Short-term urban road congestion prediction considering temporal-spatial correlation

Liqi Deng[1,a],Liying Wei [2,b*]

[1]Bejing Jiaotong University, China
[2]Bejing Jiaotong University, China

Abstract. In order to address the gaps in the study of short-term urban road congestion prediction based on Baidu map real-time road condition data, a short-term prediction model for urban road congestion based on Pearson Correlation Coefficient (PCC) and Weighted Markov Chains (WMC) is constructed by combining historical temporal correlation of urban road congestion data with spatial correlation between road sections. The model use the PCC method to filter out the spatially significantly related road sections from the upstream and downstream sections of the target road section and add them to the target road section data set as the data input of the WMC prediction model to achieve the short-term prediction of urban road congestion. The performances of the proposed models are validated by using manually collecting real-time road condition data from Baidu map. The research results show that the model integrate the spatial and temporal correlations in the urban road congestion data. Compared with other three prediction models, the prediction accuracy of the proposed model is improved by 3.096% on average, and the prediction error is reduced by 0.135 on average.

1 Introduction

Baidu Map Smart Traffic uses the City Road Congestion Index as an objective indicator of traffic congestion on urban roads, with a larger congestion index representing a higher level of congestion. According to the size of the congestion index, Baidu map classifies the congestion level of city roads into four states: smooth, slow, congested and severely congested. With Baidu map's real-time road conditions feature, we can understand the congestion status of each road section in real time so that we can plan our routes. However, when we need to plan our route some time in advance, Baidu map's real-time road condition function is useless, while its road condition prediction function can only predict the congestion status at one-hour intervals, which obviously does not meet our needs.

There is a gap in the current research on short-term prediction of urban road congestion based on Baidu maps. For the urban road congestion, which is a discrete time series in both state and time, this paper is inspired by Xie Kaibing use autocorrelation coefficients to calculate the weights of Markov chains of different orders to form weighted Markov chains for state prediction of traffic volume after division into different states[1] and Shao Chunfu et al use PCC algorithm to form a temporal-spatial combination model for traffic volume prediction[2], using PCC algorithm to explain the spatial correlation between the target road sections and the upstream and downstream road sections, and using autocorrelation coefficients to consider the correlation between different time steps, the PCC-WMC prediction model that considers both temporal and spatial correlation is proposed for urban road congestion prediction.

The Markov chain process is characterised by the fact that each transfer of states is only related to the previous state of the interconnection and is independent of the past states[3] . At present, Markov chains are more often used in the field of traffic flow prediction in the direction of transportation. Chen Shuyan et al proposed a prediction method combining wavelet analysis and Markov magic, which improved the anti-interference of the prediction model[4]; Jiang Lizhong et al proposed to combine Grey model and Markov chain model to form Grey Markov chain model for traffic flow prediction, and the experiment showed that the model has certain feasibility[5]; Li Junhuai et al proposed to combine exponential smoothing theory and Markov chain, and the experiment showed that the combined model is more accurate than the single model[6]; Xu Mengmeng et al improved the prediction accuracy of the model by correcting the residuals on the basis of the Grey Markov model[7]; Shen Qidong et al introduced traffic flow prediction influence factors from time and space to achieve traffic flow prediction of intersections[8].

* Corresponding author: [b]lywei@bjtu.edu.cn
[a]20120793@bjtu.edu.cn

2 A model for predicting urban road congestion based on temporal-spatial characteristics

2.1 Spatial feature mining

Spatial feature mining refers to considering the influence of congestion of upstream and downstream road sections on the target road section, using the inter-section correlation degree to characterize the mutual influence between the two road sections. In this paper, PCC is used to explain the correlation degree between the target road section and the upstream and downstream road sections, assuming that the time series of the target road section A is $X^A=[X_1^A,X_2^A,\cdots,X_T^A]$, and the time series of the upstream and downstream road sections B is $X^B=[X_1^B,X_2^B,\cdots,X_T^B]$, then the correlation coefficient between the two road sections is

$$r_{AB}=\frac{\sum_{i=1}^{T}(X_i^A-\overline{X^A})(X_i^B-\overline{X^B})}{\sqrt{\sum_{i=1}^{T}(X_i^A-\overline{X^A})^2}\sqrt{\sum_{i=1}^{T}(X_i^B-\overline{X^B})^2}} \tag{1}$$

Where r_{AB} is the correlation between two road sections A, B; T is the length of the time series. When r_{AB} tends to 1, it means that the correlation between the target road section and the upstream and downstream road sections is greater; when r_{AB} tends to 0, it means that the correlation between the target road section and the upstream and downstream road sections is smaller. In order to ensure the accuracy of prediction, $0.5\le r_{AB}\le 1$ is taken as the range of spatial correlation coefficient in this paper.

It should be noted that the range of values for the spatial correlation coefficient is derived from a significance test for PCC, i.e. a range of PCC values that would indicate a significant correlation between the two road segments for a given sample size. The t-test[9] is used here, with the known statistic t

$$t=r\times\sqrt{\frac{d-2}{1-r^2}} \tag{2}$$

obey the t-distribution with degrees of freedom at $(d-2)$.

where r is the calculated correlation coefficient; d is the sample size, i.e. the number of days of data counting.

At a defined level of significance α , if the statistic t satisfies

$$t>t_\alpha(d-2) \tag{3}$$

then it is considered that there is a significant correlation between the two road segments.

It is known that $d=21$,the significance level α takes 0.05, when $r=0.5$, $t\approx 2.517>t_{0.05}(21-2)=2.093$, and $t\approx 1.902<t_{0.05}(21-2)=2.093$ when $r=0.4$. Therefore when $0.5\le r_{AB}\le 1$, it indicates a significant spatial correlation between the two road segments.

2.2 Temporal feature mining

The congestion of urban roads in the target time period is not only directly related to the congestion of the previous time period, but also indirectly related to several previous time periods. In order to solve this problem, this paper adopts a weighted Markov chain to extract the temporal characteristics of the data, by constructing multiple Markov chains of different orders, using the autocorrelation coefficients of each order to obtain the weights of each Markov chain, and then using the weighted summation to calculate the prediction value of the congestion state with the highest probability.

2.2.1 Markov property test

Markov property means that past states and future states are conditionally independent if the current state is known. To apply Markov chain models to analyse and solve practical problems, it is necessary to first test whether the time series has Markov property. In fact, the test of Markov property is a test of whether the past and present states are conditionally independent, or a test of the correlation between the two. For discrete sequences, the χ^2 test[10] is a common method used for correlation tests. Thus, the χ^2 test can be used to test the Markov property of a time series.

Also assume that the target road section A time series is $X^A=[X_1^A,X_2^A,\cdots,X_T^A]$, which includes four states: smooth, slow, congested and heavily congested, corresponding to state 1, state 2, state 3 and state 4 respectively. Now use s_{ij}^* to denote the frequency of the time series in time length T from state i to state j after one step, thus obtaining the frequency matrix S_A^* of the one-step state shift of the target road section A in time length T. The value obtained by dividing row i of the matrix S_A^* by the sum of rows i is the one-step state transfer probability, which is denoted as p_{ij}^* , and its matrix is denoted as P_A^*. The value obtained by dividing the sum of column j of the matrix S_A^* by the sum of all columns of the matrix is the marginal probability, which is denoted as p_j^* , i.e.

$$p_{ij}^*=\frac{s_{ij}^*}{\sum_{j=1}^{4}s_{ij}^*} \tag{4}$$

$$p_j^* = \frac{\sum_{i=1}^{4} s_{ij}^*}{\sum_{i=1}^{4}\sum_{j=1}^{4} s_{ij}^*} \quad (5)$$

When the amount of data is sufficient, χ^2 statistics

$$\chi^2 = 2\sum_{i=1}^{4}\sum_{j=1}^{4} s_{ij}^* \left| \ln \frac{p_{ij}^*}{p_j^*} \right| \quad (6)$$

obey the χ^2 distribution with the degrees of freedom $(4-1)^2$.

At a defined level of significance α , if the statistic χ^2 satisfies

$$\chi^2 > \chi_\alpha^2((4-1)^2) \quad (7)$$

then the target road section A time series is considered to have Markov property.

2.2.2 Construction of a multi-step state transfer probability matrix

The construction of a multi-step state transfer probability matrix is an important prerequisite for the application of weighted Markov chains for prediction. After collecting some of the historical data, the overall state transfer probability matrix can be estimated from this sample data, as follows.

Firstly, assume that the time series of the target road section A is $X^A = [X_1^A, X_2^A, \cdots, X_T^A]$, which consists of four states: smooth, slow, congested and heavily congested, corresponding to state 1, state 2, state 3 and state 4 respectively.

Then use $s_{ij}^{t\,(n)}$ to denote the frequency of the time series at time t from state i to state j after n steps, to obtain the n-step state transfer frequency matrix $S_t^{A(n)}$ of the target section A at time t. The value obtained by dividing row i of the matrix $S_t^{A(n)}$ by the sum of rows i is the n-step state transfer probability, which is denoted as $p_{ij}^{t\,(n)}$, thus obtaining the n-step state transfer probability matrix of the target section A at time t, which is denoted as $P_t^{A(n)}$.

$$p_{ij}^{t\,(n)} = \frac{s_{ij}^{t\,(n)}}{\sum_{j=1}^{4} s_{ij}^{t\,(n)}} \quad (8)$$

$$P_t^{A(n)} = \begin{bmatrix} p_{11}^{t\,(n)} & p_{12}^{t\,(n)} & p_{13}^{t\,(n)} & p_{14}^{t\,(n)} \\ p_{21}^{t\,(n)} & p_{22}^{t\,(n)} & p_{23}^{t\,(n)} & p_{24}^{t\,(n)} \\ p_{31}^{t\,(n)} & p_{32}^{t\,(n)} & p_{33}^{t\,(n)} & p_{34}^{t\,(n)} \\ p_{41}^{t\,(n)} & p_{42}^{t\,(n)} & p_{43}^{t\,(n)} & p_{44}^{t\,(n)} \end{bmatrix} \quad (9)$$

The final n-step transfer probability matrix for the target section A is obtained at

$$P^{A(n)} = [P_1^{A(n)}, P_2^{A(n)}, \cdots, P_T^{A(n)}].$$

It is important to note that the state transfer frequency matrix and the state transfer probability matrix here are not the same concept as the matrix used in Section 1.2.1.

2.2.3 Calculation of weights

1) autocorrelation coefficient of each order

Assume that the target road section A time series is $X^A = [X_1^A, X_2^A, \cdots, X_T^A]$ and it consists of four states: smooth, slow, congested and heavily congested, corresponding to state 1, state 2, state 3 and state 4 respectively. The nth order autocorrelation coefficient $r^{A(n)}$ is calculated as follows.

$$r^{A(n)} = \frac{\sum_{t=1}^{T-n} (X_t^A - \overline{X^A})(X_{t+n}^A - \overline{X^A})}{\sum_{t=1}^{T-n} (X_t^A - \overline{X^A})^2} \quad (10)$$

Where X_t^A is the urban road congestion at time t of target road section A; $\overline{X^A}$ is the mean value of the time series X^A of target road section A; T denotes the length of the time series.

2) standardisation

Once the order autocorrelation coefficients have been calculated, they are standardized. The sum of the standardized autocorrelation coefficients is 1, which is then used as the weight of the Markov chain of the corresponding order. The standardisation formula for the autocorrelation coefficients of each order for the target section A is as follows.

$$\omega^{A(n)} = \frac{\left| r^{A(n)} \right|}{\sum_{n=1}^{N} \left| r^{A(n)} \right|} \quad (11)$$

Where $\omega^{A(n)}$ is the weight of the nth-order Markov chain for the target section A; N is the number of different order Markov chains.

2.2.4 Prediction module

Assume that in a Markov chain of order n, the urban road congestion of the target road section A at time t is state 1 or state 2 or state 3 or state 4, and its corresponding state vectors are [1,0,0,0] , [0,1,0,0] , [0,0,1,0], [0,0,0,1], which are denoted as $SV_t^{A(n)}$, then the predicted probability vector $PPV_t^{A(n)}$ of the target road section A at time t obtained from the Markov chain of order n is calculated as follows

$$PPV_t^{A(n)} = SV_t^{A(n)} \times P_t^{A(n)} \quad (12)$$

In Section 1.2.3, the weights of each order Markov chain $\omega^{A(n)}$ have been obtained, so the weighted predictive probability vector $WPPV_t^A$ for the target section A at moment t is calculated as follows

$$WPPV_t^A = \sum_{n=1}^{N} \omega^{A(n)} \times PPV_t^{A(n)} \tag{13}$$

By finding the state corresponding to the maximum probability in the weighted prediction probability vector, the predicted state value for the next moment can be obtained.

It is important to note that since this paper is to predict the urban road congestion for one continuous hour at five minute intervals after the current moment. Therefore, after predicting the next moment state value, it is necessary to continue to predict this predicted state value as if it were a known state.

2.3 Temporal-spatial feature prediction model

Assuming that there are m road sections in the target road network, the target road network time series matrix X can be defined as

$$X = \begin{bmatrix} X^1 \\ X^2 \\ \vdots \\ X^{m-1} \\ X^m \end{bmatrix} = \begin{bmatrix} X_1^1 & X_2^1 & \cdots & X_{T-1}^1 & X_T^1 \\ X_1^2 & X_2^2 & \cdots & X_{T-1}^2 & X_T^2 \\ \vdots & \vdots & & \vdots & \vdots \\ X_1^{m-1} & X_2^{m-1} & \cdots & X_{T-1}^{m-1} & X_T^{m-1} \\ X_1^m & X_2^m & \cdots & X_{T-1}^m & X_T^m \end{bmatrix} \tag{14}$$

Where X^m is the time series of road section m; X_t^m is the congestion state of urban roads at time t for road section m.

The congestion of urban roads at the next moment of the target road section is mainly related to the congestion of the previous moments and the significantly correlated upstream and downstream road sections. Therefore, in this paper, for the target road network time series matrix X, the PCC algorithm is used to find out the road sections with high correlation with the target road sections among the upstream and downstream road sections, and then reconstruct them into a new data set, and finally the reconstructed data set is used as the input data of the weighted Markov chain prediction model for prediction, and the specific operation of reconstructing the data here is to merge the target road section time series with the upstream and downstream significantly correlated road section time series. The model calculation process is divided into the following 9 steps.

Step 1 Build the spatial correlation coefficient matrix: use equation (1) to calculate the correlation coefficient between the two road segments and build the spatial correlation coefficient matrix according to the following equation.

$$r = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_{m-1} \\ r_m \end{bmatrix} = \begin{bmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,m-1} & r_{1,m} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,m-1} & r_{2,m} \\ \vdots & \vdots & & \vdots & \vdots \\ r_{m-1,1} & r_{m-1,2} & \cdots & r_{m-1,m-1} & r_{m-1,m} \\ r_{m,1} & r_{m,2} & \cdots & r_{m,m-1} & r_{m,m} \end{bmatrix} \tag{15}$$

Step 2 Spatial correlation coefficient ranking: the upstream and downstream sections of the target section are ranked according to their spatial correlation coefficients, from largest to smallest.

Step 3 Reconstruct the temporal-spatial feature matrix: filter out the road sections with correlation coefficients greater than 0.9, 0.8, 0.7, 0.6 and 0.5 respectively, and reconstruct them into a new data set.

Step 4 Markov property test: use equations (4), (5), (6) and (7) to test whether the time series has Markov property.

Step 5 Construct the multi-step state transfer probability matrix: count the frequency matrix of state transfers for each step of the new data set and use equations (8) and (9) to obtain the multi-step state transfer probability matrix.

Step 6 Calculation of weights: each order of autocorrelation coefficient of the target road section is calculated using equation (10) and standardized using equation (11) to obtain each order of predicted probability vector weights.

Step 7 Urban road congestion prediction: a weighted Markov chain prediction model is constructed to predict the congestion of urban roads.

Step 8 Calculate the accuracy and the Root Mean Square Error (RMSE) of the model.

Step 9 Prediction results and error analysis.

3 Case design

In order to verify the feasibility of the PCC-WMC model, this paper manually collected 96 road sections (in both directions) from the second ring to the third ring in part of the Haidian District of Beijing through the real-time road condition function of Baidu map, and the congestion of urban roads from 16:00-19:00 with 5min interval for 22 working days, a total of 36×96×22 data, and the experimental road network is shown in Figure 1 and 2. The data of the first 21d were used as the training set and the data of the last day were used as the test set to predict the congestion of urban roads at 5min intervals from 17:05 to 18:05 on the last day.

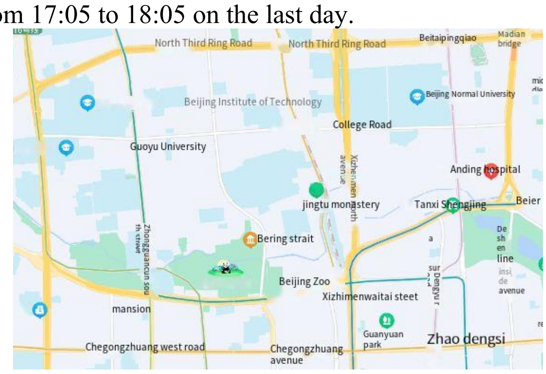

Fig. 1. Road network for verification

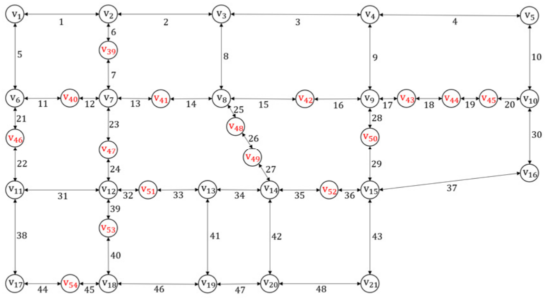

Fig. 2. Topological structure of road network

In order to quantitatively analyze the prediction results of the model, the accuracy and Root Mean Square Error (RMSE) are used as the evaluation indicators of the prediction results in this paper, and each indicator is calculated by the following formula

$$accuracy = \frac{l}{L} \times 100 \quad (16)$$

$$RMSE = \sqrt{\frac{1}{L}\sum_{i=1}^{L}(p_i - x_i)^2} \quad (17)$$

where l is the number of predicted states that are the same as the actual states in the desired prediction period; L is the overall amount of predicted data required; p_i is the predicted state value; x_i is the actual state value.

Accuracy is used to directly evaluate the correctness of the state prediction, the higher the value the better the prediction; RMSE is used to consider the size of the specific prediction error, the smaller the value the better the prediction.

4 Analysis of results

The model calculation process presented in Section 1.3 was followed to predict the congestion status of urban roads for each of the 96 road sections within the experimental road network.

Firstly, the data from the test set of section 1 is used as an example to verify whether it has Markov property. Using equations (4) and (5) in Section 1.2.1 to find the state transfer frequency matrix S_1^* and the state transfer probability matrix P_1^*

$$S_1^* = \begin{bmatrix} 17 & 1 & 0 & 0 \\ 1 & 8 & 4 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$P_1^* = \begin{bmatrix} \frac{17}{18} & \frac{1}{18} & 0 & 0 \\ \frac{1}{13} & \frac{8}{13} & \frac{4}{13} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

then from equation (6) to find $\chi^2 \approx 13.787$, check the table can be found when the significance level $\alpha = 0.05$, the statistics χ^2

$$\chi^2 \approx 13.787 > \chi^2_{0.05}(3-1)^2 = 9.488$$

therefore, the section 1 test set time series satisfies Markov property. The reason that the degrees of freedom are taken here as $(3-1)^2$ instead of $(4-1)^2$ is that the actual states in the road section 1 time series contain only three states, state 1, state 2 and state 3.

The prediction of urban road congestion is then carried out. In order to study the influence of temporal-spatial variables on the prediction results, the spatial correlation coefficient thresholds are set to 0.5, 0.6, 0.7, 0.8, 0.9 and 1 in this paper; the time steps, i.e. the number of Markov chains of different orders, are set to 1, 5, 6, 7, 8, 9 and 10. The prediction performance evaluation of the PCC-WMC prediction model under different values of temporal-spatial variables is shown in Tables 1 and 2.

In order to improve the prediction accuracy of the model, this paper uses the PCC algorithm for spatial feature mining. From the evaluation results, it can be seen that: when the time step is small (1, 5, 6), as the value of the spatial threshold value decreases, the accuracy continues to decrease and RMSE continues to increase, adding the PCC algorithm at this time instead reduces the prediction accuracy; when the time step increases to a certain number (7, 8, 9, 10). As the value of the spatial threshold is taken to decrease, the accuracy first increases to reach a peak and then gradually decreases, and RMSE first decreases and then gradually increases. In other words, when the time information used for prediction is too small, adding spatial information will not have an optimising effect and may even have a negative effect; when there is enough time information, adding the right amount of spatial information can have an optimising effect to a certain extent, but too much spatial information will also interfere with the prediction results.

Further analysis of the effect of time step on the prediction accuracy shows that under the same spatial threshold value, the overall trend of the accuracy increases and then gradually decreases as the time step increases, while the overall trend of RMSE decreases and then increases. It can be seen that, within a certain range, the model prediction accuracy increases with increasing time step. When the prediction uses too little time information, the prediction is poor; while when the time step is too long, irrelevant time information will be added, resulting in poor prediction performance.

Table 1. Accuracy under different temporal-spatial variables

accuracy		Spatial threshold					
		1	0.9	0.8	0.7	0.6	0.5
Time step	1	84.896	83.160	81.250	80.903	80.642	80.208
	5	84.983	84.722	83.507	82.813	82.813	82.552
	6	85.330	84.722	83.941	83.247	83.333	83.073
	7	85.330	86.285	84.983	84.288	84.375	84.028
	8	85.417	87.587	85.417	84.635	84.722	84.375
	9	85.417	87.587	85.503	84.722	84.896	84.549
	10	84.896	83.160	81.250	80.903	80.642	80.208

Table 2. RMSE under different temporal-spatial variables

RMSE		Spatial threshold					
		1	0.9	0.8	0.7	0.6	0.5
Time step	1	0.526	0.537	0.576	0.594	0.607	0.611
	5	0.529	0.534	0.604	0.620	0.620	0.622
	6	0.513	0.534	0.554	0.571	0.571	0.573
	7	0.516	0.424	0.515	0.534	0.533	0.536
	8	0.510	0.393	0.494	0.524	0.523	0.526
	9	0.510	0.390	0.490	0.533	0.531	0.534
	10	0.510	0.396	0.490	0.538	0.537	0.540

This trend is not obvious in Table 1 and 2, and is visualised in Figure 3 and 4. As can be seen from the graphs, the highest accuracy and the smallest RMSE, i.e. the best overall prediction, is achieved when the time step is 9 and the spatial threshold is 0.9.

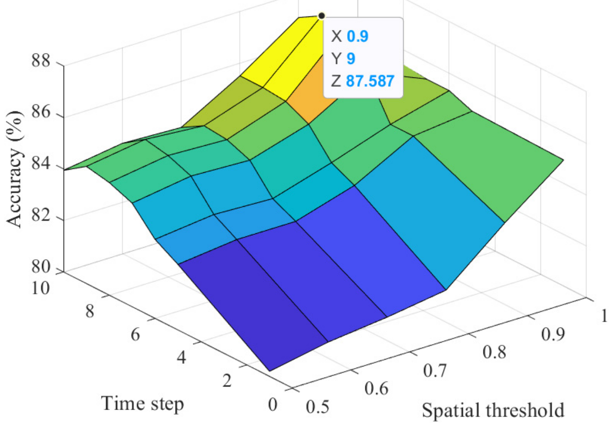

Fig. 3. Trends in accuracy

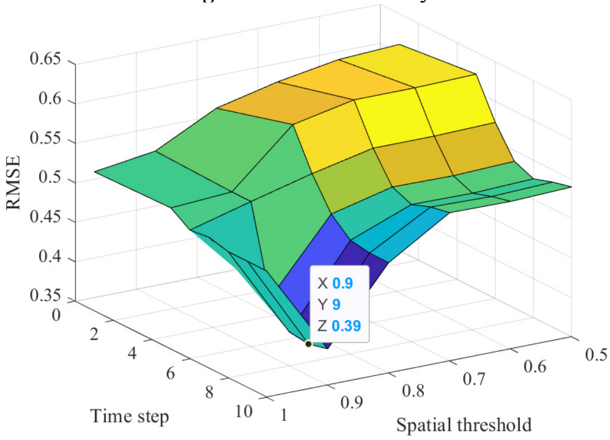

Fig. 4. Trends in RMSE

In order to compare and investigate the prediction performance of the PCC-WMC prediction model for each time period in short-term prediction, the first-order Markov chain prediction model (MC_1), the first-order Markov chain prediction model considering temporal-spatial correlation (PCC-MC_1 with the spatial threshold set to 0.9), and the weighted Markov chain prediction model (WMC with the time step set to 9) were used to predict the congestion of urban roads on each segment of the experimental road network. The prediction performance of the different models for each time period is shown in Figure 5 and 6. It can be seen that the PCC-WMC prediction model does not show superior prediction performance compared to the other three models in short time prediction, but when the prediction time increases and the prediction performance of the other three models decreases, the PCC-WMC prediction model maintains a more stable prediction performance, which makes it more advantageous in short-term prediction. Comparing the PCC-WMC and WMC prediction models, we can see that the addition of spatial information does not have a positive effect on short-term prediction, but as the prediction time increases, it gradually improves the prediction performance of the model. Overall, the PCC-WMC prediction model improved the accuracy by an average of 3.096% and reduced RMSE by an average of 0.135 compared to the other three models.

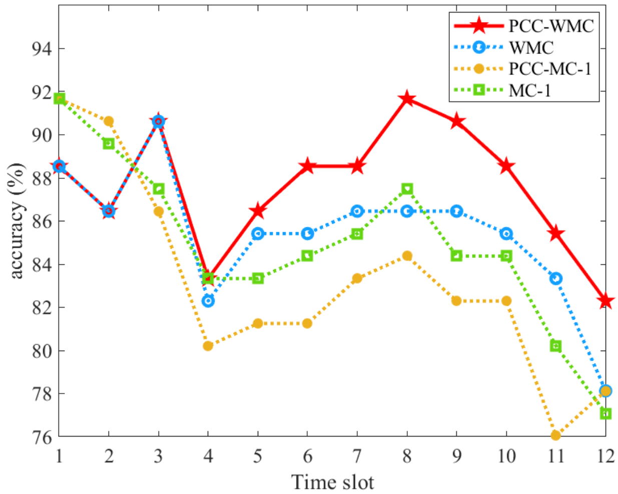

Fig. 5. Trends in accuracy by time slot of different models

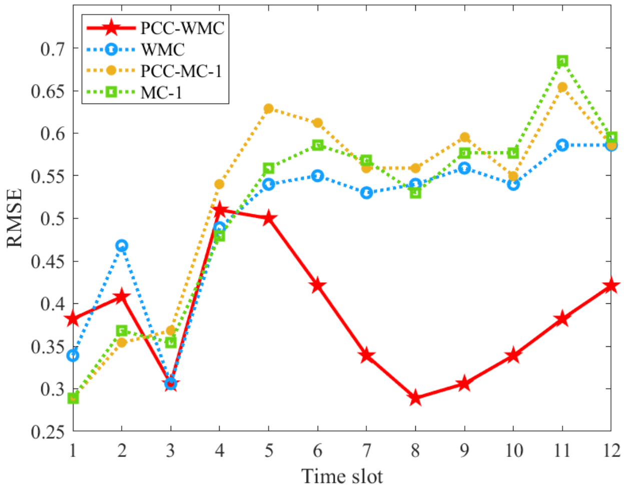

Fig. 6. Trends in RMSE by time slot of different models

5 Conclusion

1) This paper uses the PCC algorithm for spatial feature mining, which is simple to implement and the core idea is to characterize the degree of spatial correlation between road segments by dividing the covariance by the product of the standard deviation of the two time series. WMC is based on MC and uses a combination of the weights of each order obtained by using the

autocorrelation coefficient for each order of Markov chain, effectively using more time information.

2) This paper investigates the effect of time step size on prediction accuracy, which increases with increasing time step size within a certain range. When the time step size is small and there is too little time information, the prediction effect is poor; while when the time step size used for prediction is too long, the added irrelevant time information will interfere with the prediction results. The effect of the spatial threshold on the prediction accuracy was also investigated. When the time step was small, the prediction accuracy continued to decrease as the spatial threshold decreased, i.e. when there was less temporal information, the added spatial information would interfere with the prediction results; when the time step was large enough, the prediction accuracy increased and then decreased as the spatial threshold decreased, i.e. when there was a lack of spatial information, the prediction results were poor, and the prediction is equally disturbed when too much irrelevant spatial information is added. Therefore, setting the values of temporal and spatial variables reasonably can improve the prediction accuracy.

3) In this paper, we studied the trend of prediction performance of different models in short-term prediction by time period. Overall, the prediction performance of all four models tends to decrease with increasing predict length, but the prediction performance of PCC-WMC prediction model is more stable, which makes it the best in short-term prediction. The superiority of the PCC-WMC model is not obvious for short time prediction, but when the predict length is long enough, its superiority gradually becomes obvious.

4) The PCC-WMC prediction model proposed in this paper can provide a new idea for the research and application of urban road congestion prediction based on Baidu map. In the subsequent research, the prediction performance can be improved by trying to use appropriate time steps and spatial thresholds for each road segment in the road network individually, so that the temporal-spatial information used in the prediction matches each road segment one by one. In addition, to address the problem that the prediction performance of the model is not good enough for short time prediction, a combination of models can be considered for separate prediction in different time periods to further improve the prediction performance.

References

1. Chinese National Knowledge Infrastructure. (2017) Long-term prediction of traffic volume based on combined model of Markov chains. https://kns.cnki.net/kcms/detail/detail.aspx?FileName=1017060491.nh&DbName=CMFD2017.
2. Shao, C.F., Xue, S., Dong, C.J. (2021) Short-term traffic flow prediction considering temporal-spatial correlation on road network. J. Journal of Beijing Jiaotong University, 45(04): 37-43. https://kns.cnki.net/kcms/detail/11.5258.U.20210409.1020.002.html.
3. Zheng, J.H., Lin, X.H., Zheng, L.X., Zhang, S.H. (2012) Traffic congestion state prediction based on Markov chain model. J. Traffic Standardization, 281: 76-79. https://www.cnki.net/kcms/doi/10.16503/j.cnki.2095-9931.2012.22.027.html.
4. Chen, S.Y., Wang, W., Qu, F.G. (2005) Short-term traffic flow time series forecasting based on wavelet analyses-fuzzy Markov prediction model. J. Journal of Southeast University (Natural Science Edition), (04): 637-640. https://kns.cnki.net/kcms/detail/detail.aspx?FileName=DNDX200504030&DbName=CJFQ2005.
5. Jiang, L.Z., Li, C.D., Tang, B.,Zhu, H.B. (2010) Prediction of traffic volume in bridge random load based on Grey Markov chain. J. Road Traffic Science and Technology, 27(07): 128-132. https://kns.cnki.net/kcms/detail/detail.aspx?FileName=GLJK201007026&DbName=CJFQ2010.
6. Li, J.H., Gao, Z.,Wang, Z.X., Zhang, J. (2013) Short-term traffic flow forecasting based on exponential smoothing and Markov chains. J. Computer System Applications, 22(12): 132-135. https://kns.cnki.net/kcms/detail/detail.aspx?FileName=XTYY201312025&DbName=CJFQ2013.
7. Xu, M.M., Yan, L. (2017) Forecast of free traffic volume based on Gray Markov's Holiday Highway. J. Logistics Science and Technology, 40(06): 101-105. https://www.cnki.net/kcms/doi/10.13714/j.cnki.1002-3100.2017.06.029.html.
8. Shen, Q.D., Li, W.Y., Lian, G. (2021) Traffic flow prediction method based on improved Markov chain. J. Heilongjiang Transportation Science and Technology, 44(10): 178-181. https://www.cnki.net/kcms/doi/10.16402/j.cnki.issn1008-3383.2021.10.092.html.
9. Tang, Y.N., Zhou, B.CH., Xie, W.X. (2022) Equivalence analysis of three significance tests of univariate linear regression equation. J. Higher Mathematics Research, 25(06): 76-78. https://kns.cnki.net/kcms/detail/detail.aspx?FileName=XUSJ202206027&DbName=CJFQ2022.
10. Zhang, Y.F., Zhu, Y.L. (2012) The study of Markov property and its test. J. Value Engineering, 31(02): 312-313. https://www.cnki.net/kcms/doi/10.14018/j.cnki.cn13-1085/n.2012.02.006.html.