IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ARTIFICIAL INTELLIGENCE (AI)-EMPOWERED INTELLIGENT TRANSPORTATION SYSTEMS

Received November 24, 2019, accepted December 9, 2019, date of publication December 12, 2019, date of current version January 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2959125

Discrimination and Prediction of Traffic Congestion States of Urban Road Network Based on Spatio-Temporal Correlation

ZHI CHEN[1], YUAN JIANG[1], AND DEHUI SUN[1], (Member, IEEE)
[1]Beijing Key Lab of Urban Intelligent Traffic Control Technology, North China University of Technology, Beijing 100144, China
Corresponding author: Zhi Chen (chenzhi19860301@126.com)

This work was supported in part by the Beijing Natural Science Foundation under Grant 8172018.

ABSTRACT By analyzing and predicting the traffic states of urban road network, the formation of traffic congestion can be effectively alleviated, so as to improve the traffic capacity of urban road network. In this paper, firstly, we analyze and study the spatio-temporal correlation characteristics of traffic states based on the existing floating car data. At the same time, we extend the traffic conditions of urban road network from the upstream and downstream interaction to the global road network and complete the traffic congestion states discrimination of urban road network based on the spatio-temporal correlation. Secondly, according to the traffic jam aggregation and diffusion characteristics of local Moran's I, a mixed forest prediction method considering the spatio-temporal correlation characteristics of urban road traffic state is constructed by improving the existing random forest algorithm. Finally, an example is given to verify the effect of the prediction method on the short-term prediction of urban road network traffic states.

INDEX TERMS Traffic congestion, spatio-temporal correlation, local Moran's I, short-term prediction.

I. INTRODUCTION

At present, the road traffic congestion recognition algorithm was divided into two stages. In the early stage, the basic parameters such as speed, flow and occupancy rate obtained by traditional traffic point detection equipment were used to identify and judge the traffic congestion states through various algorithms or rules [1]. In fact, the study of traffic flow state prediction was an extension of traffic congestion states discrimination.

Ning Z et al. [2] constructed a method of minimizing cost offloading, proposed a two-way matching algorithm to adjust frequency spectrum and meet user delay constraints, combined with deep reinforcement learning method to optimize system state and realize distributed computation. Ning Z et al. [3] obtained the closely related social characteristics of vehicles by constructing the triangle relation structure chart, estimated the node connection probability in the network model by combining the characteristics of vehicles and associated equipment, and established the CSPD algorithm based on convolution neural network. Ning Z et al. [4] built an intelligent system framework for vehicle edge computing based on deep reinforcement learning technology, established a communication and computing state model based on Finite Markov Chain, combined with two-way matching and deep reinforcement learning methods, jointly optimized task scheduling and network resource allocation strategies to maximize the quality of user experience (QoE). Ning Z et al. [5] constructed an energy-saving scheduling framework for MEC which enabled IOV to minimize the energy consumption of RSU under the constraint of task waiting time. Combined with the task scheduling between MEC servers and the downlink energy consumption of RSU, a heuristic algorithm was established. Chen C et al. [6] constructed an online track compression framework running in the mobile environment, which includes two stages: (1) online track mapping stage, a novel compressor based on the direction change of intersection, namely direction change compression (HCC), is designed to develop a lightweight but efficient map matcher; (2) track compression stage, based on spatial direction matching (SD-matching) can match the sparse GPS points to the road

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Chen.

network and make full use of the vehicle GPS track data. Chen C et al. [7] proposed an economic fast travel service. In the first stage, the offline historical taxi track data is mined to identify the shortest travel path based on the estimated travel time under any given starting and ending point. In the second stage, an online adaptive taxi dispatching algorithm is constructed to select the route and determine the optimal travel service according to the real-time request iterative calculation Service path. Chen C et al. [8] proposed a two-stage probability framework called TripImputor, which is used to estimate the purpose of taxi travel and recommend services to passengers at the place where they get off. Ning Z et al. [9] built a three-layer VFC model to realize distributed traffic management and minimize the response time of vehicle collection and event release. Kong X J et al. [10] briefly introduced the latest technology and application of MCS in smart city. Firstly, the paper summarizes MCS in smart city and highlights its main characteristics; secondly, it introduces the architecture and supporting technology of MCS; secondly, it studies the latest application of MCS in smart city; finally, it discusses the openness and future research challenges of MCS in smart city.

Mennis J [11] proposed an ANN traffic congestion detection algorithm. Based on the basic parameters such as the running speed, flow and density of the upstream and downstream traffic flow of the detection section, the input layer, the middle layer and the output layer of BP neural network were constructed to process the data and fit the parameters to determine the road traffic congestion states. Buczak al et al. [12] proposed a new algorithm to distinguish traffic congestion based on the analysis of the fluctuation of road traffic flow parameters. Based on the analysis of a large number of traffic flow road occupancy data, the deviation of cumulative occupancy between upstream and downstream observation stations is calculated. If the fluctuation range of the difference value is large, it is determined that the traffic congestion occurs, and the difference value jumps around the zero value, then it is determined that the traffic operation is normal. Gal A and Sagi T [13] carry out traffic congestion discrimination according to T-test principle, and obtain more vehicle road occupancy data. T-test method is used to calculate the average and standard deviation of road vehicle occupancy rate in the last 10 cycles. Similar to the above standard deviation algorithm, if the result greatly deviates from the preset threshold value, the traffic congestion has occurred.

Gal A et al. [14] proposed a cellular neural network algorithm. According to the signaling data of mobile phone, a large number of data such as signal transmission time, mobile phone identification code, area identification code and vehicle stop time are obtained, and BP neural network is established. Whether there is traffic jam is determined by neural network output.

Y. Karmarianakis and P. Prasacos [15] build a new traffic congestion prediction algorithm based on spatial-temporal sequence model through highway network model verification and computer parallel computing structure, and prove that it achieves better prediction effect than other traditional prediction methods Turochy R E [16] used the truth constraint method based on fuzzy logic and neural networks. The pseudo-outer fuzzy neural network traffic flow state prediction model was established based on the forward-predictive network prediction effect of the traditional BP neural network. Quek C [17] conducted an experiment based on historical data mining of road sections, combined with real-time traffic situation changes, calculated the degree of deviation between real-time upload data of road vehicles and historical data, and the nonparametric regression prediction algorithm based on K-nearest was re-tested and improved. According to the traffic flow data of the first to the 80th roads in California, Y. Xie and Y. Zhang[18] established a wavelet neural network traffic flow state prediction model which was better than the traditional BP and RBF neural network model.

At the same time, L.VAna and L.R.Rett [19] combine the integrated neural network with support vector machine technology. Based on the forward feedback neural network prediction algorithm proposed above, the road traffic flow state prediction model is constructed, and the traffic flow state prediction simulation is carried out by using the highway floating car data as the data input. For predictions with a small number of samples, it has better prediction accuracy and simulation results. On the basis of fuzzy pattern recognition, Guo Xueting proposes a new algorithm to distinguish traffic congestion, which divides the traffic congestion into four dimensions: smooth traffic, normal traffic, congestion and congestion. Using the fuzzy pattern recognition theory, based on the three parameters of the average delay, speed and occupancy rate of vehicles at the signal intersection, the membership degree of each state is calculated to distinguish the traffic congestion status, mainly It is used to distinguish the congestion state of urban main road traffic flow.[20] Hao Yuan et al. [21] combined the urban road traffic state with the basic parameters detected by the detection equipment, based on various factors affecting the generation of road traffic congestion, using multiple classifiers to identify the traffic flow state, which has a better effect. Zou Liang et al. [22] divided the predicted road segment into three parts, obtained the running time of the floating car between each road segment through data fusion technology, extracted the data delay amount and the road segment estimated time from the sensor data, and input it into BP. A neural network model that predicts the state of road traffic flow after the simulation. Zhu Shengxue et al. [23] used the traffic flow state prediction model to be equivalent to the function estimation approximation. Based on wavelet decomposition-support vector regression calculation, wavelet decomposition and variation of traffic flow parameters are performed. A basic traffic signal and interference signal of road traffic flow are processed by establishing support vector machine, and a short-term traffic flow prediction model is established. The validity of the model is verified based on mathematical methods and real data, and the results with higher accuracy

and lower error are obtained. Peng Qiyuan et al. [24] added genetic algorithm optimization based on support vector machine regression algorithm. A GSVMR model with the advantages of minimum structural risk and easy optimization is established. The validity of the model in the short-term prediction of urban road network traffic flow is verified by the actual data collected. Guan Wei et al. [25] based on the fuzzy C-means clustering algorithm, input the road traffic speed and its variance as the feature vector of road traffic flow state analysis, and establish a traffic flow state prediction model. At the same time, the influence of various fast-track traffic flow parameters in the clustering process is discussed, which has certain feasibility and effectiveness. Yu Wanxia et al. [26] used the particle swarm optimization algorithm to optimize the calibration of fuzzy neural network parameters. Based on the fuzzy neural network, the short-term traffic flow was predicted, and its high prediction accuracy was verified by simulation. Nie Hongtao et al. [27] calibrated the structural parameters of the Radial Basis Function neural network proposed by Powell through FCM algorithm, and applied it to traffic state prediction. The experimental results showed that the model has better prediction effect and practicability.

In summary, the current research on traffic flow state prediction is analyzed. It is mainly carried out by traditional methods such as BP neural network and support vector machine. However, such algorithms usually adopt traditional traffic flow parameters as feature vectors for training prediction, and lack of consideration for the distribution of traffic flow state, often for the prediction of traffic flow state. There are significant limitations at intersections or simply upstream and downstream roads [28]. Therefore, based on the Moran 'I, this paper introduces the time dimension, constructs the spatial-temporal Moran' I, analyzes the spatio-temporal aggregation dissipative characteristics of the road network traffic flow, and combines the classification tree and the regression tree to construct a mixed forest prediction model based on the spatio-temporal state of traffic flow, and other traditions. The prediction method is compared with the error index of the Moran quadrant without adding the high correlation path as the eigenvector to prove the effect of the model prediction.

II. JUDGMENT OF CONGESTION STATES

At present, such conventional traffic flow parameters as speed, occupancy and flow rate are generally used as the evaluation indicators for road traffic congestions, but these indicators often cannot fully reflect the travel states of vehicles on the road. In addition, both at home and abroad, there is no precise quantitative indicator to calculate the threshold for traffic congestion judgment, and there is no uniform standard.

In this research, in addition to selecting the average travel speed in road sections as the basic measure of traffic conditions, the following parameters are added as supplementary indicators: common congestion coefficient CI , defined as the difference between the road section's actual travel time and its designed travel time; low-speed travel time ratio ST, defined as the ratio of the road section's low-speed travel time to the total travel time; vehicle acceleration noise AN; and average speed gradient MVG. All of the above supplementary indicators can reflect the travel changes of vehicles at the same speed, so they are good description supplementary to the road traffic flow state as reflected by V_{ave}.

$$CI = \left(\frac{T - T_0}{T_0}\right) \tag{1-1}$$

where, T indicates the actual travel time of a road section, and T_0 indicates the designed travel time of the road section.

$$ST = \frac{T_s}{T} \tag{1-2}$$

where, T_s indicates the time period during which the speed is less than 3km/h; and T indicates the entire travel time.

$$AN = \frac{1}{T - T_s} \sum_{i=1}^{n} \frac{\Delta v_i^2}{\Delta t_i} \tag{1-3}$$

where, Δt_i indicates the time used for speed changes; and n indicates the number of speed changes in a certain road section.

$$MVG = AN / \frac{1}{T} \int_0^T v dt \tag{1-4}$$

For the assessment, the main evaluation factor V_{ave} as mentioned above as well as the four supplementary evaluation factors are used. The traffic congestion state will be evaluated by using the fuzzy C-means algorithm. The steps of fuzzy C-means clustering algorithm (FCM) are as follows:

Step1: Calculate by using the ordinary K-means clustering algorithm to obtain the classification number C and the initial input matrix R;

Step2: Set the weighted index and the minimum error value as constraints;

Step3: Make separate calculations to get the clustering center matrix V and the fuzzy classification matrix R;

Step4: Evaluate the convergence effect based on the entropy value H;

Step5: Adjust the Q value to iterate and calculate, until meeting the requirements.

Finally, the classification threshold for traffic flow states is obtained:

Therefore, through the above indicators, the spatiotemporal congestion index SI is constructed.

$$SI = V_{ave} - \alpha \cdot CI - \beta \cdot ST - \chi \cdot AN - \delta \cdot MVG \tag{1-5}$$

The two most important characteristics of spatiotemporal data are autocorrelation and stability. In the spatiotemporal data of traffic flows, the spatiotemporal autocorrelation refers to the correlation of the spatial and temporal distributions of road traffic flows' attribute values (such as flow, density, or speed). The spatiotemporal stability means that the states and characteristics of road traffic flows will vary over time with changes of spatial locations.

By observing and researching the relationships among road traffics' basic parameters, such as their flows, densities and speeds, it can be found that the congestion phenomena have significant spatiotemporal diffusions. In terms of the time dimension, the traffic congestion life cycle can be divided into 4 phases, namely, generation, diffusion, dissipation and end. When the inflow traffics exceed the roads' carrying capacity, congestions begin to occur and diffuse along the main roads. When the traffics flowing into the road sections gradually decrease until less than the outflow traffics on the road sections, the congestions get into dissipation. In terms of the space dimension, it can be found that when road congestions occur, the cross-sections of roads with a reduced traffic capacity will move upstream along with the vehicle line-ups.

III. ANALYSIS OF SPATIO-TEMPORAL CORRELATION OF ROAD NETWORK TRAFFIC

After mastering the spatial-temporal transmission path of the upstream and downstream of the road congestion through the correlation coefficient method, it is necessary to conduct quantitative research on the diffusion of the traffic congestion in the road network. Based on the theory of spatial autocorrelation, Wei Wei introduced the time dimension into the traditional Moran spatio-temporal index based on spatial autocorrection and established an improved Moran's I method [29] to analyze the spatio-temporal state of urban road traffic.

The spatial autocorrelation Moran's I can be expressed as:

$$I = \frac{n \sum_{i=1}^{N} \sum_{j=1}^{n} w_{i,j} z_i z_j}{S_0 \sum_{i=1}^{n} z_i^2} \tag{2-1}$$

where: z_i represents the degree of deviation between the attribute assignment of element i and its mean, ie. $x_i = \bar{X}$. $w_{i,j}$ is the spatial weight between the elements i and j. n is equal to the number of elements. S_0 is an aggregation of all spatial weights:

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} \tag{2-2}$$

By introducing the spatial-temporal Moran 'I of the spatial-temporal dimension, the global space Moran 'I is extended to the spatial-temporal domain. Definition $y_{(p,i)}$ is an attribute value of the spatial-temporal element $ST_{(p,i)}$. Calculated as follows:

$$I = \frac{NT \sum_{p=0}^{N} \sum_{i=0}^{T} \sum_{q=0}^{N} \sum_{j=0}^{T} w_{(p,i)(q,j)} \left(y_{(p,i)} - \bar{y}\right) \left(y_{(q,j)} - \bar{y}\right)}{\sum_{p=0}^{N} \sum_{i=0}^{T} \left(y_{(p,i)} - \bar{y}\right)^2 * \sum_{p=0}^{N} \sum_{i=0}^{N} \sum_{q=0}^{N} \sum_{j=0}^{T} w_{(p,i)(q,j)}} \tag{2-3}$$

where:

$$\bar{y} = \frac{1}{NT} \sum_{p=0}^{N} \sum_{i=0}^{T} y_{(p,i)} \tag{2-4}$$

N,T are the number of spatial sequences and the number of time series, respectively. $w_{(p,i)(q,j)}$ is a weighted value of the connection relationship between the spatial-temporal elements $ST_{(p,i)}$ and $ST_{(q,j)}$.

The local Moran spatio-temporal index is calculated by:

$$I_{(p,i)} = Z_{(p,i)} W z_{(p,i)} \tag{2-5}$$

$$W z_{(p,i)} = \frac{\sum_{q=0}^{N} \sum_{j=0}^{T} w_{(p,i)(q,j)} Z_{(q,j)}}{\sum_{q=0}^{N} \sum_{j=0}^{T} w_{(p,i)(q,j)}} \tag{2-6}$$

$$Z_{(p,i)} = \frac{\left(y_{(p,i)} - \bar{y}\right)}{\sigma} \tag{2-7}$$

$$\sigma = \sqrt{\frac{\sum_{p=0}^{N} \sum_{i=0}^{T} \left(y_{(p,i)} - \bar{y}\right)^2}{NT - 1}} \tag{2-8}$$

In traffic research. The global positive correlation is reflected in the smooth state of the overall road, and the congestion states is aggregated through the spatial-temporal dimension. Global road network roads have a common feature of aggregation. The greater the correlation, the more significant the aggregation characteristics. That is to say, the global road network presents a smooth (congested) feature at a certain moment, and the traffic state of the global road network in the adjacent time period also presents a smooth (congested) situation. The global negative correlation is reversed, which is reflected in the unimpeded state of the global road network or the spatiotemporal and anomalous characteristics of the congestion states. The greater the correlation, the more significant the anomalous feature. That is to say, the road network presents a smooth (congested) feature at a certain moment, and the traffic state of the global road network at the adjacent time is in a state of congestion (smooth).

The local positive correlation reflects that the traffic flow state of the road segment is aggregated through the spatial-temporal dimension and the traffic flow state of the surrounding road segment in the adjacent time period. The local road traffic flow state has a common characteristic of the aggregation characteristics. The greater the correlation, the more significant the aggregation characteristics between the local road segments. That is to say, a certain section of the road presents a smooth (or congested) feature at a certain moment, and the road traffic state of the surrounding road section at the adjacent time also presents a smooth (congested) situation. The partial negative correlation is opposite, showing the anisotropic properties.

The Moran scatter plot is shown in Figure 1. The propagation characteristics of the road segment in the scatter plot are represented by the quadrant position of the spatial-temporal element corresponding to the road in the scatter plot:

① First quadrant. Unblocked aggregation. If the road is clear at the moment, the surrounding roads will be clear at the near time.

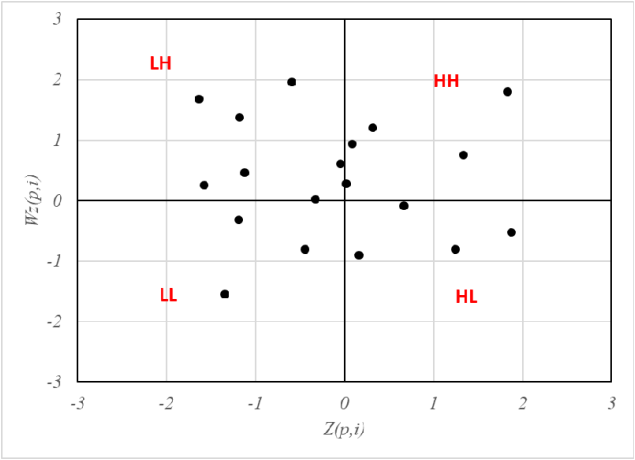


FIGURE 1. Spatio-temporal Moran scatter plot.

② Second quadrant. Unblocked. If the road is congested at the moment, the surrounding roads will be unblocked at the near moment. The road congestion comes before the surrounding road network congestion, and it lags behind the surrounding road network congestion and dissipation.

③ Third quadrant. Congestion aggregation. If the road is congested at this moment, the surrounding roads will be congested in the near future.

④ Fourth quadrant. Congestion isotropic. If the road is clear at the moment, the surrounding roads will be congested at the near moment. The road lags behind the congestion of the surrounding road network, leading to the congestion and dissipation of the surrounding road network.

IV. SPATIO-TEPORAL CHARACTERISTICS FUSION PREDICTION OF ROAD NETWORK BASED ON MIXED FOREST MODEL

A. RANDOM FOREST STRUCTURE

The key step of random forest algorithm is to build decision tree. According to the characteristics of traffic state, the CART method can be used to construct classification decision tree (predict the time and space congestion index of the road, distinguish the traffic state according to the index) and regression decision tree (directly predict the traffic state). The basic construction steps are as follows:

Step 1: Selecting sample properties. According to the high correlation path selection algorithm based on the fusion coefficient, the high correlation segment set D of the segment p is extracted: $\{D1, D2 \ldots, Dp-1, Dp, Dp+1 \ldots Dn-1, Dn\}$, and the traffic flow sample attribute is selected., including: week, time period (data smoothed to 10min, divided into $0 \sim 143$ time periods every day), driving direction (0: positive, 1: reverse), quadrant of Moran 'I of other segments in high correlation path D, and vehicle of segment p in period i average speed, average speed of vehicles of section p in period i-I, average speed of vehicles of section p in period $i+I$, spatiotemporal congestion index of vehicles of section p in period i-I, spatiotemporal congestion index of vehicles of section p in period i, spatiotemporal congestion index of vehicles of section p in period $i+I$, state of vehicle traffic flow of section p in period i (judged according to spatiotemporal congestion index above, 1-locked, 2 -Severe congestion, 3-congestion, 4-slow traffic, 5-smooth traffic). Among them, the output results are respectively the vehicle traffic flow state of segment m in n period or the spatiotemporal congestion index SI of segment m in n period. Since the spatiotemporal congestion index is based on the average speed of the road, in order to make the feature vectors independent, the classification tree uses the average speed as the feature for model training, and the regression tree uses the spatio-temporal congestion index as the feature for model training.

Step2: Selecting the split attribute. Based on the good description of the road congestion trend of the spatio-temporal Moran exponent scatter plot, the quadrant of the spatial-temporal Moran 'I of the road segment p in the i period is selected as the root node.

In the case of classification, in order to make the data pure, the output is closer to the true value. The GINI value is used to measure the purity of the node. The GINI index is between 0 and 1. 0 means that the categories are completely equal, and 1 means completely different. The more disorderly the overall inclusion category, the larger the GINI index:

$$GINI = 1 - \sum_{q \in Q} p_q^2 \tag{2-9}$$

where: Q is the set of all categories, and the probability that the sample points belong to the class q is p_q.

This step sets two classes, which are (1) Node class Node, which mainly includes split attribute; node output class, child node, depth and so on. (2) The split information class Info mainly includes the number of each class of the child nodes, the row coordinates of each child node, the split attribute of the node, the type of the attribute.

In the case of regression, the sample variance is used as an indicator to measure node purity.

$$\sigma = \sqrt{\sum_{q \in Q} (x_q - \mu)^2} = \sqrt{\sum_{q \in Q} x_q^2 - n\mu^2} \tag{2-10}$$

where: Q is the set of all classifications, μ is the mean of the prediction results in the sample set, and x_q represents the prediction result of the q-th sample.

Step 3: Continuing to split. After the first division, we continue to select new attributes for a new round of division, and calculate Gini value or sample regression variance. The selected split attribute should satisfy the constraint of minimizing the Gini value or regression variance of the child nodes. The next step is to split up. That is to minimize:

$$\text{Gain} = \sum_{q \in Q} p_q \cdot Gini_q \tag{2-11}$$

$$\text{Gain} = \sum_{q \in Q} \sigma_q \tag{2-12}$$

Step 4: Data segmentation. For discrete values, how many values are divided into several nodes. Because cart is a binary tree, that is, one of the discrete values is independent as a node, and another node is generated by the rest of the discrete values and calculate the Gini value of each division method respectively; for the continuous type attributes, the continuous type attributes are sorted first, and divided into two parts in a one size fits all manner, and N-1 cutting shall be conducted between N data to calculate the Gini value of each division method. The optimal segmentation method is determined by Gini value after data segmentation. The Q-attribute obtained is the optimal split attribute of the sample set. The q attribute value is the optimal split attribute value. Similarly, the regression tree is to select the optimal sample regression variance.

Step 5: Stopping splitting. Stop splitting when the following conditions occur in the node: (1) all observations belong to the same category; (2) all observation attribute values are consistent; (3) the depth of the decision tree reaches the set threshold; (4) the number of observations included is less than the threshold value of the number of observations that the parent node should include in the set rule; (5) the number of observations included in the lower level child node is about to be less than the set threshold value; (6) no longer has the attribute to meet the threshold value of the set split rule.

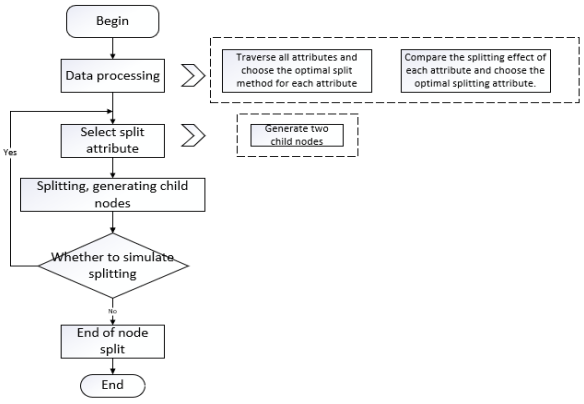


FIGURE 2. Decision tree construction process.

When the condition of stop splitting occurs, the decision tree will grow completely, which may lead to over fitting phenomenon in the case of large amount of data and more features. Therefore, the random forest model is used to vote on the output results.

The random forest algorithm is based on the foundation of bagging algorithm. The specific process steps are as follows:

Step 1: Based on Bootstrapping method, n samples (with playback) are randomly selected from the original training set. A total of k samples were taken. Generate k independent training sets with repeatable elements.

Step2: For k training sets, one-to-one training k corresponding decision tree models.

Step 3: For a single decision tree model. Suppose the number of training sample features is m. In each split, we choose the split attribute with the smallest GINI index or regression variance.

Step 4: Keeping the decision tree split until the split stop condition is reached, and do not prune the decision tree in the middle.

Step 5: Combining multiple decision trees to form a random forest. And the output results of all decision trees are finally output by voting.

Due to the different number of high correlation paths in different periods, it is necessary to fill in the missing values: firstly, the mode complement of Moran 'I in local time and space is used for coarse-grained filling, after random forest model training, the statistical similarity matrix (if two observation instances fall on the same node in the same tree more times, then two observation facts The higher the similarity of the example is), the weight of the uncertain similarity is used to vote, and the above voting scheme iterates for 4-6 times.

B. IMPROVED MIXED FOREST MODEL

Because there are two ways to distinguish traffic operation state: regression and classification, and the result of regression can be used for secondary classification. In order to enhance the accuracy of prediction results, the random forest algorithm is improved. A mixed forest model is proposed.

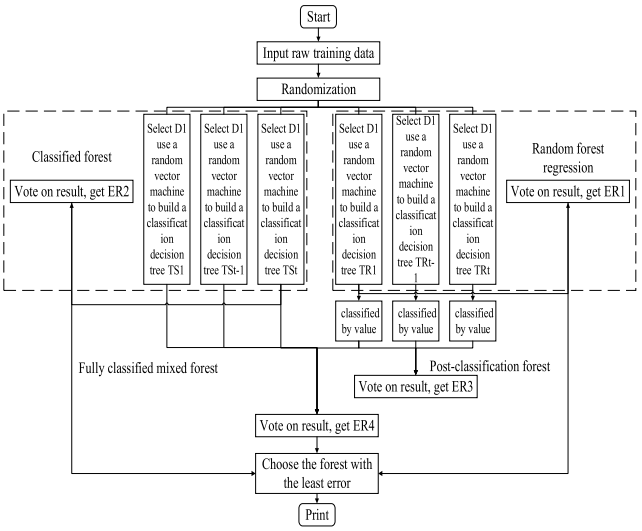


FIGURE 3. Mixed forest construction process.

The specific steps are as follows:

Step 1: Inputting the original sample, randomize the data, create random vectors, and reduce the correlation between decision trees.

Step 2: Through Bootstrapping method, m classification sample sets and n regression sample sets are randomly selected.

Step 3: Making the regression decision tree and classification decision tree grow completely without pruning branches. Build random forest by regression decision tree and classification decision tree respectively, and calculate the error, and get the classification error ER1 and regression error

ER2 respectively. All decision tree results are retained at the same time.

Step 4: According to the spatiotemporal congestion coefficient value obtained from the regression decision tree, combined with the previous traffic state discrimination standard, the predicted state is classified and transformed into a classification decision tree. Only the results obtained from the regression forest are classified into post classified forest, and the post classified error ER3 is calculated.

Step 5: To integrate classified forest and return forest, vote on the results, and get mixed error ER4.

Step 6: Selecting the forest with the smallest error and output the prediction results.

In terms of tuning, in general, the number of features is $k = \sqrt{Q}$, the maximum depth is no more than 8 layers, the tree of the tree should not be too much, and the minimum number of split samples should be more than 100 empirically.

There are two main factors that affect the accuracy of random forest algorithm: one is the tree association between any two decision trees in the forest: the stronger the independence between decision trees, the higher the accuracy of discrimination; the other is the accuracy of discrimination of a single tree: the stronger the accuracy of discrimination of a single decision tree, the lower the error rate of discrimination of the whole forest.

V. CASE ANALYSIS

This paper takes the GPS data of more than 3000 taxis in Chengdu as the research object. A congestion recognition model based on speed and trajectory is proposed. The specific range is (104.04214E-104.12958E, 30.625294N-30.72775N). In total, 142 roads upload data every 3s, totaling more than 2 billion data, with a storage scale of about 180gb, including vehicle oid, order id, time, longitude and latitude. The data is shown in Table 1:

TABLE 1. Data description.

Vehicle oid	Order id	Time.	Longitude	Latitude
1001	256	2016-11-21 07:12:42	104.07513	30.72724
1001	256	2016-11-21 07:28:12	104.07611	30.72611
1002	11	2016-11-22 09:34:02	104.07422	30.72379
1002	11	2016-11-22 10:15:26	104.07434	30.72406
.....	.....	.....	.....	.....

A. DIVISION AND EVOLUTION OF TIME AND SPACE STATE OF ROAD NETWORK TRAFFIC

1) ANALYSIS OF GLOBAL MORAN SPATIO-TEMPORAL INDEX

After data preprocessing, the data were smoothed into 10min interval data, the road space-time congestion index *SI* was calculated, and the road traffic status was judged. By calculating the correlation coefficient, the high correlation path set *D* is obtained. Based on the high correlation path to calculate the spatio-temporal Moran 'I, this paper analyzes 20 road sections around Chengdu North railway station.

TABLE 2. Example road section.

No.	Road Name	No.	Road Name
1	North Station West Second Road	11	Chenghua Street
2	North Station West First Road	12	North Third Section of Second Ring Road
3	North Second Section of First Ring Road	13	Section I of Jiefang Road
4	Second Section of Renmin North Road	14	Jiefang West Road
5	Section II of Jiefang Road	15	Section II of Jiefang Road
6	Da'an West Road	16	Jiulidi South Road
7	Wudu Road	17	North New Main Road
8	North Station East 1st Road	18	Rongbei Business Avenue
9	Fuhe Avenue	19	Baliqiao Road
10	Chenghua West Street	20	North Road of Station

Calculating the global spatial-temporal Moran' I of the road network:

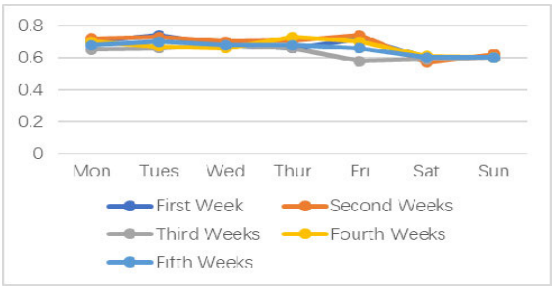


FIGURE 4. November 1st-December 6th instance road segment global spatial-temporal Moran 'I.

Due to the characteristics of a large number of commuting trips on weekends, more random trips such as shopping and tourism increased, the aggregation of the traffic state of the global road network is significantly lower than that during the working day, and the correlation of residents' travel behavior

in the space-time dimension is significantly reduced, thus showing the characteristics of random and scattered.

Taking the research data as an example, the local spatial-temporal Moran index of each road segment in the road network is calculated. The distribution is measured by the frequency of different Moran 'I in each section in 60 days. This paper analyzes and studies the local spatial-temporal evolution law of road traffic flow operation state.

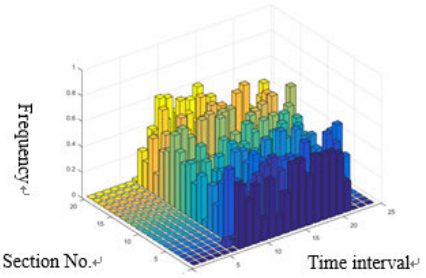


FIGURE 5. I < 0, frequency distribution.

2) MORAN' I DISTRIBUTION IN LOCAL SPATIAL-TEMPORAL

08:00-22:00 is a concentrated time period of spatial-temporal elements with local spatial-temporal Moran 'I value less than 0, including working hours and part of nighttime. The reason is that the spatial-temporal anomalies are mainly concentrated in the daytime, and the surrounding roads are intertwined and affect each other, resulting in traffic anomalies in local areas.

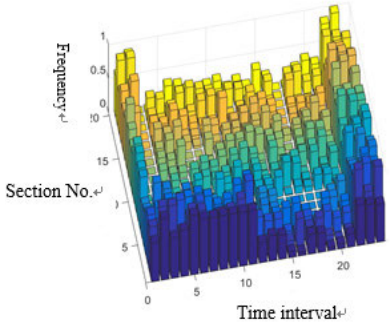


FIGURE 6. 0 ≤ I < 1, frequency distribution.

At the same time, the spatiotemporal objects are evenly distributed among index values [0,1) in all periods of the day. The main reason is that the overall road traffic states around Chengdu station is in a weak convergence trend in the spatial-temporal dimension.

In the peak period of night and morning and evening, most regions showed strong aggregation characteristics with index greater than 1. The main reason is that the roads in the early morning converge smoothly with the congestion in the peak hours.

From the perspective of spatial dimension. From 00:00 to 06:00, the traffic conditions of all sections around Chengdu North Railway Station area, from 7:00 to 9:00, from 2nd Ring North Road to Renmin North Road, from 7:00 to 11:00, from the third section of 1st ring road to CaoShi street, from

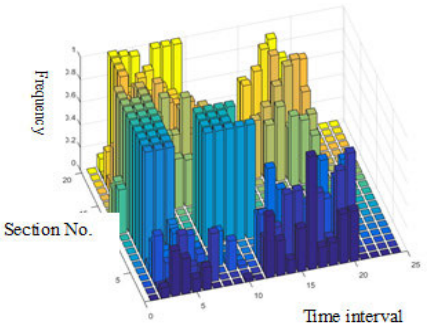


FIGURE 7. I > 1, frequency distribution.

14:00 to 20:00, from Renmin North Road to Xinhua Avenue show a strong convergence trend. According to this, it can be found that the high peak road is easy to be congested in the morning and evening.

3) ANALYSIS OF MORAN SCATTER DIAGRAM

According to the frequency of quadrant of each section in each period, the scatter diagram analysis is carried out:

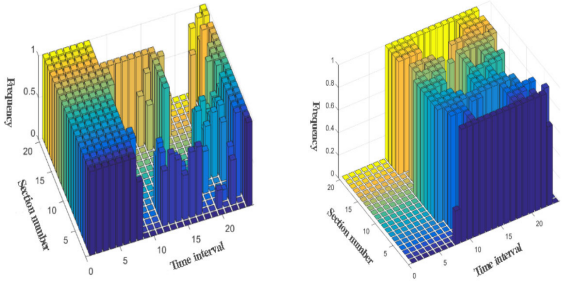


FIGURE 8. First quadrant frequency distribution (left) third quadrant frequency distribution (right).

First quadrant / third quadrant:

It can be seen from the distribution in the Figure 8 that all sections between 0:00 and 6:00, 7 sections around the Lotus pond between 12:00 and 13:00 at noon of East Jiulidi road in the early peak period, section 1-3 of Jiefang Road, and almost all sections after 22:00 are in the first quadrant of high frequency. These sections usually show obvious characteristics of smooth aggregation in the period.

From 7:00 to 11:00 in the morning, most of the roads except Jiulidi East Road and 1st Road of West station, and most of the roads from 14:00 to 19:00 except the first section of Beixing avenue to Ma'an East Road, are in the third quadrant of high frequency. These sections usually show obvious congestion aggregation characteristics in the period.

Second quadrant / fourth quadrant:

During the daytime, the temporal and spatial elements in the second and fourth quadrants are interspersed in each road section. It shows the characteristics of traffic state of the road network with different directions of traffic flow or congestion. Among them, the traffic diversion is mainly concentrated near the business district, which is prior to the road network congestion, that is, the congestion generation point. The different direction of congestion is mainly located

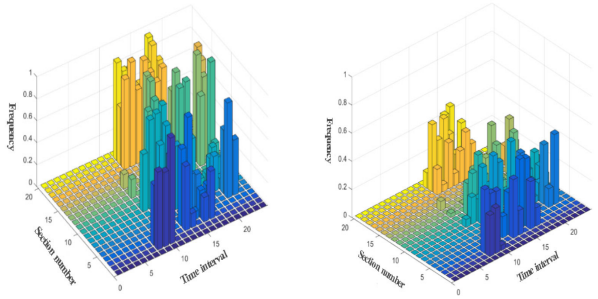


FIGURE 9. Second quadrant frequency distribution (left) fourth quadrant frequency distribution (right).

near the residential area of the main road, and the congestion is dissipated from these sections.

Generally, the traffic jam of Chengdu North railway station is serious on weekdays. But in the morning and evening peak period, it can still maintain the normal state of slow running, which is not easy to lock up. The main reason is that the sections with different directions of traffic flow and congestion are interwoven in the road network, and the congestion generation points and congestion dissipation points influence each other, so as to achieve faster congestion dissipation. Although the traffic flow is large and limited by the actual road conditions, it is easy to slow down the traffic, but the overall convergence between congestion is weak, and there is dissipation space. According to the time and space Moran scatter. The traffic flow states of the road section is divided into HH, LH, LL and HL. The road traffic congestion distribution state is pre classified. 1-unimpeded aggregation; 2-unimpeded heterogeneity; 3-congestion aggregation; 4-unimpeded heterogeneity.

B. PARAMETER SELECT

In the process of random forest construction, there are two main factors that affect the performance of the algorithm. One is the number of features used to build a single decision tree, the other is the number of decision trees built by random forest. In order to compare, different parameters are used to adjust the parameters according to the results. At the same time, in order to verify the importance of temporal and spatial characteristics of traffic in traffic state prediction, the local Moran quadrant of untrained road is added, only Moran quadrant of predicted road, Moran quadrant of predicted road and high correlation road is trained, and Moran quadrant of all roads in the road network is trained. After processing, there are 1635840 data with 21265920 eigenvalues, it is the period from 2016-11-1 to 2016-12-15, which is selected as the training set. The data between 2016-12-16 and 2016-12-31 is selected as test set. The all parameters are shown in Table 3 and Table 4.

The ER1 error analysis of classified forest is carried out.

As can be seen from the results in the Figure 10, when the Moran quadrant feature of the road network is added for training, the model prediction error value is significantly reduced, and with the increase of the number of Moran

TABLE 3. Classification tree parameters.

No	Data	Week	**Time**	Dir	[Moran]	Speed b	Speed n	Speed a	State
123	12	4	**65**	0	[3,1,3,2]	42	45	25	2

TABLE 4. Regression tree.

No	Data	Week	Time	Dir	[Moran]	SIb	SIn	SIa	State
123	12	4	65	0	[3,1,3,2]	42	45	25	2

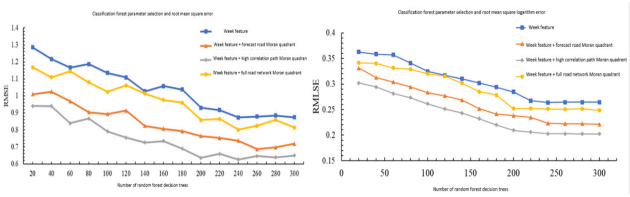


FIGURE 10. Selection of classified forest parameters and root mean square error (left) and Classified forest parameter selection and root mean square logarithm error (right).

quadrants of high correlation paths, the model prediction effect is significantly improved. However, after training the Moran quadrant of the whole network, the error increases. Because of too many eigenvectors, the training effect of the model is over fitted and the error increases. The training time of the model is prolonged obviously.

At the same time, based on the error analysis, it can be clearly seen that when the number of classified random forest decision trees reaches 240, the training effect of the model is stable. Therefore, the training week characteristics and high correlation path Moran quadrant are selected for classified forest, and the number of decision trees is 240.

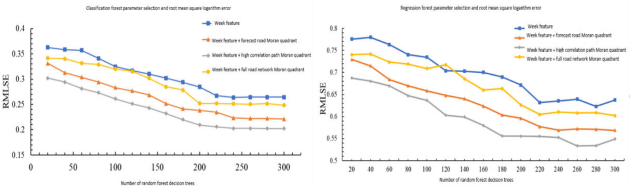


FIGURE 11. Regression forest parameter selection and root mean square error (left) and Regression forest parameter selection and root mean square logarithm error (right).

The ER2 error analysis of regression forest was carried out.

When the Moran quadrant feature of the road network is added for training which is similar to the classification tree, the prediction error of the model is significantly reduced, and with the increase of the number of Moran quadrants of high correlation path, the prediction effect of the model is significantly improved. After training the Moran quadrant of the whole network, because of too many feature vectors, the training effect of the model is over fitted, and the error increases, even the error is larger than that of only training week features. The training time of the model was prolonged obviously. However, due to the characteristics of regression

forest error value, the absolute error value of each training set is greater than that of classified forest.

At the same time, based on the error analysis, it is obvious that when the number of regression random forest decision trees reaches 260, the training effect of the model is stable and the error is the smallest. Therefore, the number of decision trees is 260.

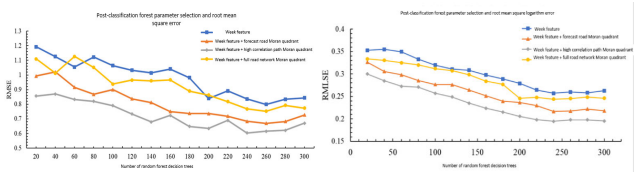


FIGURE 12. Parameter selection and root mean square error of post classified forest (left) and Parameter selection and root mean square logarithm error of post classified forest (right).

The ER3 error analysis of post classified forest was carried out.

The post classified forest still has the same error characteristics-with the addition of high correlation path Moran quadrant features for training, the error is greatly reduced, and with the increase of the number of forest decision trees, at the same time, it is reduced because the spatiotemporal congestion index obtained by regression is classified again, the prediction results can be classified, so that the error results are greatly reduced, slightly lower than the ordinary effect of forest classification.

Based on the error analysis, it can be seen clearly that when the number of post classified random forest decision trees reaches 240, the training effect of the model is stable and the error is the smallest. Therefore, the number of decision trees is 240.

The ER4 error analysis of mixed forest was carried out. The number of decision trees used to construct classified forest and return forest is 1:1.

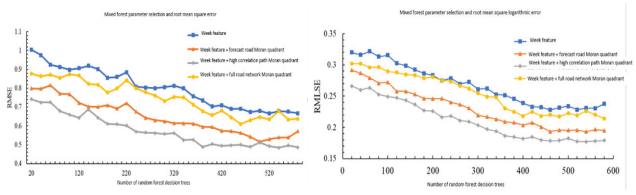


FIGURE 13. Mixed forest parameter selection and root mean square error (left) and Mixed forest parameter selection and root mean square logarithm error (right).

From the error analysis results, it can be found that the mixed forest performance has good linear characteristics compared with the pure regression forest. With the increase of the number of random forest decision trees and the Moran quadrant of high correlation path, the prediction effect of the training model is better. At the same time, compared with the pure classified forest and the post classified forest, the error is further reduced. The minimum root mean square error is reduced to about 0.5. The error of the least root mean square logarithm is less than 0.2.

Based on the error analysis, it is obvious that when the number of mixed random forest decision trees reaches 500, the training effect of the model is stable and the error is minimum. Therefore, the number of decision trees is 540.

The error, training time in the case of single machine and training time in the case of cluster and acceleration ratio were compared in the case of optimal training of all kinds of forests.

TABLE 5. Comparison of training effects of different forest types.

Forest type	The number of decision trees	RMSE	RMLSE	Single time/s	Cluster time/s	Speed up ratio
Classified forest	240	0.649337182	0.2023	4.9	3.6	1.361
Regression forest	260	1.578250302	0.533719562	5.4	5.1	1.059
Post classified forest	240	0.604034359	0.193749189	7.3	6.2	1.177
Mixed forest	540	0.500497229	0.177733053	12.5	6.3	2.083

Through comparison, we can see that the error of mixed forest model is the best in all aspects, but the training model takes the longest time. The training time of pure classified forest is the fastest, but the error index is higher than that of mixed forest. At the same time, it can be seen from the acceleration ratio that the mixed forest model is only slightly lower than the classified forest, which has an excellent ability of big data distribution cluster computing. This is because the mixed forest model integrates two types of decision trees, which not only retains the advantages of the two decision trees, but also overcomes their disadvantages. Thus, the accuracy of the mixed forest model is improved.

Therefore, the mixed forest model is selected to predict the traffic states of the road section. The number of decision trees is 240, and the training characteristics are time characteristics and high correlation path Moran quadrant.

C. MODEL TRAINING RESULTS

To evaluate the prediction performance of the designed model. According to the prediction results of four kinds of forest, the root mean square error (RMSE) and root mean square logarithm error (RMLSE) are selected to measure, and the deviation between the measured value and the model prediction value is calculated. Calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(a_i - p_i)^2} \tag{4-1}$$

$$RMLSE = \sqrt{\frac{\sum_{i=1}^{n}\left[\lg(p_i+1) - \lg(a_i+1)\right]^2}{n}} \tag{4-2}$$

where: n is the sample size, p_i is the true value of the sample, and a_i is the predicted value.

That is, for the performance of classified forests and mixed forests, the evaluation indicators are as follows:

$$S_p = t/T \tag{4-3}$$

where: T is the running time in the case of a single machine, and t is the running time of the cluster model.

TABLE 6. Comparison of regression prediction results of different prediction algorithms.

Prediction algorithm	RMSE	RMLSE
Regression forest algorithm	1.598	0.562
Decision tree algorithm	3.566	1.138
Bayesian algorithm	3.525	1.264
K-means algorithm	2.492	0.913
Support vector machine algorithm	2.695	0.984

As can be seen from the results in the table that the regression forest algorithm predicts the spatio-temporal congestion coefficient of road traffic, and the accuracy is significantly higher than other traditional algorithms. Compared with the two RMSE and RMSLE error indicators, there is a significant improvement, which indicates that the model has higher prediction accuracy than other prediction models.

For the same road on the same date, compare with the sample set without considering Moran Quadrant:

TABLE 7. Comparison of regression prediction results of different prediction algorithms.

Use sample set	RMSE	RMLSE
Moran quadrant + time factor + speed + congestion status	1.598	0.562
Time factor + speed + congestion status	3.429	1.097

As can be seen from the results in the table that the training sample set with high correlation path Moran quadrant feature is significantly higher than the training sample set without high correlation path feature in prediction accuracy, because Moran quadrant is an important indicator reflecting the road traffic agglomeration. Adding this feature to the prediction algorithm can effectively improve the accuracy of traffic congestion index prediction.

Traffic congestion was predicted using classified forest, post-classified forest and mixed forest.

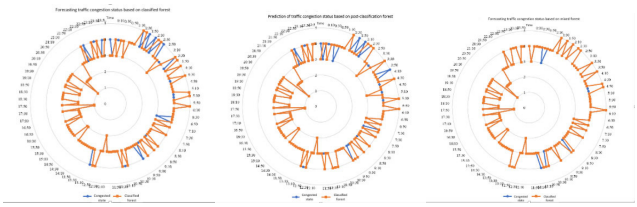


FIGURE 14. Traffic state prediction based on classified forest (left), post classified forest (middle), mixed forest (right).

In the Figure 14, the traffic congestion levels are 1-locked, 2-severe, 3-congested, 4-slow and 5-smooth respectively. The prediction results are basically consistent with the actual congestion states, reflecting the same trend of congestion states change. The prediction accuracy increases in sequence according to the order of classified forest, post classified forest and mixed forest, with the miscalculation rates of 13.28%, 11.81% and 4.17% respectively, especially the forecasting effect of the slow-down and congested state during the peak period is also good, which also reflects the low correlation between nighttime road traffic conditions and the occurrence of occasional incidents.. This model has good applicability and prediction effect for peak period and high correlation traffic state.

Compared with the traditional traffic flow parameter prediction algorithm for the same road on the same date:

TABLE 8. Comparison of classification and prediction effects of different prediction algorithms.

Prediction algorithm	RMSE	RMLSE	False positive rate
Classified forest algorithm	0.712	0.217	13.28%
Post-classification forest algorithm	0.695	0.209	11.81%
Mixed forest algorithm	0.524	0.186	4.17%
Decision tree algorithm	1.018	0.649	18.25%
Bayesian algorithm	0.747	0.312	12.67%
K-means algorithm	0.792	0.455	9.12%
Support vector machine algorithm	0.982	0.477	11.56%

As can be seen from the results in the table 6, the accuracy of hybrid forest algorithm in predicting the spatio-temporal congestion coefficient of road traffic is significantly higher than that of other traditional algorithms. Compared with other prediction models, the model in this paper has a higher prediction accuracy. For the same road and the same date, the mixed forest model was used to compare with the sample set without considering Moran quadrant.

TABLE 9. Comparison of classification prediction effect of different prediction sample sets.

Use sample set	RMSE	RMLSE	False positive rate
Moran quadrant + time factor + speed + congestion status	0.524	0.186	4.17%
Time factor + speed + congestion status	1.527	0.512	15.29%

As can be seen from the results in the table that the training sample set with high correlation path Moran quadrant feature is significantly higher than the training sample set without high correlation path feature in prediction accuracy, because Moran quadrant is an important indicator reflecting the road traffic agglomeration. Adding this feature to the prediction algorithm can effectively improve the accuracy of traffic congestion index prediction.

VI. CONCLUSION

In this paper, time dimension based on Moran' I is introduced to construct spatial-temporal Moran' I, and the

spatial-temporal aggregation and dissipation characteristics of road network traffic flow are analyzed. On the basis of the research data, the paper analyzes the traffic flow state of the example road and explores its state division and evolution characteristics. Based on the feature that the Moran quadrant of the road segment and its surrounding local spatial-temporal Moran' I reflect the aggregation of high and low values of road traffic flow state, a traffic flow state prediction model based on random forest is constructed. The random forest algorithm is improved. Based on the actual demand of traffic flow state prediction, the mixed forest model is constructed by combining classification tree and regression tree. Finally, the case data of the north second section of the first ring road in Chengdu are used for example verification. The evaluation indexes of the model are root mean square error and logarithm root mean square error. The optimal combination of eigenvectors and the number of decision trees for classified forest, regression forest, post classified forest and mixed forest algorithm are selected. At the same time, they are compared with other traditional prediction methods such as support vector machine algorithm and K-means algorithm. The results show that the model is effective and has the prospect of big data application.

REFERENCES

[1] L. Liu, C. Andris, and C. Ratti, "Uncovering cabdrivers' behavior patterns from their digital traces," *Comput., Environ. Urban Syst.*, vol. 34, no. 6, pp. 541–548, Nov. 2010.

[2] Z. Ning, Y. Li, P. Dong, X. Wang, M. S. Obaidat, X. Hu, L. Guo, Y. Guo, J. Huang, and B. Hu, "When deep reinforcement learning meets 5G vehicular networks: A distributed offloading framework for traffic big data," *IEEE Trans. Ind. Informat.*, to be published.

[3] Z. Ning, Y. Feng, M. Collotta, X. Kong, X. Wang, L. Guo, X. Hu, and B. Hu, "Deep learning in edge of vehicles: Exploring trirelationship for data transmission," *IEEE Trans. Ind. Informat.*, vol. 15, no. 10, pp. 5737–5746, Oct. 2019.

[4] Z. Ning, P. Dong, and X. Wang, "Deep reinforcement learning for vehicular edge computing: An intelligent offloading system," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 6, p. 60, 2019.

[5] Z. Ning, J. Huang, X. Wang, J. J. P. C. Rodrigues, and L. Guo, "Mobile edge computing–enabled Internet of vehicles: Toward energy–efficient scheduling," *IEEE Netw.*, vol. 33, no. 5, pp. 198–205, Sep. 2019.

[6] C. Chen, Y. Ding, X. Xie, S. Zhang, Z. Wang, and L. Feng, "TrajCompressor: An online map-matching-based trajectory compression framework leveraging vehicle heading direction and change," *IEEE Trans. Intell. Transp. Syst.*, to be published.

[7] C. Chen, D. Zhang, and X. Ma, "Crowddeliver: Planning city-wide package delivery paths leveraging the crowd of taxis," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 6, pp. 1478–1496, Jun. 2016.

[8] C. Chen, S. Jiao, S. Zhang, W. Liu, L. Feng, and Y. Wang, "TripImputor: Real–time imputing taxi trip purpose leveraging multi–sourced urban data," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 10, pp. 3292–3304, Oct. 2018.

[9] Z. Ning, J. Huang, and X. Wang, "Vehicular fog computing: Enabling real–time traffic management for smart cities," *IEEE Wireless Commun.*, vol. 26, no. 1, pp. 87–93, Feb. 2019.

[10] X. Kong, X. Liu, B. Jedari, M. Li, L. Wan, and F. Xia, "Mobile crowdsourcing in smart cities: Technologies, applications, and future challenges," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8095–8113, Oct. 2019.

[11] J. Mennis and D. Guo, "Spatial data mining and geographic knowledge discovery—An introduction," *Comput., Environ. Urban Syst.*, vol. 33, no. 6, pp. 403–408, 2009.

[12] A. L. Buczak, B. Baugher, E. Guven, L. C. Ramac-Thomas, Y. Elbert, S. M. Babin, and S. H. Lewis, "Fuzzy association rule mining and classification for the prediction of malaria in South Korea," *BMC Med. Inform. Decis. Making*, vol. 15, no. 1, Dec. 2015.

[13] A. Gal and T. Sagi, "Tuning the ensemble selection process of schema matchers," *Inf. Syst.*, vol. 35, no. 8, pp. 845–859, Dec. 2010.

[14] A. Gal, M. Katz, and T. Sagi, "Completeness and ambiguity of schema cover," in *Proc. Meaningful Internet Syst., OTM Conf.* Berlin, Germany: Springer, 2013.

[15] Y. Kamarianakis and P. Prastacos, *Space-Time Modeling of Traffic Flow*. New York, NY, USA: Pergamon, 2005.

[16] R. E. Turochy, "Enhancing short–term traffic forecasting with traffic condition information," *J. Transp. Eng.*, vol. 132, no. 6, pp. 469–474, Jun. 2006.

[17] C. Quek, M. Pasquier, and B. Lim, "POP–TRAFFIC: A novel fuzzy neural approach to road traffic analysis and prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 2, pp. 133–146, Jun. 2006.

[18] Y. Xie and Y. Zhang, "A wavelet network model for short-term traffic volume forecasting," *J. Intell. Transp. Syst. Technol., Planning, Oper.*, no. 3, pp. 23–35, 2006.

[19] L. Vanajakshi and L. R. Rilett, "Support vector machine technique for the short term prediction of travel time," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2007.

[20] G. Xueting, Q. Yanli, and L. Zhen, "Urban road congestion discrimination based on taxi GPS data," *Traffic Inf. Secur.*, vol. 31, no. 5, 2013.

[21] H. Yuan, S. Lijun, and X. Tiandong, "Analysis of traffic congestion on urban expressway and congestion threshold identification," *J. Tongji Univ., Natural Sci. Ed.*, vol. 36, no. 5, pp. 609–614, 2008.

[22] Z. Liang, X. Jianmin, and Z. Lingxiang, "Traffic state classification model of travel times based on the fusion technique," *J. Tsinghua Univ., Natural Sci. Ed.*, vol. 47, no. S2, pp. 128–136, 2007.

[23] Z. Shengxue, Z. Jun, and B. Xu, "Short-term traffic forecast based on WD and SVM," *J. Suzhou Univ. Sci. Technol., Eng. Technol. Ed.*, vol. 20, no. 3, pp. 79–82, 2007.

[24] R. Qiliang, X. Xiaosong, and P. Qiyuan, "GSVMR model on short-term forecasting of city road traffic volume," *Highway Transp. Technol.*, vol. 25, no. 2, pp. 134–138, 2008.

[25] Y. Yueming, G. Wei, and W. Jianping, "Urban expressway traffic states analysis based on video detection technique," *Traffic Inf. Secur.*, vol. 26, no. 4, pp. 1–3, 2008.

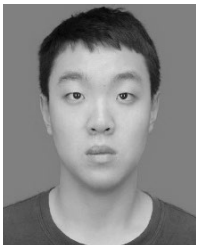
[26] Y. Wanxia, D. Taihang, and Z. Hongxing, "Fuzzy neural network model for forecas ting short-time traffic flow based on particle swarm optimization," *Microcomput. Inf.*, vol. 24, no. 4, pp. 24–25 and 185, 2008.

[27] Y. Licai, "Traffic information fusion algorithm of RBF network based on an artificial immune system and fuzzy clustering," *J. Shandong Univ., Eng. Ed.*, no. 5, pp. 114–118, 2008.

[28] K. Chen and J. Yu, "Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression approach," *Appl. Energy*, vol. 113, pp. 690–705, Jan. 2014.

[29] C. Shaokuan, W. Wei, M. Shaohua, and G. Wei, "Analysis on urban traffic status based on improved spatio-temporal Moran's I," *J. Phys.*, vol. 62, no. 14, pp. 527–533, 2013.

ZHI CHEN is currently pursuing the Ph.D. degree with the North China University of Technology, Beijing, China. Since 2011, he has been a Research Assistant with the Beijing Key Lab of Urban Intelligent Traffic Control Technology, North China University of Technology. His research interests include urban intelligent traffic control, big data mining, and traffic flow theory.



YUAN JIANG was born in 1994. He received the master's degree from the North China University of Technology. He was a Research Assistant with the North China University of Technology. His research interests include intelligent traffic and transportation research.

DEHUI SUN (Member, IEEE) was born in 1962. He received the Ph.D. degree from the University of Science and Technology Beijing. He is currently a Professor with the College of Electrical and Control Engineering, North China University of Technology. His research interests include urban intelligent traffic control, fieldbus technology and networked control theory, control technology and control engineering, embedded technology, and intelligent instrument development.