Research Article **ARR (2020) 1:7**

Advances in Research and Reviews

(ISSN:2692-1146)

Predicting Traffic Congestion Time Based on Kalman Filter Algorithm

Tianlong Wang[1]*, Xiaorui Tao[2], Jiamei Zhang[3], Yulei Li[4]

[1]College of Civil Engineering & Architecture, China Three Gorges University. [2]College of Economics & Management, China Three Gorges University. [3]College of Computer and Information Technology, China Three Gorges University. [4]College of Electrical Engineering & New Energy, China Three Gorges University, Yichang, 443002, China.

ABSTRACT

This paper mainly solved the problem of predicting the time required for vehicles to pass through congested roads. In order to obtain more accurate prediction time, a Kalman prediction model based on multiple linear regression was established in this paper. Taking the 2008 Yanan elevated road in Shanghai as an example, the measured data in this section was collected from the traffic measured data sharing network, and the above model was used to obtain good prediction results. As an improvement, we used BP neural network instead of multiple linear regression to make the prediction result more in line with the actual situation.

Keywords: Multiple linear regression equation; Kalman filter algorithm; BP neural network.

***Correspondence to Author:**
Tianlong Wang
College of Civil Engineering & Architecture, China Three Gorges University, Yichang, 443002, China.

How to cite this article:
Tianlong Wang, Xiaorui Tao, Jiamei Zhang, Yulei Li. Predicting Traffic Congestion Time Based on Kalman Filter Algorithm. Advances in Research and Reviews, 2020; 1:7

eSciPub LLC, Houston, TX USA.
Website: https://escipub.com/

1. Introduction

In the navigation software, the estimation of travel time is often an important function. Existing navigation software often obtains real-time GPS data through the taxies or the vehicles installed the software to determine the current road conditions. In the case of severe traffic jams, the speed of the car is slow, so the estimation of the speed is very inaccurate. The consequence is that the accuracy of the estimated traffic jam time is very poor. The actual time required is sometimes even several times to ten times different from the forecast time. To this end, a traffic congestion time prediction model based on Kalman filtering is established in this paper. The roads on various electronic maps now show different colors, representing different congestion conditions on the roads. The degree of road congestion can be measured by the traffic index, which can vary between 0 and 100 and is in a different value range, representing different degrees of road congestion. When the traffic index is greater than 50, the road is congested. If the traffic index is less than 50, the road is not congested and the traffic condition is good. Since the traffic data in the past can only be used to predict the traffic situation in the future, until the moment when the traffic situation is good (that is, when the traffic index is less than 50), it can be considered that the vehicle has passed the blocked road, and the time for the vehicle to pass the blocked road can be calculated. Kalman filter algorithm can be applied to obtain an optimal prediction result according to the predicted value and measurement value of each moment, and the optimal output predicted value of each step can be calculated recursively according to the historical time series, a more accurate predicted value. According to Kalman filter algorithm, the predicted value of traffic index in the past moment must be known. We can select five relevant factors: the average travel speed, the average travel time, the average delay time, the morning and evening peak periods and the weather conditions of vehicles on a certain road at a certain time, and establish a multiple linear regression equation with the traffic index as the dependent variable. A certain amount of data is selected to solve the problem, and then the test is carried out. If the test is passed, the model can be proved to be reasonable. By applying Kalman filter, the traffic index value of each moment is recursively derived, and the traffic jam time is calculated.

2. The Foundation of Model

2.1 Multiple linear regression model

The situation of road congestion in each section will be different at each time. To predict the time of road congestion, we must first predict the state of road congestion, that is, the degree of road congestion. We try to establish a traffic congestion model, and take time into account. Multiple key influencing factors such as average travel speed, average travel time, average delay time, morning and evening peak periods and weather conditions are selected as indicators of the traffic congestion model. The specific definition is as follows.

Here, we selected five kinds of data of Shanghai Yanan viaduct in different periods of sunny, rainy and snowy days (one day each) in 2008 for analysis.

(1) Average travel speed $\bar{v}$:

The average travel speed is the average travel speed of all motor vehicles in the same distance during the unit time period. The unit is *km/h*,

Average travel speed is an important indicator to describe the degree of road congestion, which is negatively correlated with the degree of congestion. Its calculation formula is shown as follows:

$$\bar{v} = (L \times N) / \sum_{i=1}^{N} t_i$$

Among them, $\bar{v}$ Represents the average travel speed, unit: *km/ h*; *L* Represents the length of the road section, unit: *km*; *N* Represents the total number of vehicles passing the section during the period, unit: vehicles; *i* Represents the vehicle number; t_i represents the time when the *i* vehicle passed the section, unit: *h*.

(2) Average travel time $\bar{t}$:

Average travel time refers to the unit walk through all the vehicles in the average amount of time, the average travel time is negative correlation with the degree of congestion, the longer the average travel time, traffic jams more serious, the shorter the average travel time, traffic is smooth, its formula is as follows:

$$\bar{t} = (L \times N) / \sum_{i=1}^{N} v_i$$

Among them, $\bar{t}$ represents the average travel time, unit: *h*; L represents the length of the road section, unit: *km*; N Represents the total number of vehicles passing the section during the period, unit: vehicles; i represents vehicle number; v_i Represents the average speed of the i vehicle passing the section, unit: km / h.

(3) Mean delay time T_i

Average delay time is the amount of time lost to external factors (inclement weather, traffic accidents, etc.) during a unit journey. The average delay time can reflect the unimpeded traffic and queuing conditions, and is an important indicator reflecting the degree of traffic congestion. Its calculation formula is as follows:

$$T_l = L / V_t - L / V_n$$

Among them, L represents the length of the road section, unit: km; V_t is the actual average speed of vehicles passing this section, unit: km / h; V_n represents the speed of vehicles passing this section of road in the free state, and can be regarded as the speed limit of this section (the maximum speed allowed for vehicles on this section of road).

(4) Morning peak: Influenced by residents' commuting, traffic jam is serious in morning and evening rush hours with strong regularity. It is generally believed that the morning rush hour is from 7:30 to 9:30. The evening peak time is 16:30 ~ 18:30. Can be defined:

$$Y_1 = \begin{cases} 1 & \text{the peak} \\ 0 & \text{off-peak} \end{cases}$$

(5) Weather impact:

Due to bad weather, the speed of vehicles on the road will be affected to varying degrees, and the traffic situation will change accordingly. Rain, snow and fog are more likely to cause traffic jams. Now it is simple to calculate. Weather conditions other than rain or shine, such as rain, snow, haze, hail and other severe weather are regarded as interference factors, and can be defined as:

$$Y_2=\begin{cases}1 & \text{sunny day}\\ 0 & \text{other weather}\end{cases}$$

According to the above five factors, five independent variables can be obtained, taking road traffic index as the dependent variable y , the five independent variables are denoted as x_1 、 x_2 、 x_3 、 x_4 、 x_5 , establish a multiple linear regression model, which can be expressed as:

$$\lg(y)=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+\beta_5x_5+\beta_6x_6$$

Among them, β_0 is a constant; β_i Is partial regression coefficient ，When controlling the linear influence of other independent variables on the dependent variable, it indicates the degree of linear influence of independent variable x_i $(i=1,2,\ldots,6)$ on the dependent variable y. The least square method is used to calculate the coefficient value, and then the significance test is carried out.

2.2 Significance verification

(1) R^2 validation

$$R^2=\frac{SSR}{SST}=\frac{\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})}{\sum_{i=1}^{n}(Y_i-\bar{Y})}$$

The value range of R^2 is between 0 and 1. The closer R^2 is to 1, the better the fitting of the regression equation will be. In general, when R^2 is greater than 0.8, the fitting condition can be considered as good.

(2) Hypothesis testing

●Put forward the hypothesis:

H_0： $\beta_1=\beta_2=\beta_3=\beta_4=\beta_5=\beta_6=0$, the linear relationship is not significant;

H_1： β_1、 β_2、 β_3、 β_4、 β_5、 β_6 ,at least one of them is not equal to 0.

●Calculate the test statistics:

The total change in the sum of squares

$SST=\sum_{i=1}^{n}(Y_i-\bar{Y})$,regression sum of squares

$SSR=\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})$, sum of squares of residuals

$SSE=\sum_{i=1}^{n}(Y_i-\hat{Y}_i)$,p is the number of regression coefficients, and n is the number of samples

$$F=\frac{SSR/p}{SSE/(n-p-1)}\sim F(p,\ n-p-1)$$

●Determine the significant level $\alpha=0.05$, and find out F_α from the table of

molecular degree p, $n-p-1$.

● Make A decision, if $F\geq F_\alpha$, reject the null hypothesis;If $F<F_\alpha$, accept the null hypothesis.

For the regression equation with multiple variables, it is necessary to analyze whether the combination of all factors is the optimal scheme even though it has passed the hypothesis test. In other words, when H_0 is rejected, β_i is not all 0, but β_i is equal to 0. Therefore, stepwise elimination method is adopted to analyze and eliminate the insignificant variables. Hypothesis testing is conducted for each β_i . If it is significant, it can be retained; if it is not significant, it needs to be removed.

2.3 Kalman filter algorithm

In road traffic, often happen in rush hour, easily blocked roads traffic congestion phenomenon, in the case of a workshop from smaller, car slowly, through satellite observations of road congestion index is not very accurate, just by GPS is unable to meet the need, when at the same time have at the moment of the measured data, and the moment to predict data, can weaken error, at the moment of the optimal estimate congestion index calculation, the predictive value of the next moment, we can achieve the goal by Kalman filtering algorithm. The moment time is given according to the definition. When the traffic is no longer crowded, the road traffic index is less than 50, and the corresponding moment is the traffic jam time we need to adjust the step size for several times.

Kalman filtering algorithm is put forward by American researchers Kalman in the 1960 s, through minimum variance, Kalman filtering algorithm is used to make optimal estimation, Kalman filter algorithm need real-time access to information and updating parameters, it can reduce the interference of noise processing method is given, and used to predict future system and so on the state of the system to eliminate interference noise constantly, and calculate the final revised forecast results. The model can obtain the prediction results through real-time updating, so it is applicable to many interferences and has good accuracy and time lines [1].

The prediction process of Kalman filtering algorithm needs to be explained by introducing some equations, which are specified as follows:

(1) State update equation

$$X(k|k-1) = AX(k-1|k-1) + BU(k) + W(k)$$

$$Z(k) = Hx(k|k-1) + V(k)$$

In the top one, $X(k|k-1)$ represents the predicted road traffic index at timek by calculating the time $k-1$of the predicted road traffic index; $X(k-1|k-1)$ represents the predicted road traffic index at time k which is predicted by the last regression model; $U(k)$ represents the amount of control at time , k can take 0; $W(k)$ represents the process noise; A is the state transition matrix; B is the control matrix; $Z(k)$ represents observations at time k ;H represents the observation matrix, can go to the identity matrix; $X(k)$ represents the real value at time k ; $V(k)$ represents observation noise. The predicted value of the road traffic index at the last moment is input into the system, and the predicted value of the road traffic index at the time k is obtained.

(2) Update the covariance matrix

$$P(k|k-1) = AP(k-1|k-1)A^T + Q$$

In the top one, $P(k|k-1)$ is covariation of $X(k|k-1)$, $P(k-1|k-1)$ is covariation of

$X(k-1|k-1)$;Q is the covariance matrix of prediction noise, in other words, it is covariance matrix of $W(k)$.

(3) Kalman gain

$$Kg(k) = P(k|k-1)H^T(HP(k|k-1)H^T + R)$$

In the top expression, H represents the observation matrix, R represents observe the covariance matrix of noise.

- An optimal estimate of the state

$$X(k|k) = X(k|k-1) + Kg(k)(Z(k) - HX(k|k-1))$$

In the top expression, $Kg(k)$ represents Kalman gain, $Z(k)$ represents measured value at time k, that's the congestion index at the moment, is the parameter of the measurement system. At this time, the optimal estimate $X(k|k)$ of road traffic index at time k is obtained.

- Update the covariance matrix of time

$$P(k|k) = (I - Kg(k)H)P(k|k-1)$$

In the top one, When the optimal estimate at time k is calculated, at the moment of $k+1$, doing next iteration, and then taking $P(k|k)$ to the second formula for $P(k-1|k-1)$, to get the covariance matrix $P(k+1|k)$, And then you recirculate it to get the congestion index at time $k+1$.

When different durations are defined, different predicted road congestion index can be obtained. To stop congestion and make the index less than 50, multiple calculations are required. When the road traffic index meets the requirements, the corresponding time length is the required congestion time.

3. Solution and Result

3.1 Solution of linear regression equation:

When the stepwise culling method is adopted for analysis, we eliminate some variables, conduct significance test on the dependent variables for the remaining variables, and finally retain a group of independent variables that have the most significant impact on the traffic index. The test results of the stepwise elimination method are shown in the following table:

Table 1: The result of phase-out

reserved	**variable** R^2
x_1、x_2、x_3、x_4、x_5	0.9271
x_1、x_2、x_3	0.9211
x_1、x_2、x_5	0.9214
x_2、x_3、x_4、x_5	0.6943
x_2、x_3、x_4	0.6117

As can be seen from the above table, when the reserved variable is x_1、x_2、x_3、x_4、x_5, the value of R^2 is the largest and the test effect is the best. In other words, when the independent variable is x_1、x_2、x_3、x_4、x_5, it has the greatest influence on the dependent variable traffic index, so no variables should be removed and all five variables should be retained.

Using MATLAB software to solve, the obtained linear regression equation is:

$$y = 119.9913 - 1.6533x_1 - 12.9156x_2 + 42.3598x_3 - 5.1566x_4 + 9.0772x_5$$

After R^2 verification and hypothesis test, the test result is:

R^2=0.9173, F=29.5726. Means the test result is qualified and the linear regression equation has a good fitting effect.

3.2 Results of Kalman filter model

After obtaining the linear regression equation, the predicted value of the traffic index at the previous moment can be obtained, and then the optimal estimated value of the traffic index at the future moment can be derived according to the measured value of the current traffic index. The specific algorithm of Kalman filter is as follows:

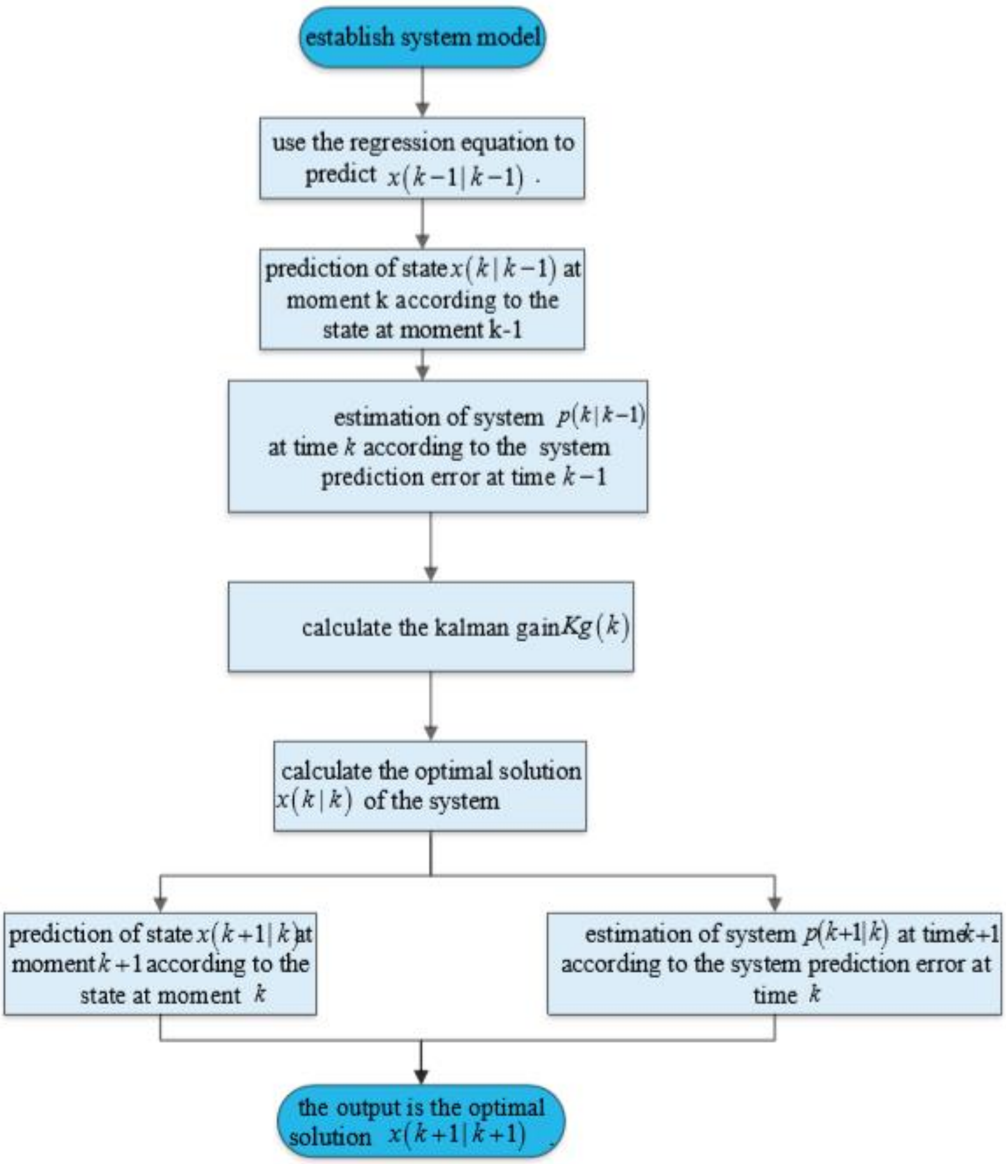

Figure 1: The flow chart of Kalman filter algorithm

Observed at 11:30 on the day when the vehicles average travel speed of 42, the average travel time was 0.1625, 0.147, the average delay time is not the morning rush, the sunny weather, road traffic index prediction is 49.5242, in 10 minutes, for step length, the next moment is the observed value of 51.5, into Kalman prediction, can be predicted after 133 minutes, no longer blocked the road.

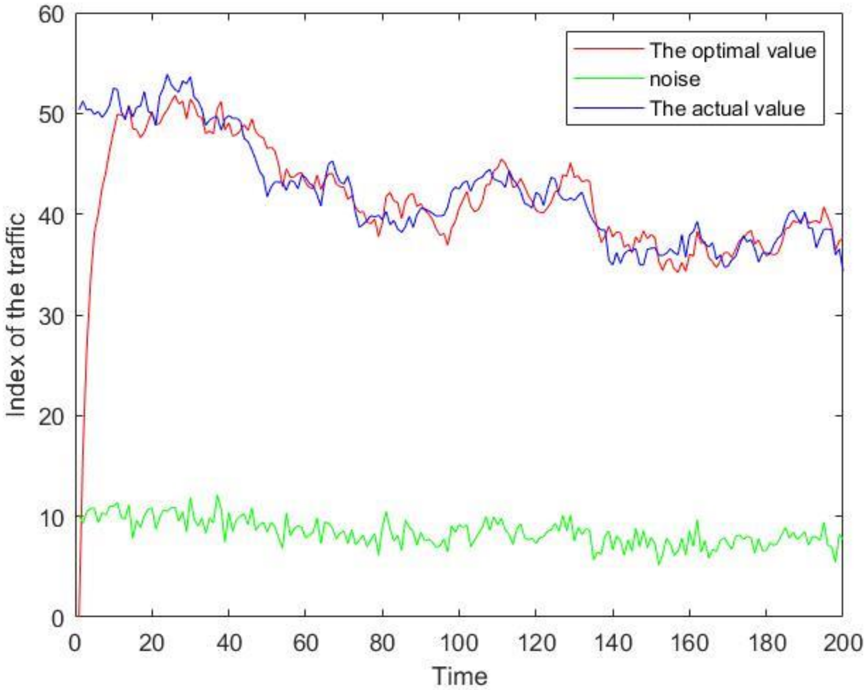

Figure 2: index values in kalman algorithm

4. Improved Model

Last model, we build the linear regression equation of Kalman filtering model: after solving the linear regression equation, the application of the five independent variables to predict a moment before traffic index prediction, then the application of Kalman filtering model introduced the predicted value of the current traffic index, the optimal estimation, the optimal transportation index future time estimates. On this basis, we improve the linear regression equation and adopt BP neural network model. We still adopt the average travel time, the average travel speed, the average delay time, whether it is rush hour, the weather is sunny these five variables as input, road traffic index as the output variable, the training of the neural network application, and correct weights and bias, after training the BP neural network model.

5. Results of BP neural network:

The BP neural network is applied to train multiple groups of data and continuously modify to obtain the optimal model. The measured value of traffic index and the fitting curve of traffic index output by the neural network model are shown in the following figure:

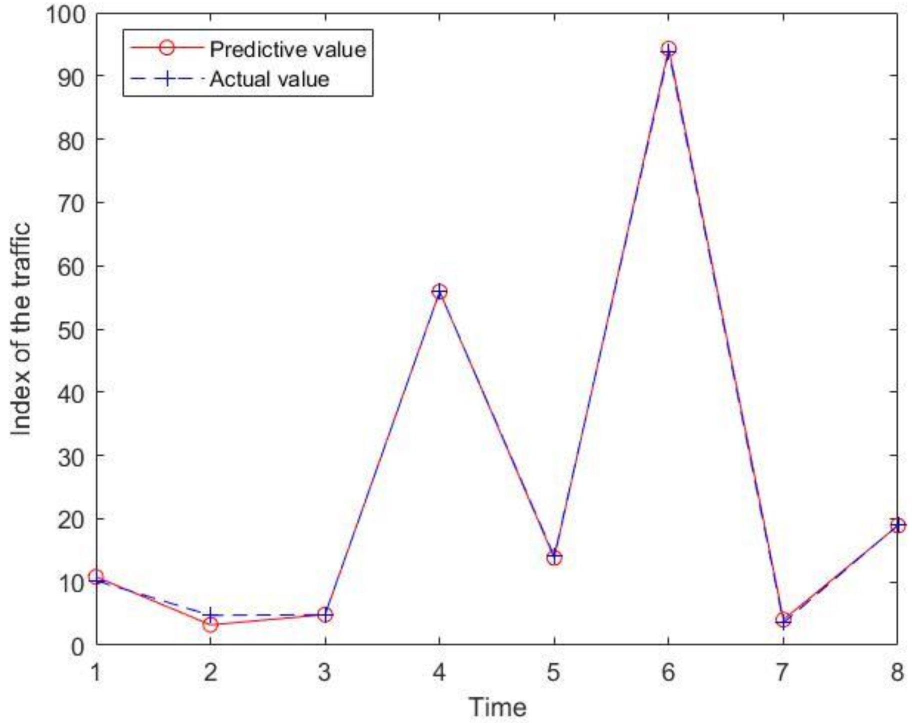

Figure 3: curve fitting the measured value of traffic index and the output value of BP neural network

As can be seen from the figure above, the predicted value of traffic index obtained by BP neural network has a small error with the measured value of traffic index, and its fitting curve is very close, which means that the trained neural network can get a better prediction result. Based on this, the Kalman filter algorithm can be carried out in the next step.

6. Conclusions

In this paper, suitable indexes are selected and a traffic congestion time prediction model based on Kalman filtering is established. Multivariate linear regression equation and BP neural network are used to predict the traffic index at the previous moment by combining the average travel time, average travel speed, average delay time, rush hour and weather at the previous moment, and then use Kalman filtering to recursively derive the future Algorithm of traffic indicators at the moment. By comparing the two, it can be concluded that the traffic indicators predicted by the BP neural network are more accurate. The model improves the accuracy of road congestion time prediction.

References

1. Zhenchao Hou. Prediction of bus journey time based on the combined model of Kalman filter and exponential smoothing method [D]. Liaoning university,2018.
2. Yuxin Long. Mathematical model of urban highway traffic jam warning [J]. Science and technology economics guide,2018,26(03):183.
3. Moggu Xu. Prediction and evaluation of traffic congestion on urban trunk roads under accident conditions [D]. Chang 'an university,2017.
4. Siqi Yang. Research and application of expressway traffic congestion prediction model [D]. South China university of technology,2017.
5. Chunjiao Dong, Chunfu Shao, Xuemei Zhou, Meng Meng, Gechengxiang Zhu. Short-time prediction Kalman filter algorithm based on traffic flow parameters [J]. Journal of southeast university (natural science edition),2014,44(02):413-419.
6. Wei Nie. Study on characteristics of expressway traffic accidents and induced control strategies

[D]. Southwest Jiaotong University,2012.

7. Chunhui Zhang, Rui Song, Yang sun. Short-term passenger flow prediction of bus stations based on Kalman filter [J]. Transportation system engineering and information,2011,11(04):154-159.
8. Zhong Zhu, Zhaosheng Yang. Real-time travel time prediction model based on Kalman filter theory [J]. Theory and practice of systems engineering,1999(09):74-78.