the parameters of quantitative classes need to eliminate the volume outline before the application and calculation, the data is pre-processed, this paper uses the standard deviation method to deal with the data without volume outline:

5.1 Data collection and analysis

This paper looks up traffic measurement data for some time period in 2008 and 2009 (up to 2011) on the traffic measurement video and data sharing platform for traffic measurement video and data sharing in four cities in Beijing, Shanghai, Xi'an and Zhengzhou. This article takes Jinshui Road in Zhengzhou City as a data sample, including this section of the road on August 25, 2008 8:35 a.m. to 4:50 p.m. car spacing, speed, flow and density, combined with these data, first sorted the workshop, and analyzed the data by *Origin* data analysis software:

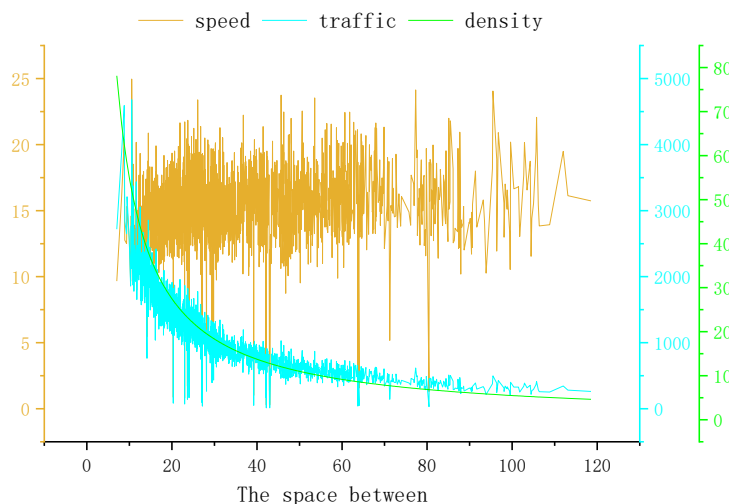


Fig.3: Relationship between car spacing and vehicle speed, flow and density

There are three Y axes in the figure, vehicle speed, flow rate and density. As can be seen from the figure, the traffic flow and density will decrease with the increase in the distance

between the vehicles, while the vehicle speed is generally stable, so the vehicle speed is analyzed separately as shown in the figure:

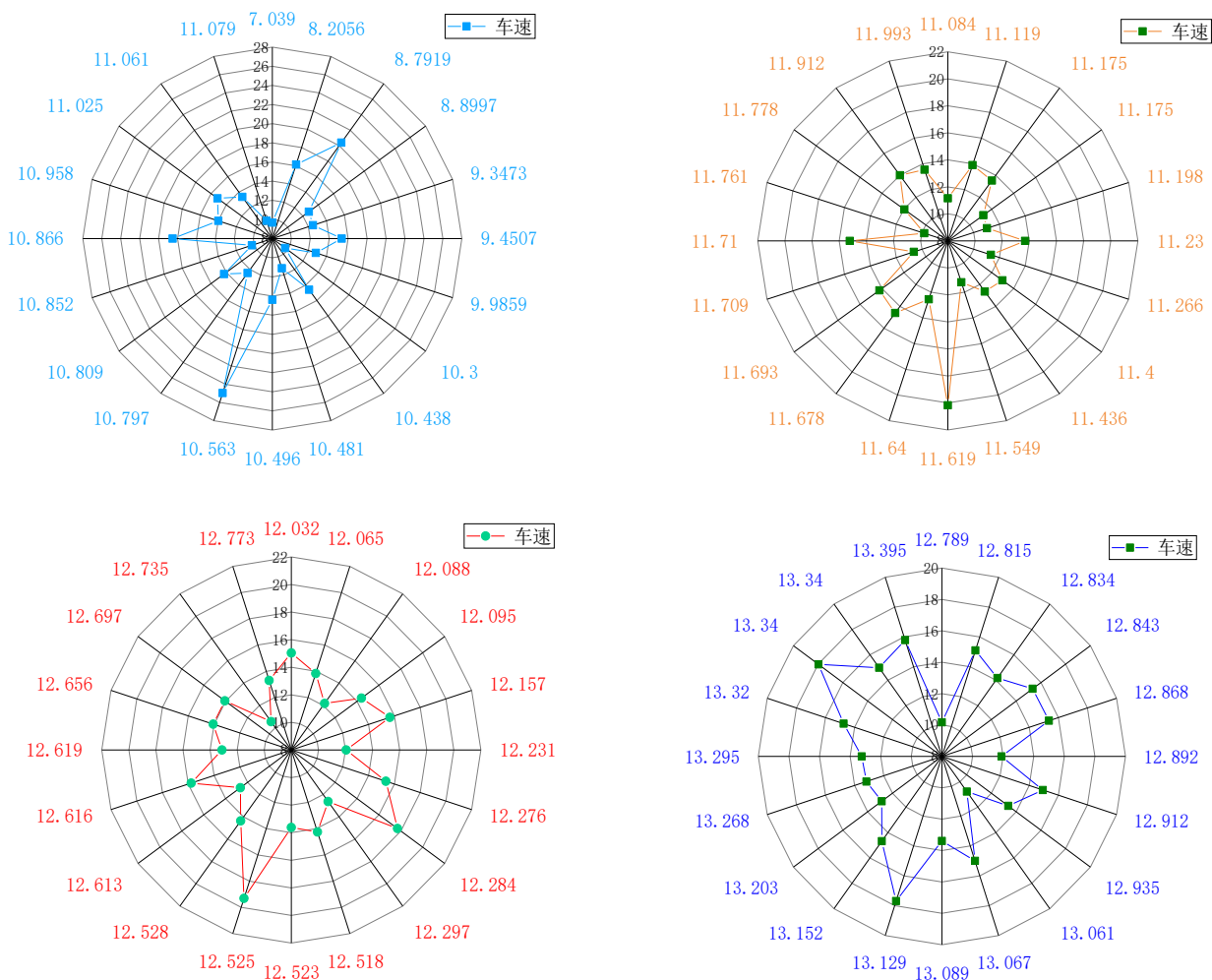


Fig.4: Analysis of the relationship between car spacing and vehicle speed

Because there is more data, this article intercepts some of the data as an analytical sample for analysis graphics. The data shown in the peripheral in the graph is that the car spacing gradually increases clockwise, and in the comparison of the four figures, it can be seen that the maximum speed of the vehicle has experienced a change of 28Km/h→22Km/h→20 Km/h as the car spacing increases.

5.2 Application of model

Based on the data of Jinshui Road in Zhengzhou City on August 26, 2008, the analysis model based on the cumulative ratio Logistic

regression and the time prediction model based on BP neural network and survival analysis function are applied, and the application results are analyzed and tested.

5.2.1 Judgment of traffic congestion:

The traffic condition of Jinshui Road of Zhengzhou on August 26, 2008 was analyzed by using the established Logistic regression analysis model based on cumulative ratio.

A. Model independent variable correlation analysis:

Table3: Parameter estimation and model calibration results

							95% Confidence	
							interval	
							Lower	Ceiling
thres	[density = 1.00]	-6.949	0.860	65.319	1	0.000	-8.635	-5.264
hold	[density = 2.00]	-4.867	0.757	41.389	1	0.000	-6.350	-3.385
location	flow	-1.761	0.230	58.633	1	0.000	-2.212	-1.311

At the significance level of 0.05, the correlation was significant. From the above table, we can get the odds ratio model of "unblocked", "congested" and "blocked", and the odds ratio model

of "unblocked, congested" and "blocked". The two models are as follows:

1.The advantage ratio model of "unblocked" versus "congested" and "blocked":

$$\ln\left(\frac{p_1}{p_2 + p_3}\right) = -6.949 - 1.761x_3$$

2.The dominance ratio model of "unblocked" and "congestion" versus " blocked ":

$$\ln\left(\frac{p_1 + p_2}{p_3}\right) = -4.867 - 1.761x_3$$

Then the calculation expression of p_1 、 p_2 、 p_3 , and can be obtained:

$$\begin{cases} p_1 = \frac{\exp(-6.949 - 1.761x_3)}{1 + \exp(-6.949 - 1.761x_3)} \\ p_2 = \frac{\exp(-4.867 - 1.761x_3)}{1 + \exp(-4.867 - 1.761x_3)} - \frac{\exp(-6.949 - 1.761x_3)}{1 + \exp(-6.949 - 1.761x_3)} \\ p_3 = 1 - \frac{\exp(-4.867 - 1.761x_3)}{1 + \exp(-4.867 - 1.761x_3)} \end{cases}$$

In the formula, p_1 、 p_2 、 p_3 are the probabilities of the traffic congestion types as "unimpeded", "congestion" and "congestion" respectively;

vely;

B. Model test:

Table4: Parallelism Test

Model	-2 Logarithmic likelihood	χ^2	Degrees of freedom	Significant
Conventional	51.990	3.194	1	0.074

As can be seen from the above table, its significance test is $0.74 > 0.05$. Through the parallel test of the model, all logistic functions in

the model are established. The model can be used to evaluate and predict traffic congestion.

C. Goodness of fit test:

Table5: Goodness of fit test

χ^2	Degrees of freedom	Significant
----------	--------------------	-------------

Pearson	41.696	7	0.062
Deviation	31.413	7	0.053

As can be seen from the table, $\text{pearson}\chi^2$ test significant $0.062 > 0.05$, Deviation test significance $0.053 > 0.05$, It shows that the fitting effect of the model is good and can be used to evaluate and predict traffic congestion.

5.2.2 Traffic jam duration forecast

Using this paper, the time prediction model based on BP neural network is established to predict the duration of traffic congestion in Jinshui Road, Zhengzhou City on August 26, 2008, and to test the accuracy of the model

prediction. Seventy percent of the data sample is used to train BP neural networks, and the remaining 30 percent of the sample data is used to test accuracy. It is important to note that when predicting time, the degree of congestion should be predicted first by the regression equation of the previous step, and then the time should be predicted in conjunction with the degree of congestion.

A. Some section average speed forecast value and actual value comparison

Table6: Table of comparison between predicted value and actual value

Predictive value m/s	10.95	16.32	14.43	16.03	15.32	13.93
Actual value m/s	10.94	16.32	14.39	16.02	15.30	13.96
Error	0.01	0.00	0.04	0.01	0.02	0.03

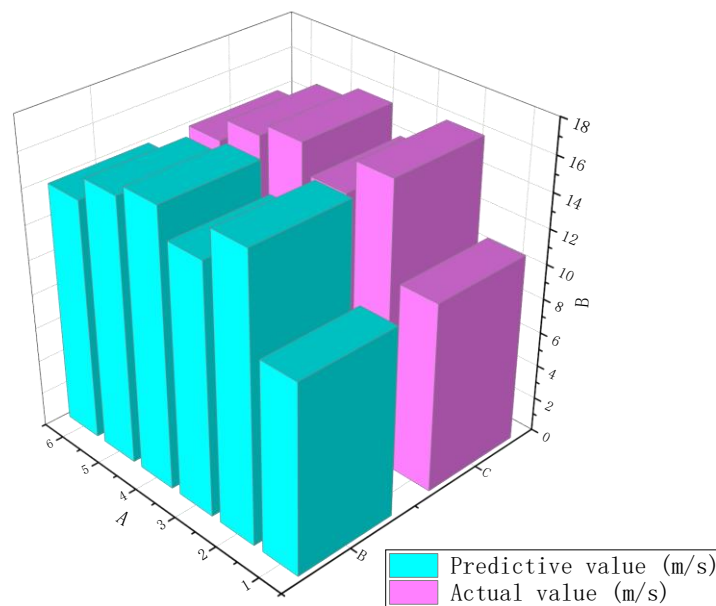


Fig.5: Comparison of predicted and actual values

The above table shows that the prediction value of the average speed of road sections obtained by the model differs very little from the actual

value of the data found, and the model is more accurate.

B. Accuracy analysis

Table7: Error analysis table

Mean error	Percentage of prediction error
------------	--------------------------------

According to the obtained error mean value and the percentage of the prediction error, the percentage of the error is less than 5%, and within the allowable range of the error, the model fitting effect is good.

C. Traffic jam duration calculation

$$t = \frac{d}{x_2}$$

In the formula:

d represents the distance from the congested area, it can be measured by GPS.

t Indicates the duration of the traffic jam;

x_2 represents the mean speed;

5.3 Results analysis

In this paper, the average speed in traffic jam is used to describe the duration of traffic jam in a fixed section. For the example combined in this paper -- Jinshui Road in Zhengzhou, the duration of traffic jam is calculated as follows:

A. The comparison between the predicted value and the actual value:

In this paper, the obtained value is compared with the actual value through MATLAB software (due to the large amount of data, part of the figure is taken here) :

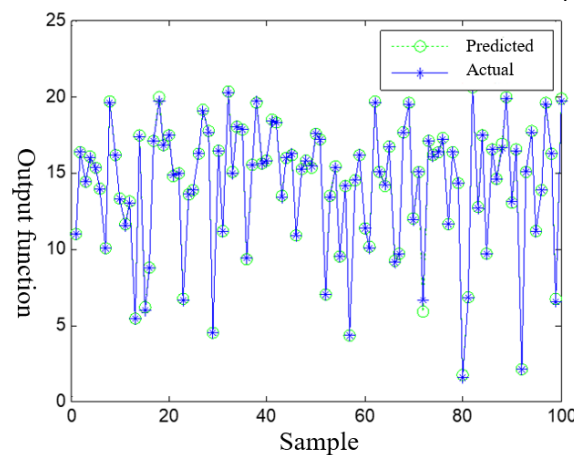


Fig.6: Comparison between predicted value and actual value

From the comparison, it can be seen that the cumulative ratio Logistic regression analysis model combined with BP neural network time prediction model, numerical combination method, standard deviation method and other

methods to obtain a relatively accurate prediction model.

B. Analysis of BP network prediction error diagram:

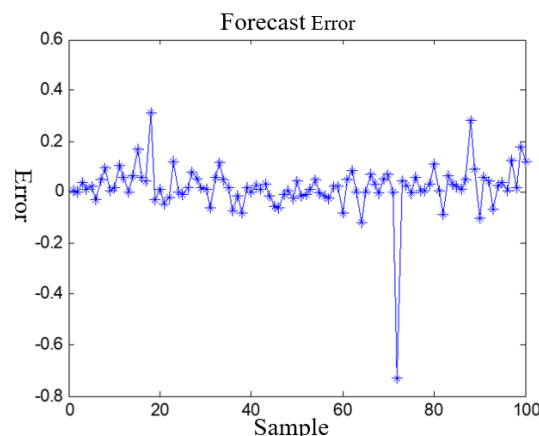


Fig.7: BP network prediction error diagram

As can be clearly seen from the figure above, most of the errors are relatively stable, and fluctuate around 0. Only one point of error is relatively obvious, and the fluctuation is many times higher than other fluctuations. If only from the perspective of data mining, it is likely to be removed as an error value. However, it can be found from the actual situation that it may be caused by special events, such as traffic acci-

dents that lead to a particularly congested road section, resulting in a difference of several times between the results of the established model and the actual value.

According to the statistical calculation results of this paper, the pie chart of relative error value size and area is obtained, as shown in the figure below:

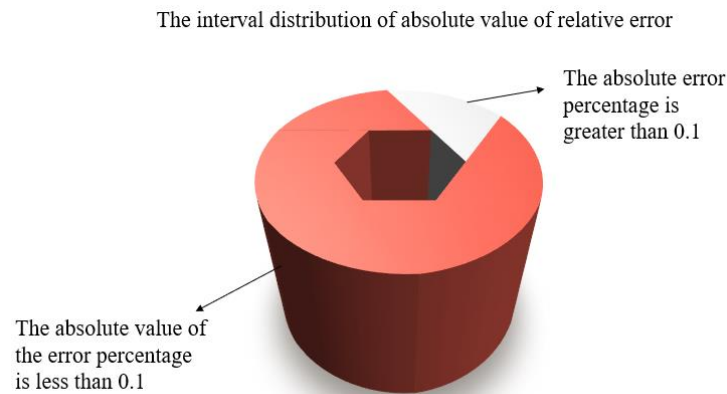


Fig.8: Relative error scale diagram

BP Neural Network has a good prediction effect, with a relative error of less than 0.1 over half.

6.1 Advantages of the model

(1) In this paper, when considering the prediction of traffic congestion time, the consideration is more comprehensive, from the time, traffic actual status and other special factors three angles of the prediction of time.

(2) In this paper, the BP neural network used to predict, compared with the traditional gray prediction and other prediction methods, has a higher degree of precision

6.2 Model shortcomings

(1) This paper is too mining the characteristics of the data itself, but the actual significance of the data itself is ignored

(2) In the use of BP neural network for prediction, the input factor sits not considering the weight size of the impact, but as an equal weight input.

7. Improvement and promotion of the model

7.1 Model improvement

(1) When predicting traffic congestion time, the weight stakes can be calculated using the entropy method, making the prediction more

accurate.

(2) This paper is only a prediction of the duration of traffic congestion, but the accuracy of the prediction is not known, the probability of such a situation is not known, therefore, we can combine the survival analysis function, the use of relevant knowledge and research on the time of each congestion section, and then solve the probability of the occurrence of traffic congestion duration.

7.2 Generalization of the model

The model established in this paper is only used to predict the congestion time of traffic roads, but it may also provide some help in the field of ship and fan flight in real life.

8. Conclusion

In this paper, a traffic jam evaluation model based on cumulative ratio Logistic regression and a traffic jam time prediction model based on BP neural network were established to predict the traffic jam time. Taking Jinshui Road in Zhengzhou City as an example, the model was applied. The model passed the rationality test, and the average error between the predicted

value and the actual value was 0.019m/s, and the prediction error rate was 0.15%, which was reasonable. Therefore, the model established in this paper can improve the prediction accuracy of congestion time, and can also be applied to shipping and other transportation industries, providing certain help to actual life.

References

1. Juan Tan, Shengchun Wang. Research on traffic jam prediction model based on deep learning[J]. Computer application research, 2015,32(10):2951-2954.
2. Jixin Liu, Xiaoyu Zeng, Minjia Yin, Xuehua Zhu. Prediction of controller stress degree based on cumulative Logistic regression model[J]. Journal of Chongqing Jiaotong University(natural science), 2019,38(03):97-102+115.
3. Yingxue Zhou, Xiaobao Yang, Mei Huan, Bingfeng Si. Study on the duration of urban road traffic congestion based on survival analysis[J]. Applying mathematics and mechanics, 2013,34(01):98-106.
4. Wencheng Wang, Xiaobao Yang, Yingxue Zhou, Mei Huan. Analysis on the duration of urban road traffic congestion based on Cox risk model[J]. Journal of Chang 'an University (natural science edition), 2015,35(S1):235-239.
5. Xuelian Zhang, Lei Yu, Menghan Liu. Traffic congestion evaluation model based on traffic demand[J]. Modern transportation technology, 2008,5(06):71-75.

