Hindawi Mathematical Problems in Engineering Volume 2019, Article ID 6571237, 10 pages https://doi.org/10.1155/2019/6571237



# Research Article

# A Road Network Traffic State Identification Method Based on Macroscopic Fundamental Diagram and Spectral Clustering and Support Vector Machine

# Xiaohui Lin

Institute of Rail Traffic, Guangdong Communication Polytechnic, Guangzhou, China

Correspondence should be addressed to Xiaohui Lin; linxh1981@163.com

Received 18 December 2018; Revised 27 March 2019; Accepted 10 April 2019; Published 21 April 2019

Academic Editor: Alexander Paz

Copyright © 2019 Xiaohui Lin. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate identification of road network traffic status is the key to improve the efficiency of urban traffic control and management. Both data mining method and MFD-based methods can divide the traffic state of road network, but each has its own advantages and disadvantages. The data mining method is oriented to traffic data with high efficiency, but it can only discriminate traffic status from microlevel, while the MFD of road network can discriminate traffic status from macrolevel, but there are still some problems, such as the fact that the discriminant method of equivalence points based on MFD lacks theoretical support or that traffic status could not be subdivided. If data mining methods and road network's MFD are combined, the accuracy of road network traffic state identification will be greatly improved. In addition, the research shows that the combination of unsupervised learning clustering analysis method (such as spectral clustering algorithm) and supervised learning machine algorithm (such as support vector machine algorithm (SVM)) is more accurate in traffic state identification. Therefore, a traffic state identification method based on MFD and spectral clustering and SVM is proposed, combining the advantages of spectral clustering algorithm and SVM algorithm. Firstly, spectral clustering algorithm is used to classify the traffic state of road network's MFD. Secondly, SVM multiclassifier is trained with the partitioned road network's MFD parameters, and the accuracy evaluation method of classification results based on obfuscation matrix is given. Finally, the connected-vehicle network simulation platform is built for empirical analysis. The results show that the classification results of spectral clustering algorithm are closer to the theoretical values, compared with K-means algorithm, and the accuracy of SVM multiclassifier is 96.3%. It can be seen that our algorithm can identify the road network traffic state more effectively from the macrolevel.

#### 1. Introduction

The traffic state of road network objectively reflects the traffic operation of road network, which is the key to improve the efficiency of urban traffic control and management. Traffic state identification method of road network has always been a research hotspot in the field of intelligent transportation system (ITS). Generally, it can be divided into two categories: the methods based on data mining and the methods based on traffic flow fundamental diagram.

1.1. The Methods Based on Data Mining. The methods based on data mining are to use machine learning algorithms such as neural network, deep learning, clustering algorithm,

support vector machine, and Bayesian method to mine data, so as to automatically identify the traffic state of the road network. For example, Mehdi (2010) et al. [1] used K-means, fuzzy c-means (FCM), clustering large applications (CLARA), and other three clustering methods to classify expressway traffic flow. The results showed that K-means clustering results are most consistent with highway capacity manual (HCM) classification. Montazeri-Gh (2011) et al. [2] used K-means algorithm to classify data sets such as average speed, acceleration, and percentage of idle time collected by floating cars, so as to discriminate the actual road network traffic state. Xia (2012) et al. [3] used Bayesian algorithm to identify highway traffic state based on traffic flow, speed, and occupancy data. Antoniou (2013) et al. [4] proposed a

location traffic state estimation method based on dynamic data driving, which used machine learning algorithm to cluster and classify traffic flow data. Yang (2014) et al. [5] used FCM clustering method to discriminate the traffic state of expressway based on the change of speed of big and small cars. Bing (2015) et al. [6] used projection pursuit technique and dynamic clustering method to construct projection index function between traffic parameters and traffic state in order to improve the accuracy of traffic state discrimination. The projection direction of projection index function was optimized by using the hybrid frog leaping algorithm. Finally, the threshold of traffic state discrimination was calibrated by simulation data. Zhang (2016) et al. [7] took speed, traffic flow, and traffic density as data sets and defined the relational relationship between multidimensional attribute information based on grey relational analysis and rough set theory, established the grey relational clustering model, and then introduced the grey relational membership ranking algorithm (GMRC) to discriminate the clustering priority, so as to analyze the degree of road network congestion. Shang (2017) et al. [8] constructed a traffic state discrimination model based on spectral clustering and stochastic subspace integration K-nearest neighbor (RS-KNN) in order to improve the accuracy of urban expressway traffic state discrimination.

1.2. The Methods Based on Traffic Flow Fundamental Diagram. The methods based on traffic flow fundamental diagram are to use the relationship between traffic flow and density, which presents a single peak parabola (called fundamental diagram) to distinguish the traffic state of the road network [9]. According to the fundamental diagram, the traffic state of the road network can be divided into two-phase traffic flow, i.e., congestion flow and noncongestion flow. Kerner (2002) et al. [10, 11] divided the traffic state of the road network into three phases: free flow, synchronous flow, and wide moving congestion according to the measured data based on the two-phase traffic flow. Guan (2007) et al. [12] divided traffic flow into four steady-state phases: free flow, harmonic flow, synchronous flow, and congestion based on the analysis of the general distribution characteristics of traffic flow speed and density in urban expressway network. Traffic state identification based on fundamental diagram has been developing slowly. With the revelation of macroscopic fundamental diagram (MFD), some scholars used MFD to discriminate road network traffic from macroscopic level. For example, Wang (2012) et al. [13] proposed a traffic condition discrimination method based on equal points of MFD parameters (referred to as "MFD equal points discrimination method"). This method determined the road network's MFD through simulation data and divided the traffic state of the road network into five congestion levels by equal points, i.e., free flow, stable flow, unstable flow, restricted flow, and forced flow. However, it directly divided traffic status by equal points, which lacked theoretical support. Zhu (2012) et al. [14] used the measured and simulated data to establish the actual road network's MFD and calibrated the traffic flow parameters of the road network and then studied the difference of the

distribution of average traffic flow and average density in the period of time, so as to determine the traffic state of the road network, i.e., free flow, congestion flow, and oversaturated flow. However, the three traffic conditions cannot be further subdivided. Xu (2013) et al. [15] divided the traffic state of the road network into free flow, cumulative flow, and congestion flow according to the observed MFD. However, there are only three traffic conditions. Yue (2014) et al. [16] established the actual road network's MFD model with the help of traffic data of remote traffic detectors and designed the macrotraffic state index of expressway based on speed mileage distribution and divided the traffic state of expressway into five grades, including smooth, basic smooth, mild congestion, moderate congestion, and congestion. However, this method only discussed and verified the feasibility of expressway traffic state discrimination, and it did not discuss whether it was suitable for the traffic state division of the whole road network. Ding (2018) et al. [17] estimated the MFD of expressway network based on floating car detection data and divided the expressway network into five states initially according to the MFD equal points discrimination method [13]. Finally, the state parameters of the network are further modified by clustering algorithm according to the real-time data of the network. Similarly, it directly divided traffic status by equal points, which lacked theoretical support.

However, both data mining method and MFD-based methods can divide the traffic state of road network, but each has its own advantages and disadvantages. The data mining method is oriented to traffic data with high efficiency, but it can only discriminate traffic status from microlevel, while the MFD of road network can discriminate traffic status from macrolevel, but there are still some problems, such as the fact that the discriminant method of equivalence points based on MFD lacks theoretical support [13, 17] or that traffic status could not be subdivided [14, 15] or that the application scope of the method is limited [16]. If data mining methods and MFD can be combined, the accuracy of road network traffic state identification will be greatly improved. With the rapid development of large data mining algorithms, K-means, FCM, spectral clustering, and other clustering analysis methods have emerged, and they have been widely used in traffic state discrimination. At the same time, supervised learning machine algorithms, represented by artificial neural network (ANN) and support vector machine (SVM), are also applied in traffic condition discrimination. Recent studies have shown that the combination of clustering analysis method and supervised learning machine algorithm is more accurate in traffic state identification [4, 18]. Clustering analysis provides necessary prior information for supervised learning algorithm, and supervised learning algorithm can ensure the real-time traffic status discrimination [8]. Among them, spectral clustering is a new clustering method based on spectral graph theory. It has been widely used in speech recognition, video segmentation, image segmentation, VLSI design, web page partitioning, text mining, and so on, but its application in the field of transportation has just begun. SVM method, which has a rigorous mathematical basis, has become a hot technology in the fields of clustering analysis, pattern recognition, state discrimination, prediction, and regression analysis and has attracted wide attention of scholars at home and abroad. Spectral clustering algorithm and SVM algorithm have been widely used in many fields because of their strong mathematical theory support, and many examples have proved their good application effect. Therefore, in view of the fact that the discriminant method of equivalence points of MFD lacks theoretical support and traffic status could not be subdivided, this paper proposes a road network traffic state identification method based on MFD and spectral clustering and SVM, combining the advantages of spectral clustering algorithm and SVM algorithm. Its basic ideas are as follows.

Firstly, FCD method is used to estimate road network's MFD and then spectral clustering algorithm is used to classify road network's MFD. The traffic state of the road network is divided into four traffic states, i.e., smooth, stationary, congested, and oversaturated. Then, SVM multiclassifier is trained with the divided MFD parameters of the road network, and the accuracy of the classification results is evaluated by the confusion matrix. Finally, the core road network intersection group of Guangzhou is taken as a research area. The microscopic traffic simulation model is built up by using the Vissim traffic simulation software, and the validity of the algorithm is validated. The specific process is shown in Figure 1.

### 2. Traffic State Identification Method Based on MFD

The concept of MFD was first proposed by Godfrey (1969) [19], but its theoretical principle was not revealed by Daganzo and Geroliminis [20, 21] until 2007. These two scholars believe that MFD is an inherent attribute of road network, which objectively reflects the general relationship between weighted traffic flow and weighted traffic density of road network. If the floating car is evenly distributed in the road network and its coverage is known, then the road network's MFD can be estimated by using the floating car data (FCD). The formula is as follows [22]:

$$k^{w} = \frac{\sum_{j=1}^{m'} t'_{j}}{\rho * T * \sum_{i=1}^{n} l_{i}}$$
(1)

$$q^{w} = \frac{\sum_{j=1}^{m'} d'_{j}}{\rho * T * \sum_{i=1}^{n} l_{i}}$$
 (2)

where  $k^w$  and  $q^w$  are the weighted traffic density (veh/km) and the weighted traffic flow (veh/h) of the road network obtained by the FCD estimation method, respectively;  $\rho$  is the ratio of floating cars; T is the acquisition cycle(s); n is total number of sections of the road network;  $l_i$  is the length of road section i (km); m' is the number of floating cars during the acquisition cycle T (veh);  $t'_j$  is the driving time of the j-th floating car during the acquisition cycle T(s); and  $d'_j$  is the driving distance of the j-th floating car during the acquisition cycle T(m).

Road network's MFD can monitor the traffic state of the road network from the macrolevel, as shown in Figure 2.

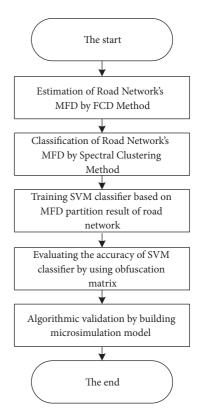


FIGURE 1: The basic research ideas.

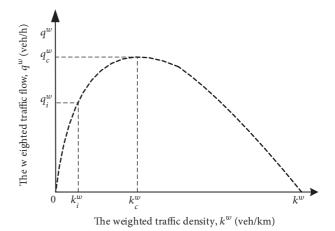


FIGURE 2: The schematic diagram of road network's MFD.

As shown in Figure 2, when the number of vehicles in the road network is less, the weighted traffic density and weighted traffic flow of the road network are small. With the increase of vehicles in the road network, the weighted traffic density of the road network increases, and the weighted traffic flow increases with the increase of the weighted traffic density. At this time, the traffic state of the road network is defined as smooth state. With the continuous increase of traffic flow in the road network, the weighted traffic density and the weighted traffic flow in the road network are also rising, but the traffic flow in the road network is relatively smooth. At this time, the traffic state of the road network is defined as

a stable state. When the traffic flow keeps entering the road network, the weighted traffic density of the road network rises again, and the traffic flow in the road network will interfere with each other. The weighted traffic flow of the road network will continue to increase with the weighted traffic density of the road network. When the weighted traffic density of the road network continues to rise to a certain range, the change of the vehicle speed in the road network will be unstable. The traffic flow of the road network reaches the maximum, and the traffic state of the road network is defined as the congestion state. When vehicles continue to pour into the road network, the weighted traffic density of the road network continues to increase, the speed of the road network and the weighted traffic flow continue to decline, the traffic demand has exceeded the supply capacity of the road network, the efficiency of the traffic operation has been declining, and the road network has entered a state of oversaturation. At this time, the traffic state of the road network is defined as oversaturated state. As the traffic flow continues to increase, the traffic density continues to increase, reaching the road network congestion density, the speed and traffic flow of the road network are close to zero, and the whole road network is paralyzed.

According to the road network's MFD, the traffic state of the road network is divided into four grades: smooth, stable, congested, and oversaturated. In order to quantify the traffic state of the road network, Wang (2012) et al. [13] analyzed the relationship among flow Q, density K, and speed V based on the road network's MFD and solved the extreme value of the road network operation index. The critical point between steady flow and unstable flow is determined as  $V_c = (2/3)V_f$ ,  $K_c = (1/3)K_j$ , where  $V_c$  is the critical velocity,  $K_c$  is the critical density,  $V_f$  is the free flow velocity, and  $K_i$  is the blocking density, and then the traffic state of the road network is divided into five grades directly according to the equivalence of velocity and density. The traffic state of the road network is free flow, stable flow, unstable flow, restricted flow, and forced flow. The corresponding evaluation indicators are shown in Table 1.

In theory, it is feasible to divide the traffic state of the road network into five grades according to the traffic density or speed of the road network. However, this method directly divides the traffic density or speed of the road network by equal points, lacking theoretical support, and whether it is reasonable needs further verification.

# 3. Road Network Traffic State Identification Method Based on MFD and Spectrum Clustering and SVM

3.1. The Division Process of Road Network's MFD Based on NJW Spectral Clustering Algorithm. Spectral clustering (SC) algorithm is a research hotspot in the field of machine learning. It is a point-to-point clustering algorithm based on spectral graph theory, which transforms the data clustering problem into the optimal graph partition problem. SC algorithm is suitable for spatial clustering problems with arbitrary shapes, compared with typical clustering algorithms

such as K-means clustering and FCM clustering, and it can obtain global optimal solutions. The graph theory based on partition criterion is most closely related to spectral clustering results. The common partition criteria are mini cut, average cut, normalized cut, min-max cut, ratio cut, MN cut, and so on. SC algorithm can be divided into two categories: twoway SC algorithm and multiway SC algorithm according to the partition criterion. Among them, two-way SC algorithm includes PF algorithm, SM algorithm, SLH algorithm, KVV algorithm, M cut algorithm, and so on. Multiway SC algorithm includes Ng-Jordan-Weiss (NJW) algorithm [23] and MS algorithm; NJW algorithm has the best effect, compared with the application effect of various SC algorithms. The basic idea of NJW algorithm is to calculate the similarity matrix of sample data sets first and then transform the similarity matrix into Laplace matrix and then construct a new vector space R by using the eigenvectors corresponding to the first k largest eigenvalues of Laplace matrix and build a one-to-one construction matrix U corresponding to the original data in the new *R* space. Finally, cluster the construction matrix *U* by using K-means algorithm [23]. Therefore, this paper will use NJW algorithm to divide road network's MFD; the specific process is as follows.

(1) It needs to define the sample data set and then determine the number of clusters K, K=4 (i.e., smooth state, stable state, congestion state, and oversaturated state); the formula is as follows:

$$X = \{x_i \mid i = 1, 2, \dots, n\}$$
 (3)

$$x_i = \{x_{i1}, x_{i2}\} \tag{4}$$

where X is the sample data set;  $x_i$  is the i-th scatter point on road network's MFD;  $x_{i1}$  is the i-th weighted traffic density of road network; and  $x_{i2}$  is the i-th weighted traffic flow of road network.

(2) Normalization of sample data: the formula is as follows:

$$x'_{ij} = \frac{x_{ij} - x_{j\min}}{x_{j\max} - x_{j\min}}$$
 (5)

where  $x_{j\max}$  and  $x_{j\min}$  are the maximum and minimum values of the *j-th* eigenvector, respectively;  $x_{ij}$  is the initial value of the first *i-th* element of the *j-th* eigenvector; and  $x'_{ij}$  is the normalized value of normalization of the *i-th* element of the *j-th* eigenvector.

(3) To calculate the similarity matrix A, the formula is as follows:

$$A_{ij} = \begin{cases} \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right), & i \neq j \\ 0, & i = j \end{cases}$$
 (6)

where  $\|x_i - x_j\|$  is the Euclidean distance between sample  $x_i$  and sample  $x_j$  and  $\sigma$  is the standard deviation of samples; the value is 0.9 in this algorithm.

(4) To calculate the Laplace matrix L, the formula is as follows:

$$L = D^{-1/2} A D^{-1/2} (7)$$

	Free flow	steady flow	unsteady flow	constrained flow	forced flow	
Density	$[0, 1/6)K_{j}$	$[1/6, 2/6)K_j$	$[2/6, 4/6)K_j$	$[4/6, 5/6)K_{j}$	$[5/6, 1]K_i$	
Speed	$[1,5/6)V_f$	$[5/6, 4/6)V_f$	$[4/6, 2/6)V_f$	$[2/6, 1/6)V_f$	$[1/6, 0]V_f$	

TABLE 1: The identification of traffic state and the classification of congestion level.

where D is the diagonalization matrix of similar matrix A, which satisfies the following conditions:

$$D_{ij} = \sum_{j} A_{ij}. (8)$$

- (5) The first K maximum eigenvalues  $(\lambda_1, \lambda_2, \ldots, \lambda_k)$  and corresponding eigenvectors  $(u_1, u_2, \ldots, u_k)$  of Laplacian matrix L are calculated. The eigenvectors are arranged in descending order according to the size of eigenvalues, and the matrix U is constructed, which is expressed as  $U = \{u_1, u_2, \ldots, u_k\} \in R^{n \times k}$ .
- (6) To normalize the row vector of matrix U to get matrix Y, the formula is as follows:

$$Y_{ij} = \frac{u_{ij}}{\sqrt{\sum_{j} u_{ij}^2}} \tag{9}$$

- (7) To take every row vector  $y_{ij} \in \mathbb{R}^k$  (i = 1, 2, ..., n) in the Y matrix as a point, K-means algorithm is used to cluster  $y_i$ , and then it gets K clustering, which are expressed as  $C_1, C_2, ..., C_k$ .
- (8) Output clustering results, which are expressed as  $A_1, A_2, \ldots, A_k, A_i = \{j \mid y_j \in C_i\}.$
- 3.2. Design of SVM Multiple Classifiers. SVM was proposed by Cortes and Vapnik in 1995, which is mainly used to solve the problem of data classification. This method has strict theoretical basis and belongs to supervised machine learning method. It has been widely used in many fields, such as pattern recognition, data classification, data mining, computational intelligence, prediction and analysis, traffic condition identification, traffic sign recognition, vehicle type recognition, traffic demand prediction, traffic accident severity, pedestrian detection, and so on.

SVM multiclassifier is implemented by the LIBSVM software package developed by Professor Lin Chih-Jen of Taiwan University. This software package belongs to One-Versus-One SVMs, whose basic principle is to combine K samples in pair arbitrary combinations, design a SVM for each combination sample, and form k\*(k-1)/2 SVMs. When it classifies an unknown sample, each SVM decides the category of the unknown sample, votes on the corresponding category, and finally decides that the category of the unknown sample is the category with the most votes. The basic idea of each SVM classifier is to find an optimal hyperplane and segment the sample data according to positive and negative examples to maximize the classification margin. The algorithm flow is as follows [24].

It assumes that there is a training sample set  $\{x_i, y_i \mid i = 1, 2, ..., l\}$ ,  $x \in \mathbb{R}^n$ ,  $y \in \{1, -1\}$ , where l is the number of samples and n is the dimension of x.

(1) The training sample set is linearly separable, that is, there exists a plane which can divide the training sample into positive and negative categories. This plane is called hyperplane, and its classification function formula is as follows:

$$q(x) = w^{T}x + b = 0 (10)$$

where w is the slope of the hyperplane and  $w^T$  is the transposition of the slope; b is a constant; x is a multidimensional vector.

At this time, the classification interval is  $2/\|w\|$ . When  $\|w\|$  is the smallest, the classification interval is the largest. The problem is transformed into a minimization problem for easy solution. The objective function and constraints are established as follows:

min 
$$\frac{\|w\|^2}{2}$$
  
s.t.  $y_i(wx_i + b) - 1 \ge 0, \quad i = 1, 2, ..., l$  (11)

(2) When the training sample set is linearly inseparable, it needs to introduce the kernel function K to map the low-dimensional sample set to the high-dimensional one, so that the results calculated in the low-dimensional space are the same as those calculated in the high-dimensional space. Then the linearly inseparable is transformed into linearly separable. Its classification function formula is as follows:

$$f(x) = \sum_{i=1}^{l} a_i y_i K(x, x_i) + b = 0$$
 (12)

where K is a kernel function. The common kernels are polynomial kernels, Gauss kernels, linear kernels, string kernels, etc.;  $a_i$  is the Lagrange coefficient.

There may be data points that are far from the normal position or the outlier data points of Category 2 are mixed in Category 1 area. In order to deal with this situation, individual data points are allowed to deviate from the hyperplane to some extent; i.e., nonnegative relaxation variable  $\xi_i$  is introduced. The original objective function and constraints are transformed into

min 
$$\frac{\|w\|^2}{2} + C \sum_{i=1}^{l} \xi_i$$
 (13)  
s.t.  $y_i (wx_i + b) \ge 1 - \xi_i, \quad i = 1, 2, ..., l$ 

where C is the penalty coefficient, indicating the importance of the loss caused by outlier sample data points. The smaller the C value, the smaller the penalty for misclassification and the smaller the loss to the objective function.

	The predicted value				
The actual value	1	2	•••	n	The total line
1	$x_{11}$	$x_{12}$		$x_{1n}$	$x_{1+}$
2	$x_{21}$	$x_{22}$	•••	$x_{2n}$	$x_{2+}$
•••	•••	•••	•••	•••	• • •
n	$x_{n1}$	$x_{n2}$		$x_{nn}$	$\boldsymbol{x}_{n+}$
The total column	$x_{+1}$	$x_{+2}$		$x_{+n}$	N

TABLE 2: The confusion matrix of the classification results of SVM multiple classifiers.

*3.3. The Accuracy Evaluation of SVM Multiple Classifiers.* The test classification results of the SVM classifier are transformed into the confusion matrix, as shown in Table 2.

According to the confusion matrix, the classification accuracy of SVM multiclassifier can be evaluated from five evaluation indicators: accuracy, recall, omission, prediction, and commission.

(1) Accuracy. The accuracy rate is the correct ratio in all test samples; the formula is as follows:

$$Accuracy = \sum_{k=1}^{n} \frac{x_{kk}}{N}$$
 (14)

(2) Recall. The actual recall rate refers to the proportion of the correct predicted value of the actual category to the total number of the actual categories in the *i-th* actual classification; the formula is as follows:

$$Recall_i = \frac{x_{ii}}{x_{i+}} \tag{15}$$

(3) Omission. The actual omission rate refers to the proportion of the actual category value to the total number of the actual category value in the *i-th* actual classification; the formula is as follows:

$$omission_i = \frac{x_{i+} - x_{ii}}{x_{i+}} \tag{16}$$

(4) *Precision*. The prediction accuracy rate is the proportion of the correct value of the prediction to the total number of samples of the prediction category in the *j-th* prediction category; the formula is as follows:

$$Precision_j = \frac{x_{jj}}{x_{+j}}$$
 (17)

(5) Commission. The commission refers to the ratio of the value of the error prediction to the total number of samples in the *j-th* prediction category; the formula is as follows:

$$commission_j = \frac{x_{+j} - x_{jj}}{x_{+j}}$$
 (18)

### 4. Empirical Analysis

4.1. Experimental Platform Construction. The intersections of Tianhe District core road network in Guangzhou are

chosen as the experimental area [25], as shown in Figure 3. A microtraffic simulation platform for connected-vehicle network is built by using VISSIM simulation software and based on the road network layout, the actual road lane layouts, the intersection plane layout, the traffic flow data, the intersection signal control schemes, the traffic organizations, and other information.

The MFD estimation accuracy of road network can reach 97% when the coverage of networked vehicles reaches 42% after many simulations. Therefore, the coverage rate of networked vehicles is set to 42%, and the related data of each networked vehicle are read every 15 seconds. The statistical period of data is 120 seconds, and the simulation time is 3240 seconds. The simulation results of the online vehicle data file (\*. fzp) are imported into EXCEL file, and the FCD estimation method is realized by VBA macroprogramming. Finally 270 sets of the weighted traffic flow  $(q^w)$  and the weighted traffic density  $(k^w)$  are obtained, and the MFD of the simulation network is drawn, as shown in Figure 4.

The data points of road network's MFD are fitted by function, and the critical weighted traffic density  $(k_c)$  and the critical weighted traffic flow  $(q_c)$  are calculated, as shown in Table 3.

#### 4.2. Analysis of Experimental Results

(1) The Partition Result of Road Network's MFD Based on NJW Spectral Clustering Algorithm. The NJW algorithm is programmed in Matlab software. The parameter of the road network's MFD is input into the program, and the result of traffic state division is output, as shown in Figure 5. In order to better analyze the effect of spectral clustering algorithm, K-means algorithm is selected as a comparison to cluster the MFD parameters of road network, as shown in Figure 6.

As shown in Figures 5 and 6, the partition result of road network's MFD based on K-means algorithm only considers the size of MFD parameters but does not consider the specific location of MFD scatters. It incorrectly incorporates part of the congestion scatters into the oversaturated state and individual oversaturated scatters into the congested state, while the partition result of road network's MFD based on spectral clustering is more reasonable. It can distinguish traffic conditions into four better states, i.e., smooth, stationary, congested, and oversaturated. The scope of traffic state partition of the two clustering methods is obtained by Figures 5 and 6. At the same time, according to the MFD fitting function of the whole road network in Table 1, the equal

TABLE 3: The fitting function of MFD.

Fitting formula	$k_c$ (veh/km)	$q_c$ (veh/h)
$y = 0.00990x^3 - 1.5525x^2 + 77.591x - 115.92$	41.32	1138

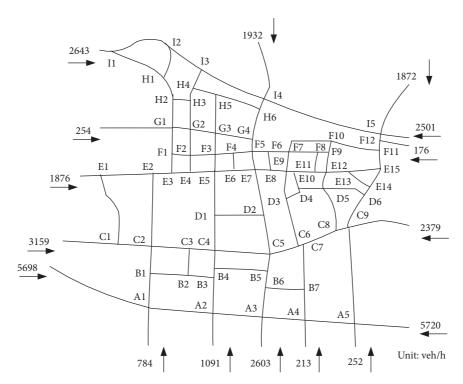


FIGURE 3: Road network layout diagram of vehicle networking simulation platform.

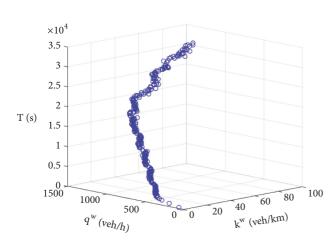
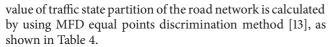


FIGURE 4: The 3D diagram of simulation road network's MFD.



As shown in Table 4, compared with the K-means-based traffic state partition results, the result of road network traffic state partition based on spectral clustering is close to the equal value. It shows that the results of spectral clustering algorithm and equivalence method are feasible, but

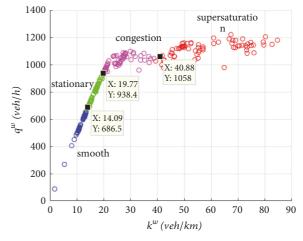


FIGURE 5: The division results of road network's MFD based on spectral clustering.

the equivalence method is directly divided by equivalence points. It is too idealized and lacks theoretical support. The spectral clustering algorithm has detailed and reliable theory as support, and it is more scientific.

(2) SVM Multiclassifier Training. According to the above results of road network traffic state partition based on spectral clustering, 270 sets of parameters of road network's MFD are

Traffic state	Partition based on K-means algorithm	Partition based on spectral clustering algorithm	Equal value
Smooth	[0, 12.26)	[0, 14.09)	[0, 10.83)
Stationary	[12.26, 16.9)	[14.09, 19.77)	[10.83, 21.67)
Congestion	[16.9, 20.69)	[19.77, 40.88)	[21.67, 43.33)
Oversaturated	$[20.69, k_j)$	$[40.88, k_j)$	$[43.33, k_j)$

TABLE 5: The confusion matrix and the accuracy evaluation index of the classification results by SVM multiclassifiers.

				predicted valu	e		
Actual Value	1	2	3	4	total	recall	omission
1	28	1	0	2	31	90.3%	9.7%
2	0	35	1	0	36	97.2%	2.8%
3	0	0	33	1	34	97.1%	2.9%
4	0	0	0	34	34	100.0%	0.0%
total	28	36	34	37	135	96.3%	3.7%
precision	100.0%	97.2%	97.1%	91.9%			
commission	0.0%	2.8%	2.9%	8.1%			

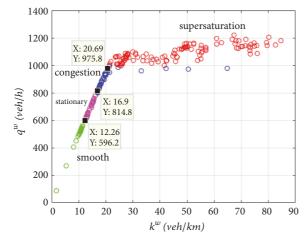
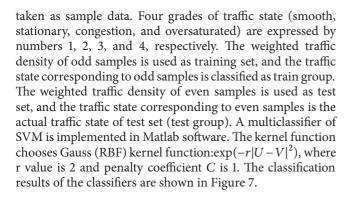
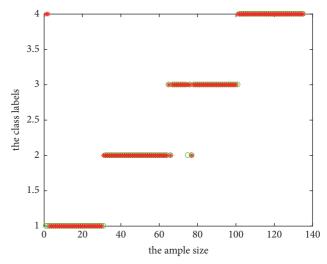


FIGURE 6: The division results of road network's MFD based on K-means.



(3) Accuracy Evaluation of SVM Multiclassifier. The classification result of the classifier is sorted out into confusion matrices, and the accuracy evaluation indexes of the classifier are calculated, as shown in Table 5.



- the classification of the actual test sets
- \* the classification of the predictive test sets

FIGURE 7: The classification results of SVM multiclassifier.

As shown in Table 5, it can be seen that the prediction accuracy of Category 1 is 100%, the actual recall rate is 90.3%, the prediction accuracy of Category 2 is 97.2%, the actual recall rate is 97.2%, the prediction accuracy of Category 3 is 97.1%, the actual recall rate is 97.1%, the prediction accuracy of Category 4 is 91.9%, and the actual recall rate is 100%. From the overall prediction results, the accuracy of SVM multiclassifier is 96.3%. Therefore, the results of SVM multiclassifier are ideal.

#### 5. Conclusion

In view of the fact that the discriminant method of equivalence points of MFD lacks theoretical support and traffic status could not be subdivided, this paper proposes a road

network traffic state identification method based on MFD and spectral clustering and SVM, combining the advantages of spectral clustering algorithm and SVM algorithm. Based on the results of empirical analysis, this paper draws the following conclusions:

- (1) Spectral clustering-based traffic state partition results are closer to the theoretical values, compared with K-means-based traffic state partition results.
- (2) The accuracy of SVM multiclassifier is up to 96.3%; it can be seen that the classification results are ideal.
- (3) This paper combines the advantages of spectral clustering algorithm and SVM algorithm, which can identify the traffic state of road network more effectively and automatically from the macrolevel on the basis of MFD.
- (4) It is noteworthy that this paper estimates the road network's MFD by using the data of the simulation environment of the networked vehicle and assumes that the coverage rate of the networked vehicle is as high as 42% and is evenly distributed in the road network, but, in the actual road network, it is impossible to obtain such a high coverage rate of the networked vehicle, and there is an uneven distribution of the networked vehicle. Therefore, in practical application, a large amount of traffic data is needed to accurately estimate road network's MFD.

## **Data Availability**

The data used to support the findings of this study are available from the corresponding author upon request.

### **Conflicts of Interest**

The author declares that there are no conflicts of interest.

### Acknowledgments

This paper is jointly funded by Natural Science Foundation of Guangdong Province (2016A030313786), Guangdong Province higher education outstanding young teachers training program in 2015 (YQ2015184), special innovative research projects of Guangdong Provincial Department of Education in 2017 (2017GKTSCX018), key scientific research projects of Guangdong Communication Polytechnic in 2018 (2018-01-001), special fund project for Guangdong university students' science and technology innovation in 2018 (pdjhb0756), and Guangdong science and technology project in 2016 (2016A020210123).

#### References

- [1] M. Azimi and Y. Zhang, "Categorizing freeway flow conditions by using clustering methods," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2173, no. 1, pp. 105–114, 2010.
- [2] M. Montazeri-Gh and A. Fotouhi, "Traffic condition recognition using the k-means clustering method," *Scientia Iranica*, vol. 18, no. 4, pp. 930–937, 2011.

- [3] J. Xia, W. Huang, and J. Guo, "A clustering approach to online freeway traffic state identification using ITS data," *KSCE Journal of Civil Engineering*, vol. 16, no. 3, pp. 426–432, 2012.
- [4] C. Antoniou, H. N. Koutsopoulos, and G. Yannis, "Dynamic data-driven local traffic state estimation and prediction," *Trans*portation Research Part C: Emerging Technologies, vol. 34, pp. 89–107, 2013.
- [5] Q. F. Yang, M. H. Ma, S. D. Liang et al., "Freeway traffic state identification based on toll data," *Journal of South China University of Technology (Natural Science Edition)*, vol. 42, no. 12, pp. 51–57, 2014.
- [6] Q. Bing, B. Gong, Z. Yang, C. Lin, and X. Qu, "Traffic state identification for urban expressway based on projection pursuit dynamic cluster model," *Xinan Jiaotong Daxue Xuebao/Journal* of Southwest Jiaotong University, vol. 50, no. 6, pp. 1164–1169, 2015.
- [7] Y. Zhang, N. Ye, R. Wang, and R. Malekian, "A method for traffic congestion clustering judgment based on grey relational analysis," *ISPRS International Journal of Geo-Information*, vol. 5, no. 5, p. 71, 2016.
- [8] Q. Shang, C. Y. Lin, Z. S. Yang et al., "Traffic state identification for urban expressway based on spectral clustering and RS-KNN," *Journal of South China University of Technology (Natural Science Edition)*, vol. 45, no. 6, pp. 52–58, 2017.
- [9] W. Wei, Analytical Method of Traffic Condition Characteristic and Congestion Propagation Rules Based on Practical Measured Data, Beijing Jiaotong University, 2017.
- [10] B. S. Kerner, S. L. Klenov, and D. E. Wolf, "Cellular automata approach to three-phase traffic theory," *Journal of Physics A: Mathematical and General*, vol. 35, no. 47, pp. 9971–10013, 2002.
- [11] B. S. Kerner, "Three-phase traffic theory and highway capacity," Physica A: Statistical Mechanics and Its Applications, vol. 333, no. 1, pp. 379–440, 2002.
- [12] W. Guan and S. He, "Phase identification of urban freeway traffic based on statistical properties," *Journal of Transportation Systems Engineering and Information Technology*, vol. 7, no. 5, pp. 42–50, 2007.
- [13] F. J. Wang, W. Wei, and D. H. Wang, "Urban road network traffic state identification and monitoring based on macroscopic fundamental diagram," in *Proceedings of the presented at the 7nd Annual Conference of ITS China*, Beijing, China, 2012.
- [14] L. Zhu, L. Yu, and G. H. Song, "MFD-based investigation into macroscopic traffic state of urban networks and its influencing factors," *Journal of South China University of Technology (Natu*ral Science Edition), vol. 40, no. 11, pp. 138–146, 2012 (Chinese).
- [15] F. Xu, Z. He, Z. Sha, W. Sun, and L. Zhuang, "Traffic state evaluation based on macroscopic fundamental diagram of urban road network," *Procedia - Social and Behavioral Sciences*, vol. 96, no. 1, pp. 480–489, 2013.
- [16] Y. Y. Yue, L. Yu, L. Zhu et al., "Macroscopic model for evaluating traffic conditions on the expressway based on speedspecific VKT distributions," *Journal of Transportation Systems Engineering and Information Technology*, vol. 14, no. 4, pp. 85– 92, 2014.
- [17] H. Ding, L. Y. Zhu, C. B. Jiang et al., "Traffic state identification for cities freeway network based on MFD," *Journal of Chongqing Jiaotong University (Natural Science)*, pp. 1–8, 2018.
- [18] C. Deng, F. Wang, H. Shi, and G. Tan, "Real-time freeway traffic state estimation based on cluster analysis and multiclass support vector machine," in *Proceedings of the International Workshop on Intelligent Systems and Applications, ISA '09*, China, 2009.

- [19] J. W. Godfrey, "The mechanism of a traffic network," *Traffic Engineering & Control*, vol. 11, no. 7, pp. 323–327, 1969.
- [20] C. F. Daganzo, "Urban gridlock: macroscopic modeling and mitigation approaches," *Transportation Research Part B: Meth-odological*, vol. 41, no. 1, pp. 49–62, 2007.
- [21] N. Geroliminis and C. F. Daganzo, "Existence of urban-scale macroscopic fundamental diagrams: some experimental findings," *Transportation Research Part B: Methodological*, vol. 42, no. 9, pp. 759–770, 2008.
- [22] A. S. Nagle and V. V. Gayah, "Accuracy of networkwide traffic states estimated from mobile probe data," *Transportation Research Record Journal of the Transportation Research Board*, vol. 2421, no. 2421, pp. 1–11, 2014.
- [23] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: analysis and an algorithm," in *Advances in Neural Information Processing Systems*, vol. 14, MIT Press, 2001.
- [24] D. Wang, C. Lu, W. Jiang et al., "Study on PSO-based decisiontree SVM multi-class classification method," *Journal of Electronic Measurement and Instrumentation*, vol. 29, no. 4, pp. 611– 615, 2015.
- [25] X. Lin, J. Xu, and W. Zhou, "A dynamic partitioning method for the multilayer boundary for urban oversaturated homogeneous traffic networks based on macroscopic fundamental diagrams in connected-vehicle network," *Advances in Transportation Studies*, in press.

















Submit your manuscripts at www.hindawi.com























