第八届“认证杯”数学中国

数学建模国际赛

承　诺　书

我们仔细阅读了第八届“认证杯”数学中国数学建模国际赛的竞赛规则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的, 如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们允许数学中国网站(www.madio.net)公布论文，以供网友之间学习交流，数学中国网站以非商业目的的论文交流不需要提前取得我们的同意。

我们的参赛队号为：

我们选择的题目是：　　C 题

参赛队员　(签名) ：

队员 1：刘静博

队员 2：孔子庆

队员 3：江润强

参赛队教练员 (签名)：　王蓓

第八届“认证杯”数学中国

数学建模国际赛

承　诺　书

我们仔细阅读了第八届“认证杯”数学中国数学建模国际赛的竞赛规则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们允许数学中国网站(www.madio.net)公布论文，以供网友之间学习交流，数学中国网站以非商业目的的论文交流不需要提前取得我们的同意。

我们的参赛队号为：

我们选择的题目是：

参赛队员　(签名)　：

队员 1：刘静博

队员 2：孔振

队员 3：江润强

参赛队教练员 (签名)：[illegible]

第八届“认证杯”数学中国

数学建模国际赛

编 号 专 用 页

参赛队伍的参赛队号：（请各个参赛队提前填写好）：

3424

竞赛统一编号（由竞赛组委会送至评委团前编号）：

竞赛评阅编号（由竞赛评委团评阅前进行编号）：

Research on traffic jam time prediction based on speed

Abstract:

In the case of serious traffic jam, the existing navigation software is very inaccurate in the estimation of speed, which leads to the inability to accurately predict the time of traffic jam, with huge error. Therefore, aiming at the above problems, we put forward a mathematical model based on speed prediction to accurately predict the time required to cross the crowded road section.

As for the choice of parameters to measure the traffic state, based on the summary of some previous theories and research, this paper chooses the most direct speed to measure the vehicle traffic state as the traffic state parameter. For the judgment of traffic state, the interim technology of operation monitoring and service of China's highway network is selected to divide the road state into five states according to the speed: unblocked, basically unblocked, mildly congested, moderately congested and severely congested.

As for the establishment of the mathematical model of speed prediction, this paper discussed the speed prediction of weekday and weekend respectively based on the difference of the state and trend of car speed in weekday and weekend. The data set was taken from the GPS information of 14,000 taxis in chengdu in 2014. Speed prediction model is divided into KNN - VA with KNN - RBF two forecasting model, based on Euclidean distance similarity metric of KNN algorithm selection and sample high similarity of sample under test combination, among them to choose the samples of the KNN - NA predicted using the speed of the sample average as speed, KNN - RBF of selected samples after processing, the moving average filtering of the speed by using the RBF network forecast.

For the predicted value, root mean square error MSE and absolute error were adopted for error analysis. Knn-va predicted the model's root mean square error 0.009 and absolute error 0.2168069 within a week on August 18, 2014, while the model's root mean square error MSE for Saturday, August 23, 2014 was 0.013018 and absolute error 0.33823. Knn-rbf's root mean square error MSE of the model was 0.000814153 within the working week of August 18, 2014, and the root mean square error MSE of the model root mean square error model of the prediction model on Saturday, August 23, 2014 was 0.00143321. On the whole, knn-rbf model was better than knn-va model.

For the determination and analysis of traffic congestion degree and congestion time, the time period from 12 to 18 on August 18, 2014 was taken for analysis. 24 points were sampled in 15min. The accuracy of knn-rbf speed prediction method in predicting the traffic condition was 91.7%. Taking the time of 2014-08-2308:00 as an example, the coordinate latitude and longitude of the section (30.6214, 104.094), (30.6195, 104.058) was predicted to take 6.985974246636439 minute.

Key words: KNN; RBF; Speed predict; Error analysis;Time predict

Contents

1 Introduction

1.1 Background

In navigation software, travel time estimation is often an important function. Existing navigation software often USES taxis or other vehicles installed with the software to obtain real-time GPS data to determine the current road conditions. In heavy traffic, cars are slow, so estimates of speed are wildly inaccurate. This has led to poor accuracy in predicting the time of a traffic jam, with the actual time sometimes several to ten times less than the predicted time.

Intelligent TransportationSystem can integrate the information, automation and intelligence of the TransportationSystem. This technology has been widely studied and applied in related fields. Based on the above background, this paper establishes a speed prediction model and a judgment method for traffic condition, so as to accurately predict the time required for crossing the crowded road section

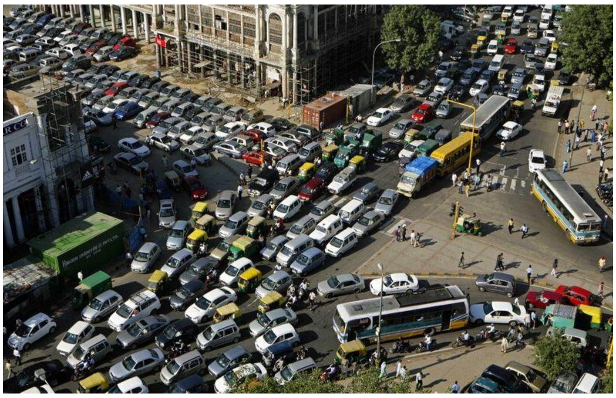

Fig. 1 A picture of the traffic jam

1.2 Previous Research

At present, given that the prediction accuracy is low in the case of severe traffic congestion, a large number of scholars have conducted a large number of studies on the prediction of traffic congestion. For the prediction of traffic jam time, it mainly starts from two aspects, namely the prediction selection of traffic parameters and the establishment of prediction model.

For the prediction and selection of traffic parameters, the traffic state is generally predicted from the parameters of vehicle speed, vehicle density, traffic flow, etc. These prediction parameters and indicators can reflect the actual traffic situation in a real way.

For the establishment of prediction model, there are mainly the combination of single parameter model and different models. Most single-parameter models are established in common use models such as neural network model, time series model and grey theory model. In order to solve the problems of single parameter prediction model with single applicable conditions, high data requirements and poor generalization ability, a variety of prediction models are generally integrated, such as fuzzy neural network model and wavelet neural network model, etc. This kind of model also has a large time cost and increased computational complexity.

1.3 Our Work

1. Analyze and compare the research and application of traffic parameters such as speed, traffic flow density and saturation in traffic state prediction. From the integration of the two perspectives of current use and recommended use, we take speed as the prediction parameter of traffic state

2. From the perspective of prediction based on speed, this paper analyzes the relationship between the change of speed and its size and the degree of traffic congestion. According to the detection requirements of China's road network for traffic conditions, we divide road congestion into unblocked, basically unblocked, slightly congested, poisoned congested and seriously congested.

3. Through data analysis, we finally established two velocity prediction models. One is knn-va. KNN algorithm is used to select k samples, and the average value of the past velocity of these k samples is used to predict the future velocity. One is knn-rbf. Also, k samples are selected by KNN algorithm to predict the velocity in the target period through RBF neural network

4 in view of the inconsistencies between the prediction results of knn-va and knn-rbf models, it is considered that the root-mean-square error function of time interval is used as the evaluation standard of the prediction model.

2 Analysis of Overall and Key Points

2.1 Overall Analysis

For, data were collected using traditional traffic condition prediction model, only short-term prediction, cannot completely solve the problem of traffic state forecasting, therefore, we through the historical data of time series as input, to predict the future period of speed, which can not only predict the future traffic condition, also can be calculated through the crowded roads needed to spend time.

2.2 Key Points Analysis

2.2.1 Analysis of traffic parameters

According to the results of data analysis, when the traffic is crowded, the most obvious response of all traffic parameters is to select the vehicle speed.

In order to determine the accuracy and reliability of many traffic parameters, many researchers draw conclusions through a large number of questionnaires. The survey results show that in the identification of vehicle congestion state, the prediction of traffic state based on speed occupies the highest proportion, which is much higher than other parameters. Among the traffic status identification parameters recommended by scholars, the utilization rate of speed parameter accounts for 89%, much higher than other parameters. From the statistical results of the bar chart, it can be seen that among the traffic status identification parameters, the highest recommendation degree is the speed parameter. It is reasonable to select the speed as the measurement index of traffic congestion.

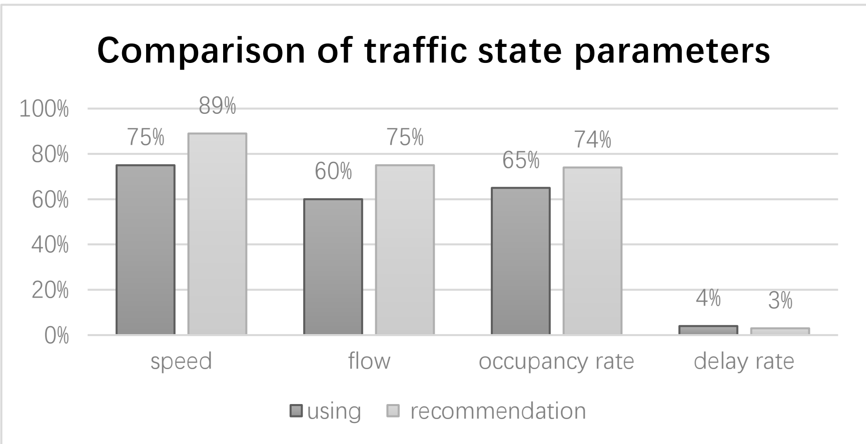

Fig. 2 Comparison of traffic state parameters

2.2.2 Traffic status determination

Based on the model of speed prediction, this paper divides the road state into five states, namely unblocked, basically unblocked, mildly congested, moderately congested and severely congested. And set k1,k2,k3 and k4 as the threshold of section travel speed. By comparing the parameters of observed data with the set threshold, the current traffic status can be identified. Since the traditional discriminant method can only make short-term prediction based on the observed data and cannot meet the requirements of modern traffic prediction, it is necessary to establish a prediction model with historical data as samples and speed as parameters. The traffic state prediction model can be obtained as follows:

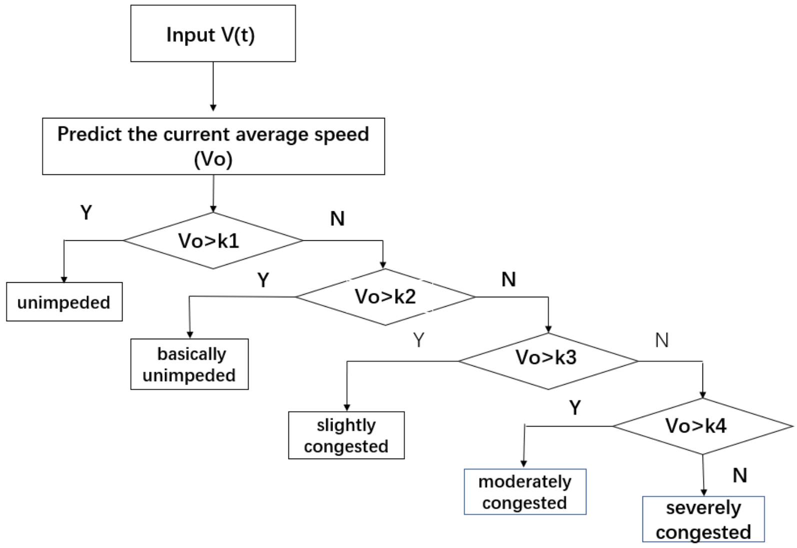

Fig. 3 Congestion based on speed

2.2.3 KNN

KNN is a machine learning algorithm that can be used for both classification and regression. For a given test sample, find K training samples closest to the training set based on the distance measurement, and then make prediction based on the information of K "neighbors". In the classification task, voting method can be used to select the category marker that appears most in these K samples as the prediction result. You can also select the first tag in the sequence in descending order as the category of the test sample. The flow chart of KNN algorithm is as follows:

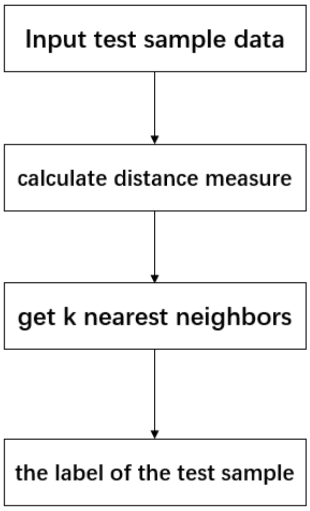

Fig. 4 KNN algorithm

2.2.4 RBF

Radial basis function neural network is a three-layer forward network with a single hidden layer. The first layer is the input layer, which is composed of signal source nodes. The second layer to hidden layer, hidden layer node number depending on the described problems need, transformation function of neurons in hidden layer radial basis function is the center of radial symmetry and attenuation of nonnegative linear function, the function is local response function, local response of the concrete is embodied in its visible to hidden layer of transformation is different from its network. The output layer adjusts the linear weight and adopts the linear optimization strategy, so the learning speed is faster. The network structure is as follows:

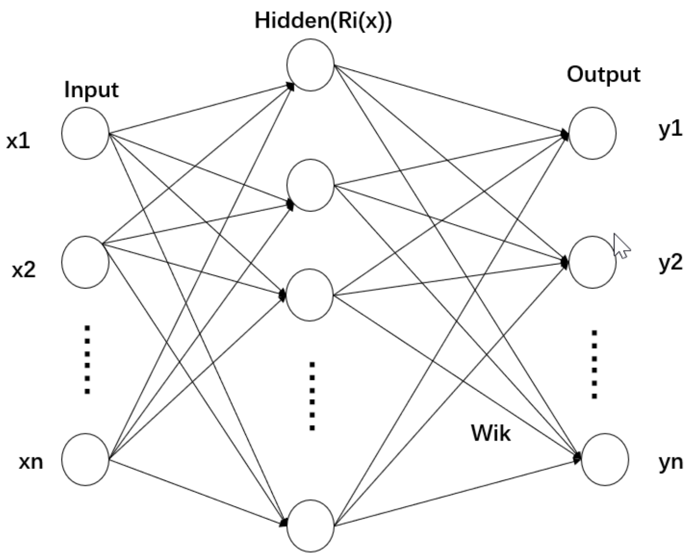

Fig. 5 RBF algorithm

3 Assumptions

Assumption I:

To judge the congestion degree of traffic state, it is necessary to set the congestion degree limit through traffic state parameters. The standards for parameter setting of traffic state vary from country to country. The standards of the model in this paper meet the requirements of the interim technology of operation detection and service of China's highway network.

Assumption II:

In the process of data analysis and data mining, we often need to know the size of the differences between individuals, so as to evaluate the similarity and category of individuals. The main methods to measure differences among individuals are distance measurement and similarity measurement. Distance measurement is used to measure the spatial distance of individuals. The farther the distance is, the greater the difference between individuals will be. Representative methods include Euclidean distance, minkowski distance, Manhattan distance, chebyshev distance and so on. Similarity measurement is to calculate the degree of similarity between individuals. Contrary to distance measurement, the smaller the value of similarity measurement is, the smaller the degree of similarity between individuals is, and the larger the difference is. The typical methods are cosine similarity of vector space, Pearson correlation coefficient and so on. In this paper, Euclidean distance is used as the measurement standard of KNN algorithm.

Assumption III:

There are some outliers and missing values in the vehicle speed data set. The process of data preprocessing is the process of data cleaning, in which missing values include deletion method, filling method, regression method, maximum likelihood method and so on. Considering that velocity is time-related data, the missing value is supplemented by the average value of historical data. The processing methods of outliers are generally deletion method and correction method. In the same way, considering the object of the problem, the data is filtered by sliding average, and the result is taken as the velocity value of the current moment.

4 Symbols and Definitions

In the section, we use some symbols for constructing the model as follows:

Table 1 Symbols and Definitions

d	Euclidean distance
V	velocity
Y	velocity prediction results
Vmin	minimum velocity
Vmax	maximum velocity
msev	KNN-VA mean square error
mser	KNN-RBF mean square error
yv	KNN-VA prediction result
yr	KNN-RBF prediction result
KNN	k-nearest neighbor
RBF	Radial Basis Function Neural Network

5 Model

The overall process of the speed-based traffic state prediction model is as follows: data preprocessing, sample extraction, sample selection, model prediction and error analysis.

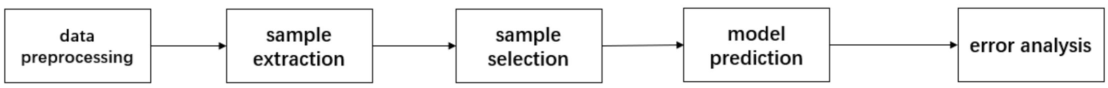

Fig. 6 The whole process

The data set used in this model is more than 1.4 billion GPS records of 1.4+ 10,000 taxi-taxis in chengdu in 2014, from Aug. 03 to Aug. 30, 2014. Statistics include taxi ID, latitude and longitude, passenger status, and time.

Data pretreatment steps are as follows:

(1) we ignored the influence of passenger carrying status on the data prediction, and chose taxi ID, longitude, latitude and time as our sample data

(2) by mapping the GPS longitude and latitude information of the taxi throughout the day, we drew the taxi route map of the city

(3) through data analysis, we finally chose the crowded path in the center of the city, as shown in Fig.7. And according to the latitude and longitude filter in the path within the taxi data.

(4) first, we separate the data with taxi ID.

(5) missing values were processed for each taxi data fragment, and duplicate lines were deleted. Due to the delay in GPS data collection, the data were not arranged in order of time, so we rearranged them according to time

(6) check each taxi data fragment and eliminate the fragments with less than 3 pieces of data. The purpose of this is to ensure that the subsequent speed calculation will not report errors and make the data more accurate

(7) calculate the speed corresponding to each time point for each taxi data segment according to its adjacent longitude and latitude interval and time difference

(8) merge all taxi data fragments, aggregate the merged data, average the speed at the repetition time point, and finally get the speed data on the selected path.

(9) finally, the processed data are removed by sliding filtering.

Fig. 7 Area map of target location

Take July 28, 2014 solstice, August 3, 2014 from 6pm to 24pm as an example. As can be seen from the variation diagram of speed and time, the speed trend of weekdays on Monday and Friday is basically the same, while the speed trend of weekends may have some deviation. From 6:00 a.m. to 7:00 a.m. on weekdays, the road passes smoothly without any traffic jam. From 7 o 'clock to 9 o 'clock, the speed of the road slowed down gradually, and congestion began to appear, which may be caused by people having to work early in the morning. From 9:00 to 12:00 in the morning, the road traffic speed is relatively stable, belonging to the normal road traffic. From 12:00 to 14:00, the communication speed of the road will decrease obviously, which may be caused by people's work and activities in the afternoon after taking a nap at noon. From 14:00 to 17:00 in the afternoon, the road traffic was stable, but from

17:00 to 20:00 in the afternoon, the road traffic speed decreased significantly, and a typical evening peak appeared. From 20:00 to 24:00, the road traffic was gradually smooth and the traffic speed increased continuously.

The details are shown in the following figure:

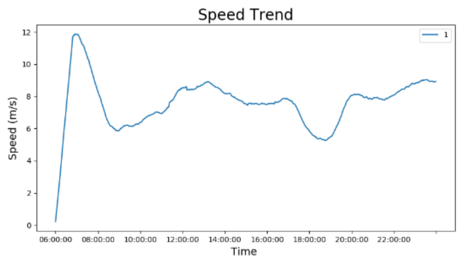

Fig.8 Monday

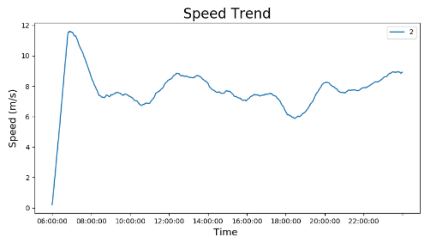

Fig. 9 Tuesday

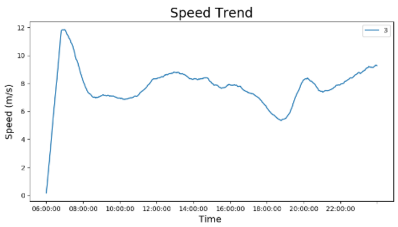

Fig. 10 Wednesday

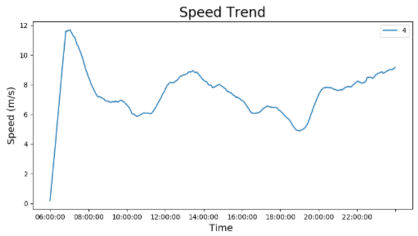

Fig.11 Thursday

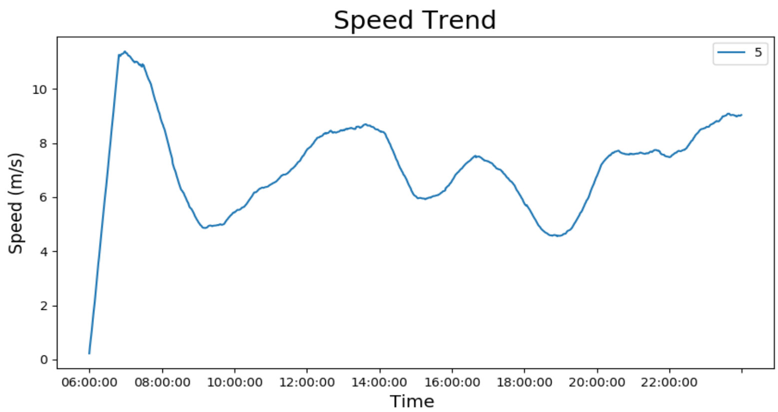

Fig. 12 Friday

As can be seen from the graph of time speed from Monday to Friday, the speed trend of each day of working day is similar with a certain rule, and there is an early peak and an evening peak in each day. In the comparison between the speed of the morning peak and the evening peak, it can be seen that the duration of the morning peak is slightly higher than that of the evening peak, and the speed of the evening peak is lower than that of the morning peak. Therefore, according to the speed of change trend chart, it can be seen working days crossing speed has a certain regularity and similarity, the speed of the weekend change quite gentle no typical peak early and late peak phenomenon, at the same time to 14:00 time 12:00 at noon, road traffic speed increase gradually, people may be due to the weekend at home with his family did not go out.

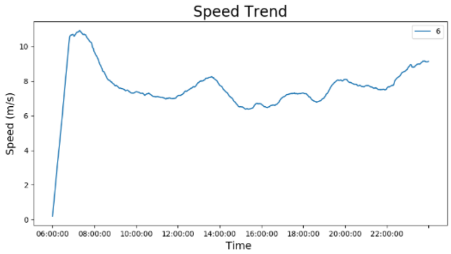

Fig. 13 Saturday

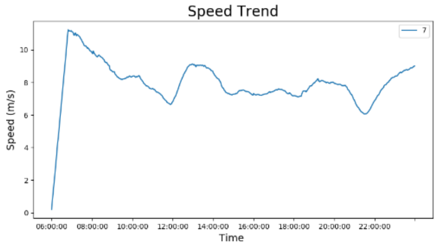

Fig.14 Sunday

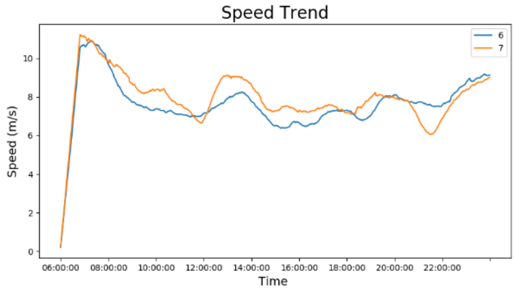

Fig. 15 Weekend

According to the velocity variation of Saturday and Sunday, the velocity variation trend of weekend is more gentle than that of weekday, and there is a certain difference. Through the analysis and comparison of the speed difference between weekdays and weekends in a week, it can be seen that there is a big difference between the speed change of weekdays and that of weekends. Therefore, modeling of weekdays and weekends can be considered for analysis and discussion respectively.

Through the analysis and comparison of the speed difference between weekdays and weekends in a week, it can be seen that there is a big difference between the speed change of weekdays and that of weekends. Therefore, modeling of weekdays and weekends can be considered for analysis and discussion respectively.

5.1 KNN-VA based velocity prediction model

5.1.1 model theory

The speed prediction model based on knn-va is mainly operated in two steps. First, KNN algorithm is used to screen samples of the original segmented data, select K groups of data that meet the requirements, and then calculate the historical average value of the selected data, and the result is the predicted value of the speed. The flow diagram of the algorithm is as follows:

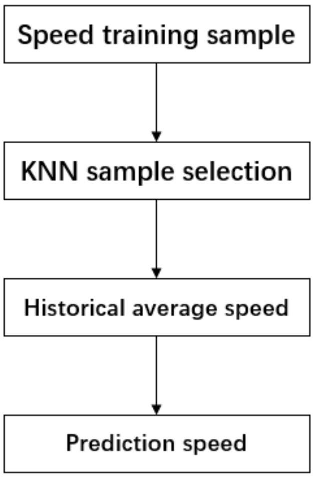

Fig. 16 KNN-VA model

Introduction of model

(1) select n samples from the training set samples and construct the training sample set V1 of KNN algorithm

$$V_1 = \{v_1(t), v_2(t), ..., v_n(t)\}$$

(2) The determination of the nearest neighbor number k in KNN algorithm. Firstly, an initial value is randomly assigned to k. In the process of iteration, the parameter k is continuously adjusted according to the result of KNN algorithm until the result reaches the optimal value.

(3) select the sample most similar to the test sample from the training sample, and use the Euclidean distance as the measurement standard. Where, v(t) is the velocity of the sample to be tested the day before, and x(t) is the velocity data of the training sample. The Euclidean distance between them can be defined as:

$$d_i = \sqrt{\sum_{i=1}^{n} (v(t) - x_i(t))^2} \quad i = 1, ..., n$$

(4)select k samples that are most similar to the prediction samples :

$$V_2 = \{v_1(t), v_2(t), ..., v_k(t)\}$$

(5) average the similar samples obtained in step (4), that is the predicted value Y of the velocity.

$$Y = \frac{1}{k} \sum_{i=1}^{k} v_i$$

5.1.2 Results and analysis

Prediction of working day: take the speed value on August 18, 2014 as the prediction target, and use knn-va model to predict the speed on August 18. The specific process is to use KNN algorithm to select data samples with high similarity between the 6-day test set and the selected data, and take the average of the data screened by KNN as the final predicted value.

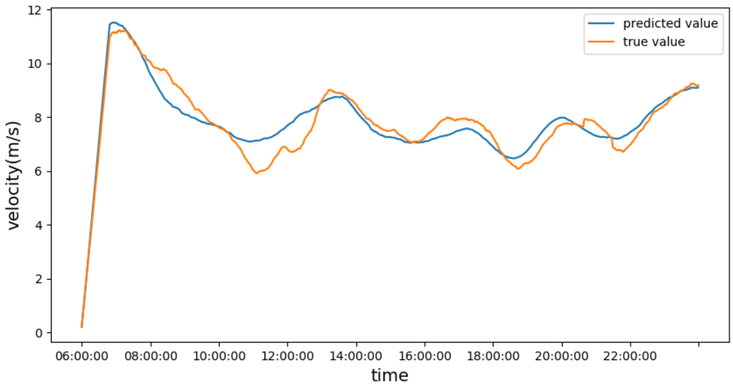

Fig.17 prediction results of KNN-VA model

As can be seen from the predicted results, the velocity trend between 6 and 8 points in the early stage was basically the same, and that between 22 and 24 points was also basically the same. The predicted velocity value in the middle is slightly different from the real value, and the root mean square error MSE of the model is 0.0146, mean absolute error 0.37187.

Prediction of weekend : take the weekend speed of August 23, 2014 as the prediction target, and use knn-va model to predict the speed of August 23. The specific process is to use KNN algorithm to select data samples with high similarity between the 3-day test set and the selected data, and take the average of the data screened by KNN as the final predicted value.

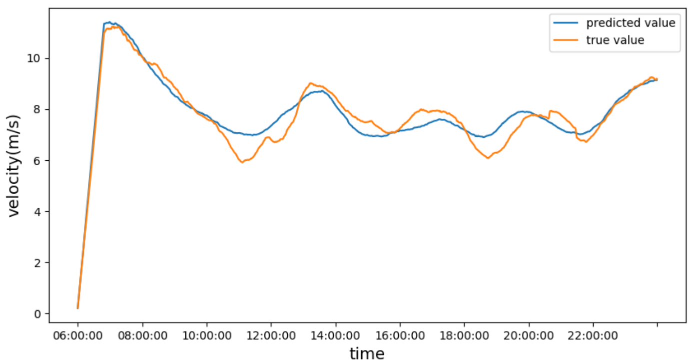

Fig.18 KNN-VA weekend model prediction results

As can be seen from the predicted results, the velocity trend between 6 and 8 points in the early stage was basically the same, and that between 22 and 24 points was also basically the same. The predicted speed in the middle is slightly different from the real value, in which the root-mean-square error mse is 0.013018 and the absolute error is 0.33823.

5.2 Velocity prediction model based on KNN-RBF

5.2.1 model theory

Based on K neighbor - radial basis neural network algorithm speed prediction, first on the speed of the data set, using KNN algorithm to filter samples, get qualified K set of samples, and then the K set of data training RBF neural network model, will stay prediction data input has been trained in the model, complete forecast for speed. The flow diagram of the algorithm is as follows:

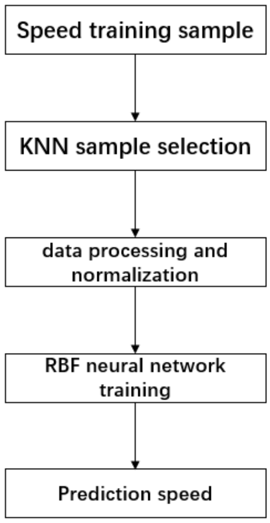

Fig. 19 KNN-RBF model

Introduction of model

(1) select n samples from the training set samples and construct the training sample set V1 of KNN algorithm

$$V_1 = \{v_1(t), v_2(t), ..., v_n(t)\}$$

(2) select k samples that are most similar to the prediction samples:

$$V_2 = \{v_1(t), v_2(t), ..., v_k(t)\}$$

(3) moving average filtering: perform arithmetic average operation of k velocity values using the first-in, first-out principle to obtain the filtered velocity value.

(4) data normalization

$$v_i = \frac{v - v_{\min}}{v_{\max} - v_{\min}}$$

(5) send the normalized training samples to radial basis neural network for training.

(6) use the trained radial basis neural network to predict the speed

5.2.2 Results and analysis

Prediction of working day speed: the speed on August 18, 2014 was selected as the predicted value, and knn-rbf model was used to predict the speed of the selected section on August 18. First, KNN algorithm was used to select the data of 6 days, and the selected speed data was filtered by moving average filtering. Then, RBF neural network was used for radial basis function optimization training, and the optimized model was trained. The date to be tested was input into the model to obtain the predicted speed value.

Moving average filtering speed:

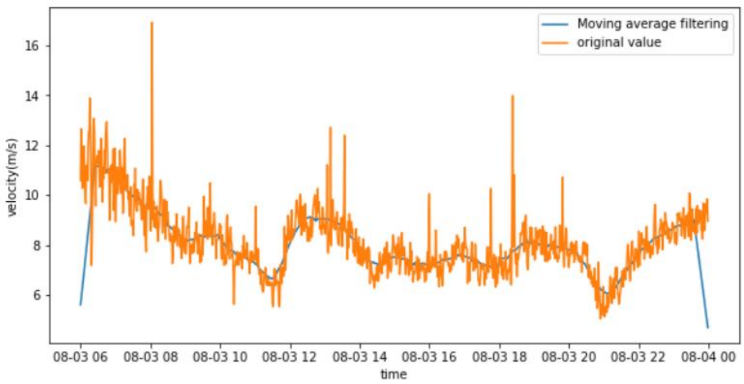

Fig 20 Average speed sliding filter

It can be seen from the moving average filtered velocity image that the filtered velocity curve is smoother. Some outliers have been removed, and the regularity of the resulting velocity curve is stronger.

The predicted and true values obtained after training are as follows:

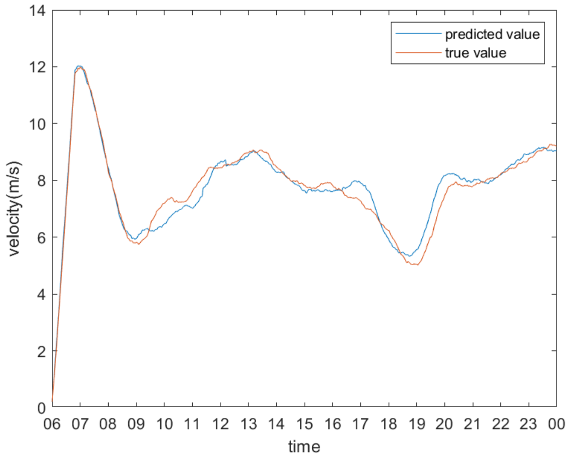

Fig. 21 KNN-RBF model prediction

As can be seen from the predicted results, the velocity trend from 6 to 9 in the early stage was basically the same, and the velocity trend from 22 to 24 was also basically the same. There was a deviation between the predicted speed value and the real value between 9 am and 10 PM. The calculated root mean square error MSE of the model was 0.000814153, which showed high accuracy of the model.

Prediction of weekend speed: the speed on August 23, 2014 was selected as the predicted value, and knn-rbf model was used to predict the speed of the selected section on August 23. The specific process is to use KNN algorithm to select the data of 6 days, continue the sliding average filtering for the selected speed data, and finally use RBF neural network to conduct radial basis function optimization training.

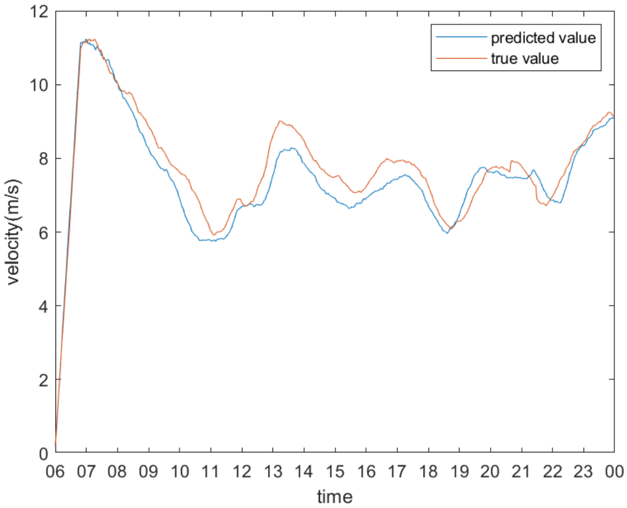

Fig 22 KNN-RBF weekend model prediction

As can be seen from the predicted results, the velocity trend between 6 and 7 points in the early stage was basically the same, and that between 22 and 24 points was also basically the same. From 6 am to 10 PM, the predicted value was slightly lower than the real value, but the error was relatively low. The calculated root mean square error MSE of the model was 0.00143321, indicating that the model has a strong practical value.

5.3 The comparison of KNN-VA and KNN-RBF model

Combined with the above prediction mean variance error MSE, the comparison of the two methods is made:

Table 2: Comparison of results

error	KNN-VA	KNN-RBF
Working days(8.18)	0.009	0.000814153
the weekend(8.23)	0.013018	0.00143321

By comparing the root mean square error of the two models, it can be seen that the mean square error of knn-rbf is less than knn-va and the error is less than 2%. The result is effective and reliable.

5.4 Determine the level of congestion

Traffic condition prediction results to determine the key lies in the different traffic status between the choice of the threshold speed, the model of highway traffic state in accordance with the provisions of the ministry of transportation grade division, road traffic state regulations be divided into five parts, respectively is marked as 1, 2, 3, 4, 5, in turn, on behalf of the flow, the basic flow, mild congestion, moderate congestion, severe congestion. Details are shown in the following table:

Table 3. Congestion classification

Traffic state	Velocity threshold	Congestion level
unimpeded	v ≥ 25m/s	1
bsically nimpeded	19.4m/s≤v≤25m/s	2
mild congestion	13.89m/s≤v≤19.4m/s	3
moderate congestion	8.3m/s≤v≤13.89m/s	4
severe congestion	0≤v≤8.3m/s	5

By comparing the speed prediction results at each moment with the speed threshold in the above table, the degree and grade of road congestion at this moment can be determined, thus providing theoretical support for the time prediction of subsequent congested sections.

5.4.1 Results and analysis

Through statistical analysis, we decided to choose on August 18, 2014, 12 to 18 points analysis, the time period as a result of our data sampling frequency is 1 min, if to traffic state estimation of each moment, sampling point will be very large, and the practical value is low, by reading lots of literature and practical experience of life, in the end we choose 15 min as the sampling interval, both to get enough sample points at the same time also has a strong application value, through the analysis of selected sample points, the results are as follows:

Table4: Congestion result

Time	True（m/s）	Rate	Predit（m/s）	Rate	Accuracy
1	8.45	4	8.65	4	Same
2	8.57	4	8.53	4	Same
3	8.64	4	8.56	4	Same
4	8.82	4	8.72	4	Same
5	8.98	4	8.89	4	Same
6	8.97	4	9.02	4	Same
7	9.03	4	8.77	4	Same
8	8.85	4	8.85	4	Same
9	8.48	4	8.28	5	dirrerent
10	8.23	5	8.30	4	dirrerent
11	8.00	5	8.00	5	Same
12	7.98	5	7.83	5	Same
13	7.82	5	7.64	5	Same
14	7.72	5	7.66	5	Same
15	7.78	5	7.66	5	Same
16	7.91	5	7.68	5	Same
17	7.82	5	7.60	5	Same
18	7.69	5	7.66	5	Same
19	7.46	5	7.75	5	Same
20	7.38	5	7.97	5	Same
21	7.27	5	7.92	5	Same
22	7.00	5	7.65	5	Same
23	6.83	5	7.13	5	Same
24	6.46	5	6.38	5	Same

According to table 4, we will predict the result of congestion level with the actual value of the congestion level, by looking at the last column of the table, it can be seen that used in this article, based on k nearest neighbour and the speed of RBF neural network combining with the real forecasting model of the predicted results are basically the same, after calculation, the model prediction accuracy reached 91.7%. From observation, we can see that the prediction results of the model are deviated at the 9th and 10th points. For the ninth point, the actual velocity was near the threshold of velocity, which caused the deviation of the crowding degree predicted by the model, but the accuracy of the velocity value predicted by the model reached 97.6%. Same thing for point 10. To sum up, it is reasonable and feasible to judge the traffic state by predicting the speed.

6 Time prediction

6.1 Algorithm thought

Through crowding in front to predict time and choose the sections, we chose to target at the peak period of road in the night, calculate the prediction by the time it takes for the entire section, the first of all, we through the road of latitude and longitude information, calculation of total length to the road of L (m), as a result of our speed sampling interval to 1 minute, so we assume that in a minute, car is moving at a constant speed, so we are calculated by means of discrete summation. For the last part of the journey, since it may not meet the uniform speed of a sub-station, we calculate the passage time by dividing the remaining journey by the speed. The specific calculation steps are shown below.

Step1. Calculate the length of the section as L(m) according to the longitude and latitude at both ends of the selected section.

Step2. Assume that the initial time after this section is t_init, and obtain the speed corresponding to this time point through the trained model.

Step3. Since the time sampling interval is 1 minute, the distance traveled in the interval of 1 minute can be calculated according to the speed corresponding to each time point, and the distance traveled s (m) can be calculated by using the discrete summation method.

Step4. When calculating the last segment of the road, because the current one-minute driving distance is larger than the length of the remaining segment, the difference between L and s is calculated as diff.

Step5. Use the following formula to calculate the time required to pass a section with length of diff t_diff. Define the velocity at the previous moment as v and the velocity at the current moment as v.

$$T\_diff = (L - s)/n$$

Step6. Define the starting time of the last section of the journey as t_last, and then the total length of the final journey through this section is t_sum.

$$T\_sum = t\_last - t\_init + t\_diff$$

6.2 Result

Taking the time 2014-08-2308:00:00 as an example, the coordinate latitude and longitude of the section passed were (30.6214, 104.094), (30.6195, 104.058), the whole length was 3451.31 meters, and the prediction time was 6.99 minutes.

7 Conclusion

According to the traffic state prediction model established by us, the traffic state prediction can be realized by relying on the vehicle speed in the process of driving. State of the model to select the most direct response vehicle speed in order to establish the prediction model parameters, implementation for predicting the future period of speed, the prediction model for processing good training samples by KNN algorithm to choose the highest test sample similarity of samples, to k neighbor sample data for training model, the speed of the prediction is a kind of using the data of k neighbor passing speed as the average of the speed of the next moment, one kind is by RBF neural network to k neighbor samples training to generate the speed prediction model. Two basic models have been established, and the improvement of the model starts with two basic models to predict the final speed. By comparing the predicted speed with the highway traffic status classification standard stipulated by the ministry of transportation, the congestion degree of vehicles can be estimated approximately.

8 Strengths and Weaknesses

8.1 Strengths

1. In many of the traffic state parameter selection of the most can reflect the speed of the vehicle information information to predict parameters, due to the speed of information acquisition is more convenient, so the universality of this method is stronger, at the same time due to the speed of road passage can objectively reactions, so use it as a road congestion degree evaluation factors, strong interpretability.

2. In this paper, using the speed as prediction parameters, KNN as sample, using two samples selected K forecast, two different model to forecast the velocity, the two models, one is based on the average of the past speed, one is based on RBF neural network prediction, this paper, from two different models were analyzed, and the reliability of the model prediction is higher.

3. Adopt the highway traffic status classification standard stipulated by the ministry of communications to classify the congestion degree of the road into five levels, so as to determine the congestion degree of the road, and select the period of the congestion of the road, which has a strong theoretical basis.

8.2 Weaknesses

1. The data in this paper selected a high-speed section that was approximately straight line. The establishment and prediction of the model had certain limitations, and the model was not verified under different road conditions, such as crossroad, crossroad and urban inner ring.

2. In the selection of traffic parameters, speed is selected as the only traffic state parameter for analysis, which makes it convenient to process information. At the same time, the robustness and generalization performance of the model are often poor, and different combinations of traffic parameters and comprehensive analysis and evaluation of different parameters are lacking.

3. The established model does not take into account the influence of some special circumstances, such as the occurrence of traffic accidents in abnormal weather.

9 Future

According to the current research results, although the prediction results of some time periods can meet the requirements, due to the randomness of event occurrence, the accuracy of data prediction cannot be guaranteed in the long run, and the following work still needs to be done in the future

(1) for the selection of prediction model, multiple prediction model fusion can be considered to improve the accuracy of the model, and a variety of random factors can be comprehensively considered to improve the generalization ability of the model. Besides sliding filtering method, other denoising methods can be tried to reduce the data information loss caused by filtering.

(2) in this paper, only knn-va and knn-rbf models were used to predict the velocity, and the prediction results within a week and at the weekend were obtained and compared. In addition to the above models, there must be more accurate models, which need to be explored continuously.

(3) in this paper, KNN- and knn-rbf models were used to predict the speed of a certain section of road, and the results were compared. The predicted value of the speed was used to calculate the time of the crowded section, which had certain credibility.

(4) compared with the neural network, the data collected in this paper is not enough to support a sufficiently accurate prediction model, and a large amount of data training is still needed to meet the requirements.

References

[1] Hu W, Wang H, Yan L. An actual urban traffic simulation model for predicting and avoiding traffic congestion[C]//IEEE, International Conference on Intelligent Transportation Systems. IEEE, 2014:2681-2686.

[2] Yigit H.A weighting approach for KNN classifier[C]//IEEE, International Conference on Electronics,Computer and Computation.IEEE,2013:228-231.

[3] Meng K,Dong Z Y,Wang D H.A self-adaptive RBF neural network classifier for transformer fault analysis[J].IEEE Transactions on Power Systems,2010,25(3):1350-1360.

[4] Zhao N, Yuan J B, Han X U. Survey on Intelligent Transportation System[J]. Computer Science,2014, 28(1):1-3.

[5] Jiang G Y,Jiang L H, Wang J F. Identification method of urban expressway traffic congestion [J]. Journal of transportation engineering,2006, 6 (3) : 87-91.

[6] Hu W, Wang H, Yan L. An actual urban traffic simulation model for predicting and avoiding traffic congestion[C]//IEEE, International Conference on Intelligent Transportation Systems. IEEE, 2014:2681-2686.

[7] Chowdhury D, Santen L, Schadschneider A. Statistical physics of vehicular traffic and some related systems[J]. Physics Reports, 2000, 329(4): 199-329.

[8] Ran h. ritten,Eibe Frank. Translated by dong Lin et al. Practical machine learning technology in data mining [M]. Beijing: China machine press, 2006:211-277

[9] Hassan T. Abdelwahab, Traffic Micro-simulation Model for Design and Operational Analysis of Barrier Toll Stations[J], Ain Shams Engineering Journal, 2016.

[10] Wang jiangfeng. Study on automatic identification of traffic congestion on expressway [D]. Jilin university, 2004.

[11] Ni xuanghua. Evaluation and prediction of traffic congestion based on data [D]. Zhejiang university of technology, 2014.

[12] Lin Ruihui, The Overall Planning and Design of Expressway Toll Plaza[J], Popular Science & Technology, 2006, 06: 21.

[13] Wang sheng. Traffic flow prediction based on artificial neural network [D]. Tianjin: nankai university, 2003.

[14] Zhuang bin, Yang xiaoguang, li keping. Discriminant criterion and detection algorithm of road traffic congestion [J]. Chinese journal of highways, 2006, 19(3): 82-86.

[15] Guan manquan. Analysis of causes and countermeasures of traffic congestion from road factors [J]. Journal of people's public security university of China, 2005, 5(3): 4-8.

[16] Hu qizhou, liu yingshun, guo tangyi. Spatial and temporal pattern recognition model for traffic congestion situation monitoring [J]. Transportation system engineering and information, 2012, 12(3): 41-45.

[17] Sun lifen. Research on decision support system for urban traffic congestion dredge [D]. Wuhan: huazhong university of science and technology, 2005.

[18] Jiang longhui. Research on short-term prediction method of urban expressway traffic parameters based on data fusion [D]. Jilin: jilin university, 2003.

[19] Wu bing, li linbo. Evolutionary dynamic analysis of traffic congestion [J]. Chinese journal of highways, 2006, 3(19): 106-110.

[20] Lin fei. Countermeasures and key technologies of traffic congestion in China's big cities [D]. Xi 'an: chang 'an university, 2006.