

Research Article

Application of Data Science Technologies in Intelligent Prediction of Traffic Congestion

Xu Yang , **Shixin Luo**, **Keyan Gao**, **Tingting Qiao**, and **Xiaoya Chen**

School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China

Correspondence should be addressed to Xu Yang; yangxu@tsinghua.edu.cn

Received 22 November 2018; Accepted 8 April 2019; Published 15 April 2019

Academic Editor: Jose E. Naranjo

Copyright © 2019 Xu Yang et al. 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.

In recent years, with the rapid development of economy, more and more urban residents, while owning their own motor vehicles, are also troubled by the traffic congestion caused by the backward traffic facilities or traffic management methods. The loss of productivity, car accidents, high emissions, and environmental pollution caused by traffic congestion has become a huge and increasingly heavy burden on all countries in the world. Therefore, the prediction of urban road network traffic flow and the rapid and accurate evaluation of traffic congestion are of great significance to the study of urban traffic solutions. This paper focuses on how to apply data science technologies on vehicular networks data to present a prediction method for traffic congestion based on both real-time and predicted traffic data. Two evaluation frameworks are established, and existing methods are used to compare and evaluate the accuracy and efficiency of the presented method.

1. Introduction

The problem of urban traffic congestion, from the beginning of the birth of motor vehicles to the widespread popularity of motor vehicles, has been an important issue that puzzles countries all over the world [1]. A traffic economic theory analysis considers that the upgrading of basic transportation facilities is an efficient way to solve congestion. However, in fact, this problem cannot be completely solved by simply building better roads and facilities [2–4]. The Intelligent Transportation System (ITS) technology, which has emerged in recent years, has proved to be a scientific and efficient solution [5]. ITS widely uses information technology, traffic engineering, and behavioral science to reveal the rules of urban traffic, calculate traffic flow in real time, and try to guide vehicles and their surrounding vehicles to avoid traffic congestion. ITS has promoted the emergence of various solutions to modern urban traffic problems [5–9].

The rapid development and deployment of ITS in recent years has produced a lot of real-time and historical traffic information. Generally speaking, traffic data are obtained using equipment and systems, such as GPS and GIS. Real-time traffic data can be used to estimate the traffic conditions of the surrounding roads for drivers, such as road

congestion. Accurate and fast traffic information prediction is very important for many applications in Advanced Traffic Information System (ATIS) and vehicle energy management [10–13]. In ATIS, if traffic congestion information can be accurately predicted, it can be used to reduce the uncertainty of future traffic state, improve traffic mobility, and provide drivers with reliable travel time estimates, expected delays, and optional routes to their destinations. Electric and hybrid electric vehicles are the future trend of motor vehicle development. Complying with this trend, if we can achieve accurate and rapid prediction of congestion on the expected route, then it can be very helpful to achieve optimal fuel economy on the vehicle, as well as to estimate the range of the vehicle [11, 14–18].

Modern traffic control and guidance systems are usually designed for large-scale road networks, requiring real-time, high-resolution data. And the current traffic research still cannot fully express the essential rules of traffic congestion formation, nor can it predict the sudden traffic congestion. Traffic flow in urban traffic is a complex dynamic nonlinear process, and traffic congestion is the aggregation of traffic flow in limited time and space. In 2009, Flynn et al. [19] carried out theoretical work on the use of macro-traffic flow theory to model traffic congestion and obtained some basic

results of congestion prediction. The result is thought to be a creative solution to the traffic equation in -50s. Based on current research, traffic congestion in road networks is expected to be accurately evaluated in real time and used to support traffic control.

This paper focuses on the development of methods to accurately and quickly predict traffic congestion in a specific road network, based on data from vehicle networks. In this paper, we have shown a kind of referential congestion degree evaluation method. The main reference value of congestion evaluation is the occupancy rate of the road in a specific time period.

The implementation of this method is based on two kinds of traffic big data: one is road real-time traffic information; the other is the prediction of the traffic flow of the corresponding road network based on the traffic flow prediction algorithm [20]. The accumulated real-time data will be used alone or in conjunction with the corresponding prediction data to form the big data group. Our method will use these two types of data to obtain the congestion degree of the corresponding road network quickly and accurately and to judge the congestion degree of the current and future road networks.

The contribution of our work includes the following:

- (1) A congestion evaluation algorithm based on road occupancy rate;
- (2) An evaluation framework for comparison of prediction of road congestion;
- (3) A visual evaluation framework for comparison of prediction of road congestion based on Simulation of Urban MObility (SUMO) [21] and its java branch SUMODLER.

Both frameworks can use a large amount of data to evaluate and compare existing methods and our congestion evaluation methods. The first framework focuses on the horizontal comparison between the congestion evaluation algorithm described in this paper and several different algorithms under the same road conditions. It has a uniform input and output. Different congestion evaluation algorithms are used to analyze the input data, and the congestion level is obtained. The unified criterion is also reclassified according to its original criteria to facilitate the horizontal comparison between the algorithms. The second framework, the visual evaluation framework, focuses on evaluating the performance of the congestion evaluation algorithm mentioned in this paper on the overall congestion evaluation of the road network. The XML file exported by SUMO is used. By using the congestion evaluation algorithm described in this paper, the congestion grades of each road are divided, and the change of congestion grade of all roads in the concerned time period is plotted in the road network.

The rest of the paper is organized as follows: Section 2 discusses related works; Section 3 will give introduction of background. The framework of our algorithm is presented in Section 4. Experiment results are presented in Section 5. We give final conclusion in Section 6.

2. Related Works

The existing criteria for road congestion evaluation are mainly applied uniformly across countries, regions, cities, etc, typical examples of which are as follows:

- (1) *Road Congestion Evaluation Standard Based on Road Speed*. INRIX Index is a road congestion evaluation standard that has been used in most European countries in recent years. The standard is calculated on the basis of the speed of the vehicle flow on the road. INRIX metrics in specified intervals (≤ 15 minutes). The index is based on the ratio of the expected vehicle speed to the actual average vehicle speed and the length of each section of the road as the weight value to calculate the whole road network on a monthly basis. Thus, the overall congestion index of the road network is obtained.
- (2) *Road Congestion Evaluation Standard Based on Road Traffic Density*. The Texas Road congestion Index Roadway Congestion Index (RCI) describes the intensity and persistence of congestion. The index is defined as the average daily traffic volume of different grades of roads multiplied by the average of the corresponding weight of the road. The RCI needs to count the traffic flow of each section and then calculate the number of miles per day by referring to the road length and lane. However, because the traffic flow does not always correspond to the actual traffic situation, the result is not necessarily accurate. For example, in the case of the same flow rate, the traffic state can be very smooth, or it can be very congested.
- (3) *Road Congestion Evaluation Standard Based on Road Traffic Volume*. In Japan, congestion is measured by Degrees of Congestion (DC). It is defined as the ratio of the real-time or historical traffic volume of a certain section to the evaluation reference volume corresponding to 24 hours or 12 hours a day. The DC index is based on the ratio of the actual traffic volume to the standard traffic volume in 12h, but its statistical time is longer and the range is larger, so it can only reflect the overall traffic congestion in the statistical time macroscopically.
- (4) *Road Congestion Evaluation Standard Based on Mileage Ratio of Congestion*. The evaluation index system of urban road traffic operation in Beijing takes the number of vehicle kilometers as the weight to calculate the proportion of traffic mileage of each grade road and get the proportion of traffic network congestion mileage of the whole city. Then the road traffic operation index is obtained according to a certain transformation model, and the whole road traffic running condition is reflected by the index. The index is based on the weight of the number of kilometers of the vehicle to calculate the proportion of the length of the congested road, but only the most congested section is considered in the calculation of the proportion of the length of the congested road.

The local and global effects of other less congested sections are not considered.

- (5) *Road Congestion Evaluation Standard Based on Travel Time Ratio*. The urban traffic operation index of Shenzhen is calculated based on the travel time ratio. Travel time ratio refers to the ratio of actual travel time to expected travel time, that is, the more time spent when the road is smooth. On this basis, the conversion relationship between the travel time ratio and the traffic operation index is obtained through the expert scoring. The index is intuitive and easy to understand.

Generally speaking, there are two problems in the existing evaluation criteria of road congestion:

- (1) The data collection cycle is too long, and the statistics are usually conducted daily or even monthly, without the ability to obtain real-time congestion levels;
- (2) The calculation of those existing evaluation criteria is either too macroscopic and thus cannot represent the characteristics of each road in the road network under the micro state; or in a microscope, it can reflect the congestion of a particular road, but the calculation is more complex, and the high complexity is often not conducive to the implementation of congestion evaluation.

In order to deal with a large number of real-time data, with small computational resource consumption and time consumption to get the microscopic road network congestion degree of each road, we need a more simple, direct, and efficient method to calculate and evaluate traffic congestion levels based on microscopic traffic flow characteristics.

Lozano et al. propose a method based on monitoring images, which uses the surveillance images at intersections to judge the service level of the shot sections and then get the congestion level of the roads [22]. Congestion evaluation algorithm based on surveillance image has general applicability to traditional traffic management, which is because surveillance camera is widely used in urban road construction as a necessary infrastructure for road construction in recent years. Therefore, the algorithm for infrastructure configuration requirements is low. However, because of the image processing, the requirements of image transmission are higher than the ordinary data transmission, and the image data processing which is far larger than the ordinary data processing will consume more computing resources. Therefore, the difficulty of real-time congestion calculation for all roads in urban road network is much more difficult than that of common data.

Zhang et al. propose a method based on macroscopic traffic flow characteristic-vehicle density [23]. This method uses the real-time road vehicle density obtained from the high-sensitivity magnetic force sensor distributed throughout the road network and calculates the congestion coefficient by using the mathematical model derived by themselves. Then the congestion coefficient is used to judge the road congestion grade. The congestion evaluation algorithm based on vehicle

density requires a sensor to collect the vehicle density; that is, it needs the support of additional infrastructure, so it still needs time to use in practice. But because the vehicle density itself is the macroscopic characteristic of the traffic flow, it can accurately evaluate the overall congestion grade of the road. This paper also takes the algorithm as a reference to improve it. A detailed description will be given in Section 4.

Liu et al. use logistic regression analysis to realize congestion grading [24]. Based on the dependent variable of congestion intensity, five independent variables are selected to represent the traffic state and build the model. After the significant verification of independent variables, it is determined that the travel speed should be used as independent variable, and the maximum likelihood method is used to solve the congestion index of different grades of roads and to grade it. The emphasis of this method is to find out the most influential factors of road congestion assessment, and it has certain accuracy.

3. Background

3.1. Microscopic Traffic Flow Characteristics. Road traffic flows consist of vehicles and their corresponding drivers, each of which has its own characteristics. A traffic flow is microscopic when it is considered to consist of a flow of a particular vehicle. Therefore, the dynamic of traffic flow is caused by the potential interaction between driver and vehicle; that is, the dynamic is largely determined by the physical characteristics of each driver and vehicle. The complexity of the algorithm can be greatly increased by considering the influence of drivers. However, the joint effects of drivers and vehicles on traffic flow should not be ignored. Nonetheless, it is still the usual practice to consider only the characteristics of traffic flow related to vehicles.

Taking vehicle i as an example, the microscopic characteristics of traffic flow include

- (a) Length of vehicle i : l_i
- (b) Longitudinal position of vehicle i on the road: x_i
- (c) Speed of vehicle i : v_i
- (d) Acceleration of vehicle i : a_i

In addition, because traffic flow is composed of vehicles, the microscopic characteristics of traffic flow also need to consider the relationship between vehicles.

3.2. Macroscopic Traffic Flow Characteristics. When a large number of vehicles are considered at the same time, four important measurements of macroscopic traffic flow characteristics are discussed:

- (a) *Vehicle Density*. It represents the degree of congestion in a particular section of the road, that is, the number of vehicles per unit length of a lane occupied under a given condition. The calculation of vehicle density usually ignores the influence of traffic composition and vehicle length and only considers the abstract "number of vehicles."

- (b) *Vehicle Flow*. Unlike the spatial measurement of vehicle density, vehicle flow is a time measurement, which is usually recorded as the frequency of traffic flow, that is, the number of vehicles passing through a particular point per unit time.
- (c) *Road Occupancy Rate*. Although the calculation of vehicle density is very important, it can be obtained by double induction coil detector (DLDD) sampling, but in most cases, the vehicle speed cannot be obtained in time to calculate the vehicle density, for example, when using a single inductive coil detector. In this case, it is possible to store the time-point measurements-road occupancy rate, which represents the ratio of the time when the vehicle passes through the detector to the total sampling time.
- (d) *Vehicle Average Speed*. It represents the average speed of a traffic flow over a period of time. Velocity is a vector that contains direction and size, while average velocity is the norm of speed; that is, it only represents size.

3.3. *SUMO*. SUMO [21] is a micro, multi-modal space continuous, time discrete traffic flow simulation platform. The implementation of SUMO began in 2001 and was first released in 2002 and developed by the German Institute of Transportation Systems and the German Aerospace Center. The SUMO project itself is an open source traffic simulation package, including the simulation application itself and the supporting tools for road network import and requirement modeling. SUMO can assist in observing a large number of research topics, mainly in traffic management and vehicular communications.

On the one hand, SUMO can simulate the traffic state change of macroscopic road network by constructing road network and loading vehicle flow data and express it visually. On the other hand, it can be accurate to every vehicle, every road, and every intersection at a specific point in time traffic state information.

3.4. *Traffic Flow Forecasting Algorithm Based on Transition Probability*. In [20] we have presented a traffic flow forecasting algorithm based on transition probability. This method first analyzes the transfer probability of upstream of the target road and then makes the prediction of the traffic flow at the next time by using the traffic flow equation. Newton Interior-Point Method is used to obtain the optimal value of parameters. Finally, it uses the proposed model to predict the traffic flow at the next time. We have not included details of that algorithm in this paper; readers who are interested can find out more in [20].

The output of this algorithm will be the congestion evaluation algorithm presented in this paper, thus helping us to predict congestion in the future.

4. Framework of the Congestion Evaluation Algorithm

In view of the existing standards and algorithms mentioned above, the evaluation algorithm described in this paper aims

to be able to calculate and evaluate traffic congestion levels based on microscopic traffic flow characteristics. At the same time, the computational complexity is low to adapt to the real-time data processing, so that the shortcomings of the above methods can be effectively overcome.

First, we give definitions for some parameters used in this paper:

- (1) ρ_i^x : the vehicle density of road x at time i ;
- (2) o_i^x : the road occupancy of road x at time i ;
- (3) $\overline{\rho}_I^x$: average vehicle density of road x within time period I ;
- (4) \overline{o}_I^x : average road occupancy of road x within time period I ;
- (5) $\overline{\rho}_I^x(m)$: average vehicle density at sampling time point m of road x within time period I ;
- (6) $\overline{o}_I^x(m)$: average road occupancy at sampling time point m of road x within time period I .

4.1. *Calculating Congestion Factor Based on Vehicle Density*. The road can be regarded as share resource for vehicle and traffic flow link, and according to Jains fairness index for shared computer systems, the quantitative congestion factor can be defined based on the traffic congestion model [23], as

$$C_{cf}^i(t) = \frac{(\sum_{m=1}^n \rho(x_m))^2}{n \sum_{m=1}^n (\rho(x_m))^2} \quad (1)$$

Here i indicates the lane number, x is the locations coordinate with origin starting from stop line, and the vehicle density is sampled in n discrete values with fixed frequency. The congestion factor indicates the general congestion state on the whole road segment, which is a number between 0 and 1, and larger value means more crowded.

If we treat the vehicle density of all the sampling points x_m as the average vehicle density of road x , then (1) could be transformed into

$$C_{cf}^x(t) = \frac{(\sum_{m=1}^n \overline{\rho}_I^x(m))^2}{n \sum_{m=1}^n (\overline{\rho}_I^x(m))^2} \quad (2)$$

It can be proved that the value of congestion factor calculated by (2) is between 0 and 1 and indicates the change of vehicle density in a road. The greater the absolute difference between the vehicle densities at each sampling point in a time period t , the closer the value of congestion factor to 0; conversely, the smaller the absolute difference, the closer the value of congestion factor to 1. Because of the actual situation, the smoother the road, the greater the change of the traffic flow; on the contrary, the change of the traffic flow will be very small when the road is seriously congested. Thus, we can conclude that the value of congestion factor is proportional to the traffic congestion to some extent.

Referring to [23], we present the road congestion classification based on the congestion factor as shown in Table 1.

TABLE 1: Decide congestion factor based on vehicle density.

Congestion Grade	Congestion Factor C
No Congestion	$0 \leq C < 0.3$
Slight Congestion	$0.3 \leq C < 0.8$
Congestion	$0.8 \leq C < 1$

TABLE 2: Preliminary result of congestion grade based on road occupancy.

Preliminary Congestion Grade	Congestion Factor C
No Congestion	$0 \leq C < 0.3$
Slight Congestion	$0.3 \leq C < 0.8$
Might be Congestion	$0.8 \leq C < 1$

4.2. *Calculating Congestion Factor Based on Road Occupancy.* However, calculating congestion factor based on vehicle density has two issues:

- (i) The calculation of vehicle density is based on the ideal condition, so the influence of the length of vehicle and the composition of vehicle in the vehicle flow on the result is not taken into account, so there will be deviation between the calculation result and the actual situation.
- (ii) When the congestion factor is considered based on the vehicle density, the existence of special circumstances is not considered; that is, when the road is extremely smooth, there may also be a small difference in the density of vehicles on the road. In other words, the vehicle density on the road will be kept at a relatively small range. In this case, there will be a deviation in determining the congestion factor based on the difference of the vehicle density.

So, we present a method to calculate congestion factor based on road occupancy.

In view of the first problem, road occupancy can be used to characterize road congestion at a particular point in time, as is the vehicle density, and the calculation of road occupancy takes into account the factor of vehicle length, so we use $\bar{o}_t^x(m)$ to replace $\bar{\rho}_t^x(m)$ in (2) to improve accuracy:

$$C_{cf}^x(t) = \frac{\left(\sum_{m=1}^n \bar{o}_t^x(m)\right)^2}{n \sum_{m=1}^n \left(\bar{o}_t^x(m)\right)^2} \quad (3)$$

Consider the second issues, after we use (3) to obtain the congestion factor C in time period t road x . According to Table 2, the preliminary evaluation results are obtained, and then combining with the average road occupancy of road x in time period t , the congestion grade of road x in time period t is determined according to Table 3.

4.3. *Modification for Situation Where Inputs Are Traffic Flow.* In practical application, the occupancy of multiple sampling points on the road can be directly obtained by using a single inductive coil detector, and the average occupancy of the

road can be obtained. For other sampling methods and the traffic flow prediction algorithm as input in this paper, only traffic flow information can be obtained, so for the case where the input is traffic flow, according to the traffic flow theory [25], the road occupancy rate can be derived from the vehicle density and the average vehicle length:

$$\bar{o}_t^x = \bar{l} \cdot \bar{\rho}_t^x \quad (4)$$

Here \bar{l} is the average length of a total of N vehicles in road x . If we assume \bar{l} is 0.005km, then (4) is revised to

$$\bar{o}_t^x = 0.005 \cdot \bar{\rho}_t^x \quad (5)$$

Based on the relationship among vehicle density, vehicle speed, and vehicle flow, the following formulas can be deduced:

$$\bar{\rho}_t^x = \frac{\bar{f}_t^x}{v_0} \quad (6)$$

Here, \bar{f}_t^x is the average traffic flow in road x in time period t , and v_0 is the average speed of the road x within the time period t , where the maximum speed limit of the vehicle specified on the road x is taken directly. Based on (3), (5), and (6), we gain

$$C_{cf}^x(t) = \frac{\left(\sum_{m=1}^n 0.005 \left(\bar{f}_t^x(m)/v_0\right)\right)^2}{n \sum_{m=1}^n \left(0.005 \left(\bar{f}_t^x(m)/v_0\right)\right)^2} \quad (7)$$

$$\bar{o}_t^x = 0.005 \frac{\bar{f}_t^x}{v_0} \quad (8)$$

Therefore, (7) is the formula for calculating the congestion factor when the traffic flow is taken as input, and (8) is the formula for calculating the average road occupancy for the final judgment.

5. Experiment Results

5.1. *Experiment Framework.* In this paper, two comparative evaluation frameworks are designed and implemented.

The first framework focuses on the crosswise comparison between the congestion evaluation algorithms described in this paper and several different algorithms under the same road conditions. It has uniform input and output. The frame inputs one or more real-time traffic streams of the same road, the same sampling frequency, and one or more sampling frequencies corresponding to the real-time traffic flow. BP neural network is used to predict traffic flow. The unified output is to obtain the congestion grade judgment obtained from the input data through different congestion evaluation algorithms. The unified evaluation criteria are also reclassified according to their original criteria to facilitate horizontal comparison.

The structure of this framework is shown in Figure 1. The framework constructs a unified input and output mode around different congestion evaluating algorithms. Input

TABLE 3: Result of congestion grade based on road occupancy.

Congestion Grade	Average Road Occupancy \bar{o}_t^x	Preliminary Congestion Grade
No Congestion	-	No Congestion
Slight Congestion	-	Slight Congestion
No Congestion	$0 \leq \bar{o}_t^x < 0.5$	Might be Congestion
Slight Congestion	$0.5 \leq \bar{o}_t^x < 0.8$	
Congestion	$0.8 \leq \bar{o}_t^x < 1$	

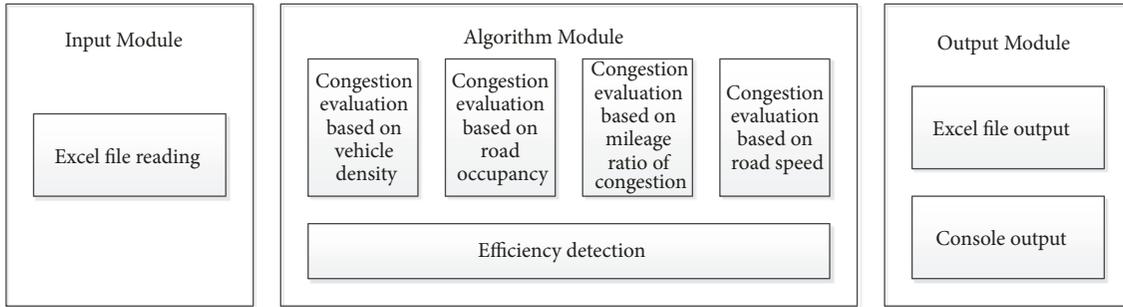


FIGURE 1: The framework for crosswise comparison between different congestion evaluation algorithms.

format supports XML file reading and text file reading; output supports Excel file output, console output, and so on.

The second framework focuses on evaluating the performance of the congestion evaluation algorithm proposed in this paper based on real-time data. First, SUMO is used to generate the traffic flow simulation of a specific urban road network, and then the road information of each sampling point on each road is sampled by SUMO, to get the road occupancy of each sampling time point in the road network at a fixed sampling frequency, and further use the congestion evaluation algorithm described in this paper to classify each road congestion grade. Finally, the traffic congestion level of all roads in the time period is plotted on the road network.

The structure of this framework is shown in Figure 2. The framework is developed based on SUMO Modeler, an open source Java branch of SUMO, which is used to test the real-time rendering effect of congestion evaluation algorithm based on road occupancy proposed in this paper for the overall congestion of urban road network. The framework designed in this paper not only supports the method of traffic modeling in graphic form by reading the road network in XML format generated by SUMO and traffic data, as in SUMO Modeler, but also increases the reading and parsing of SUMO road sampling files, road-network congestion grade calculation based on sampling point, and road-network congestion grade rendering.

The modification of our work is mostly in the File Parsing Module, Congestion Drawing Module, and Graph Drawing Module. The interface of our framework is shown in Figure 3. The improvements made according to the evaluation requirements of the algorithm described in this paper mainly include (1) adding the function of sampling file reading, (2) increasing congestion drawing layer, and (3) increasing congestion rendering playback.

There are four partitions in the interface:

- (1) Menu bars and toolbars, including selection tools, congestion playback tools, and modeling tools
- (2) Road network, layer, and congestion rendering display area
- (3) Layer selection area
- (4) Road retrieval and road attribute display area

5.2. Result for Crosswise Comparison of Different Congestion Evaluation Algorithm

5.2.1. Comparison Method. In order to compare and evaluate the congestion evaluation algorithm proposed in this paper with other algorithms, the method of combining manual verification with machine verification is adopted:

- (i) Manual verification is to sample the traffic data of concerned road time period, determine the current congestion level artificially, and compare with the results of the four algorithms to verify the accuracy of the algorithm.
- (ii) Machine verification is to use the Efficiency Detection Module (Figure 1) to compare the efficiency of each algorithm in the same input, calculate the time consumed, and then evaluate the efficiency of the algorithm.

The two kinds of verification results will generate Excel files through output module, import Matlab and use its drawing function, and display them in the form of broken line diagram and bar chart, so as to facilitate the comparison. Three different algorithms are chosen to compare to our methods (denoted as CRO, congestion evaluation based on road occupancy), which are congestion evaluation based on mileage ratio of congestion (CMRC), congestion evaluation

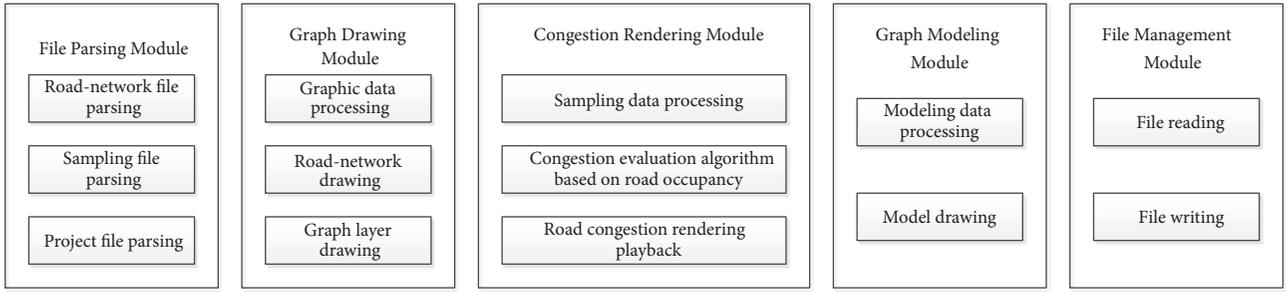


FIGURE 2: The framework for evaluating the performance of the proposed algorithm.

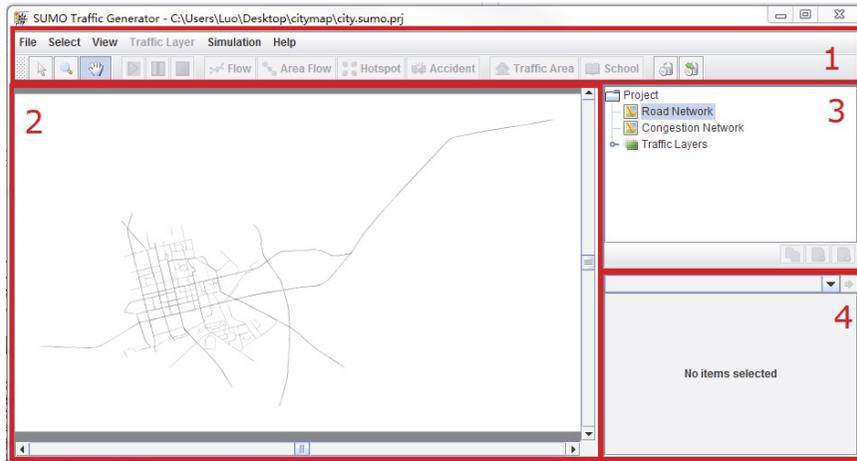


FIGURE 3: The interface of our framework for performance evaluation.

TABLE 4: Interpreting of evaluation result to road congestion grade.

Congestion Evaluation Result	Congestion Grade
0	No Congestion
1	Slight Congestion
2	Congestion

based on road speed (CRS), and congestion evaluation based on vehicle density (CVD).

As mentioned earlier, the congestion evaluation algorithm based on road occupancy rate divides the road congestion rating into three levels, corresponding to Table 4.

The test data used in the evaluation of road congestion grade are different and correspond to the two input forms in the input module, respectively, and are shown in Table 5. Here, the road data for Cologne, Germany, came from the real traffic data from 6:00 to 8:00 in the German city of Cologne, where two roads were taken out. Karamay’s two roads use the same length of time (that is, two hours) of SUMO traffic simulation data. The sampling rate of 4 roads is 1 s, so the total data scale is 7200.

We divided the traffic flow data of each road into two groups: one is the training group, that is, the training data as the neural network input, a total of 4,200, and the remaining group as a prediction group, a total of 3000, as the input of the trained neural network. In prediction, a 1-minute sampling

and prediction frequency is used to predict the next moment, that is, a minute after the current moment.

5.2.2. Result Evaluation. In the experiment, for CVD and CRO, since they have three different congestion grade levels, we will use the parameters shown in Table 4, while for CMRC and CRS, since they only can output two levels of congestion grade, we will use T (true) to denote congestion, and F (false) to denote no congestion. As in CVD and CRO, we will need to first generate the congestion factor and then use it to decide the final congestion grade; we put them in the result tables, too.

The experiment results for road -37932733#6 are shown in Table 6. The experiment results for road 188982779#3 are shown in Table 7. The experiment results for road 24592809 are shown in Table 8. The experiment results for road 23572355#2 are shown in Table 9.

Based on the comparison results, we draw the comparison chart of algorithm accuracy as shown in Figure 4. Here, the classification of CRO and CVD is strictly differentiated; that is, when the output of those two algorithms is the intermediate state, slight congestion, then it is treated as an inaccurate result directly.

From these comparison results, we can see that the accuracies of the CRS algorithm and the CMRC algorithm are close. The congestion factors calculated by the two algorithms based on the macroscopic traffic flow characteristics are very

TABLE 5: Description of experimental data.

City	Road ID	Algorithm	Source	Scale(No. of data)	Description
Karamay City, China	-37932733#6	CMIRC	SUMO Sampling File	7200	Simulated sampling data of average road speed
		CRS	Text File		Simulated real-time and predicted traffic flow
	188982779#3	CMIRC	SUMO Sampling File	7200	Simulated sampling data of average road speed
		CRS	Text File		Simulated real-time and predicted traffic flow
Cologne, Germany	24592809	CMIRC	SUMO Sampling File	7200	Real sampling data of average road speed
		CRS	Text File		Real real-time and predicted traffic flow
	23572355#2	CMIRC	SUMO Sampling File	7200	Real sampling data of average road speed
		CRS	Text File		Real real-time and predicted traffic flow

TABLE 6: Comparison result for road -37932733#6 in Karamay City, China.

Index	Congestion Factor of CRO	Congestion Factor of CVD	CRO	CVD	CRS	CMRC	Human Verification
0	1.0	1.0	0	2	T	T	F
1	0.7876923076923074	0.7876923076923077	1	1	T	T	F
2	0.8243902439024391	0.8243902439024391	0	2	T	T	F
3	0.9043010752688169	0.9043010752688172	0	2	T	T	F
4	0.9523809523809519	0.9523809523809523	0	2	T	T	F
5	0.0	0.0	0	0	F	F	F

TABLE 7: Comparison result for road 188982779#3 in Karamay City, China.

Index	Congestion Factor of CRO	Congestion Factor of CVD	CRO	CVD	CRS	CMRC	Human Verification
0	0.0	0.0	0	0	F	F	F
1	0.3	0.30000000000000004	0	1	T	T	F
2	0.6000000000000001	0.6000000000000001	1	1	T	T	F
3	0.3	0.30000000000000004	0	1	T	T	F
4	0.7	0.7	1	1	T	T	F
5	0.0	0.0	0	0	F	F	F

TABLE 8: Comparison result for road 24592809 in Cologne, Germany.

Index	Congestion Factor of CRO	Congestion Factor of CVD	CRO	CVD	CRS	CMRC	Human Verification
0	1.0	1.0	0	2	T	T	F
1	0.8047619047619045	0.8047619047619047	0	2	T	T	F
2	0.8981308411214957	0.8981308411214953	0	2	T	T	F
3	0.941176470588235	0.9411764705882354	0	2	T	T	F
4	0.9580498866213155	0.9580498866213153	0	2	T	T	F
5	0.0	0.0	0	0	F	F	F

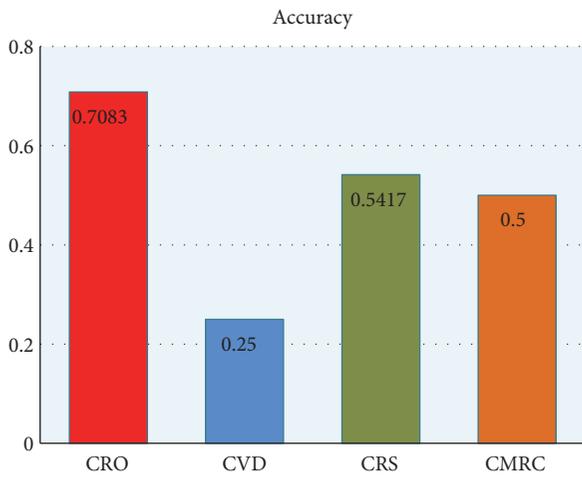


FIGURE 4: Comparison of accuracy for different congestion evaluation algorithms.

close to each other, too. However, the accuracy of the CRO algorithm proposed in this paper has been improved obviously. The main reason is that the CRO algorithm proposed in this paper combines the average road occupancy rate with the congestion factor and thus could consider special cases (described in detail earlier), so the accuracy is relatively high.

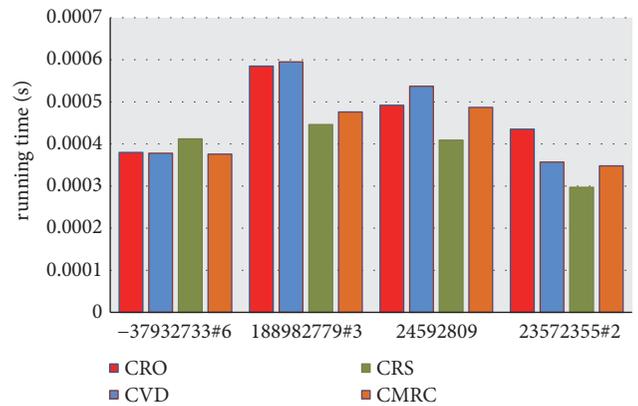


FIGURE 5: Comparison of running time for different congestion evaluation algorithms.

From the efficiency detection module in the framework, the algorithm running time comparison diagram is shown in Figure 5. It can be seen from the diagram that two algorithms based on macroscopic traffic flow characteristics, namely, CVD algorithm and CRO algorithm, consume about the same time. The CRS algorithm and the CMRC algorithm consume the same time and are lower than the first two algorithms. But in fact, the overall time consumption of all these four algorithms is lower than that of data receiving,

TABLE 9: Comparison result for road 23572355#2 in Cologne, Germany.

Index	Congestion Factor of CRO	Congestion Factor of CVD	CRO	CVD	CRS	CMRC	Human Verification
0	0.0	0.0	0	0	F	F	F
1	0.6230769230769231	0.623076923076923	1	1	F	F	F
2	0.7	0.7	1	1	F	F	F
3	0.39999999999999997	0.4	1	1	F	F	F
4	0.5451612903225806	0.5451612903225806	1	1	F	F	F
5	0.0	0.0	0	0	F	F	F

TABLE 10: Data for performance evaluation for CRO algorithm.

Road-network	Scale of road-network (No. of roads)	Scale of Sampling data (No. of records)	File size (KB)
Karamay	740	740000	112992
Cologne	457	457000	80618

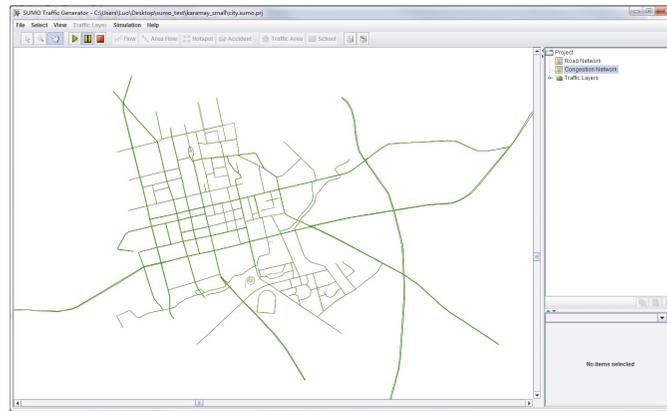


FIGURE 6: Whole road network of Karamay.

processing, road rendering, and display. The time consumed by these processes is described in detail in Figure 5.

5.3. Result for Performance Evaluation of the Proposed Congestion Evaluation Algorithm

5.3.1. Comparison Method. The performance of the proposed CRO algorithm is evaluated by manual verification, the main purpose of which is to evaluate the impact of the data size on the real-time rendering of road congestion. After loading the road network and sampling data, the congestion of the road network is played to observe whether the interval between the timing thread and the time used to complete the actual drawing will be out of sync with the increase in the amount of data. The feasibility of applying the CRO algorithm to the whole urban road network is judged.

In the evaluation of the performance of the CRO algorithm, the road networks used are two different sizes of Karamay and Cologne urban road networks. The complexity of road connection and the number of roads are different between the two road networks. The SUMO sampling data used are based on simulated and real traffic data from these networks, respectively. Each road corresponds to 1000 time units of sampling data, so the scale of sampling data is

proportional to the number of roads. The evaluation data is shown in Table 10.

5.3.2. Result for Karamay. The road network of this evaluation case is the road network of some streets in the center of Karamay, Xinjiang, including 740 roads. The road traffic data are sampled in seconds, a total of 1000 time sampling points. By loading SUMO road network data and SUMO sampling data separately, and then using the algorithm to judge the road congestion grade of all sampling points, the congestion level is played at intervals of 1 s. The road network is shown in Figure 6, and it could be enlarged as shown in Figure 7.

We have built two kinds of sampling data:

- (1) Normal situation: that is, one moving vehicle is generated in the road network every 1 second;
- (2) Extreme situation: every 0.1 seconds, a moving vehicle is generated in the road network to simulate traffic congestion.

In the course of evaluation, the road numbered 238549234#0 of SUMO network is selected to record the change of congestion grade. The change of congestion grade under normal situation is shown in Figure 8, the traffic flow in the road continues to change, and occasionally there

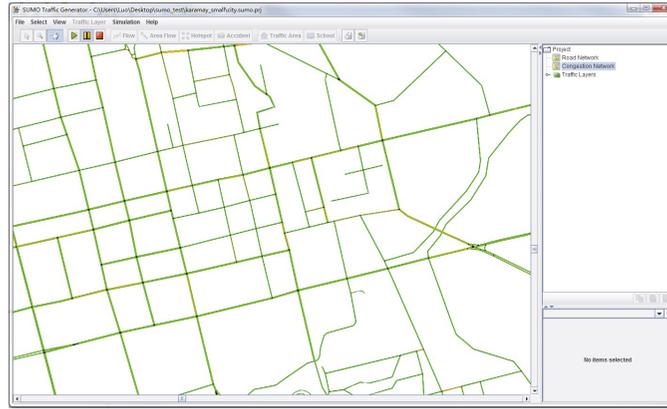


FIGURE 7: Detail of road network of Karamay.

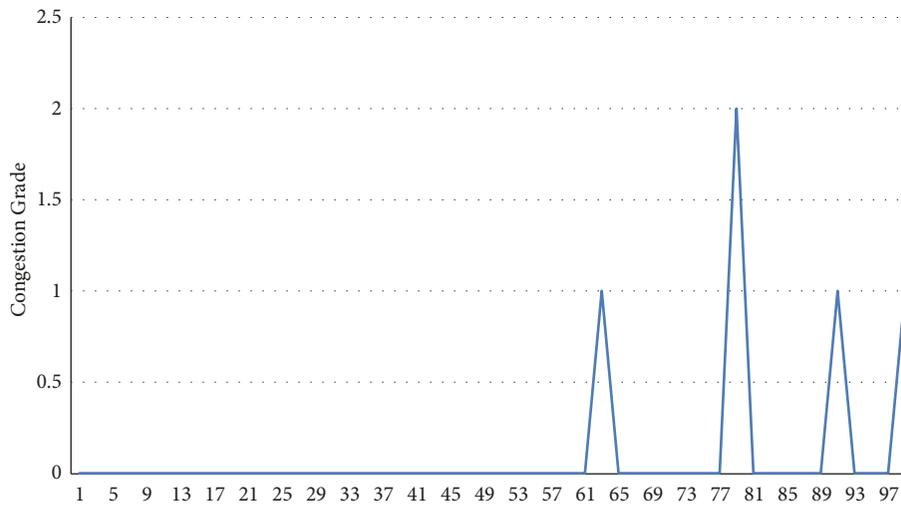


FIGURE 8: Change of congestion grade under normal situation.

is a heavy traffic situation. The change of congestion grade under extreme situation is shown in Figure 9. As the number of generated vehicles increases gradually until the vehicle cannot enter and exit normally the occupancy rate continues to increase and the road falls into a state of congestion which shows that the congestion level fluctuates from 0 at the beginning to 2 at the end.

Java Visual VM is used to evaluate the virtual machine and the program in different runtimes, as shown in Table 11.

5.3.3. *Result for Cologne.* The road network of this evaluation case is the road network of some streets in the center of Cologne, Germany, including 457 roads. The difference between this road network and the Karamay road network is that this road network has a high degree of interleaving and a high degree of road connectivity, and the complexity of simulation and rendering is higher than that of the former.

Congestion grade rendering result is shown in Figure 10. Java Visual VM is used to evaluate the virtual machine and the program in different runtimes, as shown in Table 12.

6. Conclusion

In this paper, an algorithm for evaluating congestion level is proposed. The main reference value of congestion evaluation is the occupancy rate of roads at a specific time. We use two comparison frameworks to verify the algorithm and compare it with many other algorithms. The road network data used in the comparison are, respectively, from Cologne, Germany, and Karamay, Xinjiang, China. Traffic data are derived from real data and simulated data generated by SUMO.

Based on the experimental results, the results can be summarized as follows: the congestion evaluation method proposed in this paper can accurately express the congestion degree of road network based on real-time and predicted traffic data on the basis of high efficiency and fast original intention. This conclusion is supported by the following facts:

- (1) For a single road, the processing time of 3000 seconds' sampling data is less than 0.0005 seconds on average, and the processing time of real-time processing and



FIGURE 9: Change of congestion grade under extreme situation.

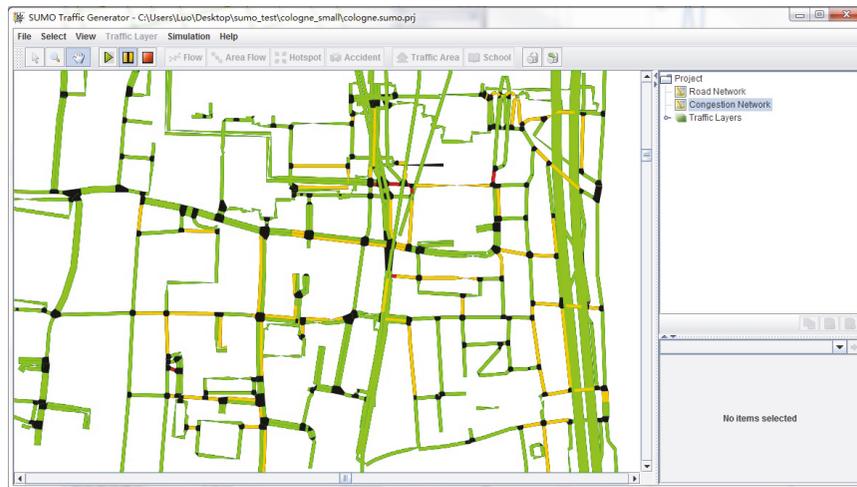


FIGURE 10: Congestion grade rendering for Cologne.

real-time rendering of more than 70000 sampled data is less than 6s;

- (2) Without training cost, the algorithm itself is based on real-time or predictive traffic data in a short period of time, without the need for a large number of data as training sets and a large amount of time for training;
- (3) Accuracy is higher than other algorithms compared;
- (4) It can support congestion calculation and real-time rendering for different cities, different road networks, different road structures, different traffic modes, and

different sampling data types. It is robust and can support larger road network and sample data processing and rendering when the hardware is extended.

Data Availability

Requests for data, 6/12 months after publication of this article, will be considered by the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

TABLE 11: Monitoring table of occupying Resources of Karamay Network Operation Program.

Running phase	CPU occupancy rate	Duration	Memory occupancy
initialization	0.1%	Negligible	53MB
parsing/loading road network file	average 1.7%, peak 20%	1s	105MB
parsing/loading sampling file	average 10%, peak 31%	5s	1400MB
congestion grade rendering	average < 5%, peak 12%	equal to sampling time	200MB

TABLE 12: Monitoring table of occupying Resources of Cologne Network Operation Program.

Running phase	CPU occupancy rate	Duration	Memory occupancy
initialization	0.1%	Negligible	19MB
parsing/loading road network file	average 1.7%, peak 7.2%	1s	134MB
parsing/loading sampling file	average 18%, peak 34%	6s	800MB
congestion grade rendering	average < 5%, peak 15%	equal to sampling time	83MB

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant No. 91846303 and the National Natural Science Foundation of China under Grant No. 61502032.

References

- [1] IBM, "Frustration rising: Ibm 2011 commuter pain survey," <http://www.ibm.com/us/en/>.
- [2] D. Metz, "The myth of travel time saving," *Transport Reviews*, vol. 28, no. 3, pp. 321–336, 2008.
- [3] G. Orosz, R. E. Wilson, and G. Stpn, "Traffic jams: dynamics and control," *Philosophical Transactions Mathematical Physical & Engineering Sciences*, vol. 368, no. 1928, pp. 4455–4479, 2010.
- [4] E. Naone, "Gps data on beijing cabs reveals the cause of traffic jams," *MIT Technology Review*, <https://www.technologyreview.com/s/425553/gps-data-on-beijing-cabs-reveals-the-cause-of-traffic-jams/>, Accessed 20190414.
- [5] G. Dimitrakopoulos and P. Demestichas, "Intelligent transportation systems," *IEEE Vehicular Technology Magazine*, vol. 5, no. 1, pp. 77–84, 2010.
- [6] X. Wang, Z. Ning, and L. Wang, "Offloading in internet of vehicles: a fog-enabled real-time traffic management system," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 10, pp. 4568–4578, 2018.
- [7] X. Kong, F. Xia, Z. Ning et al., "Mobility dataset generation for vehicular social networks based on floating car data," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 3874–3886, 2018.
- [8] A. Faro, D. Giordano, and C. Spampinato, "Integrating location tracking, traffic monitoring and semantics in a layered ITS architecture," *IET Intelligent Transport Systems*, vol. 5, no. 3, pp. 197–206, 2011.
- [9] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: a survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1624–1639, 2011.
- [10] R. Deng and H. Liang, "Whether to charge an electric vehicle or not? A near-optimal online approach," in *Proceedings of the IEEE Power and Energy Society General Meeting, PESGM '16*, pp. 1–5, 2016.
- [11] Q. Gong, Y. Li, and Z.-R. Peng, "Trip-based optimal power management of plug-in hybrid electric vehicles," in *Proceedings of the American Control Conference, ACC '08*, pp. 3225–3230, 2008.
- [12] H. Dia, "An object-oriented neural network approach to short-term traffic forecasting," *European Journal of Operational Research*, vol. 131, no. 2, pp. 253–261, 2001.
- [13] G.-E. Katsargyri, I. V. Kolmanovskiy, J. Michelini et al., "Optimally controlling hybrid electric vehicles using path forecasting," in *Proceedings of the 2009 American Control Conference, ACC '09*, pp. 4613–4617, 2009.
- [14] J. Park, D. Li, Y. L. Murphey et al., "Real time vehicle speed prediction using a neural network traffic model," in *Proceedings of the International Joint Conference on Neural Network, IJCNN '11*, pp. 2991–2996, 2011.
- [15] Y. Bin, Y. Li, Q. Gong, and Z.-R. Peng, "Multi-information integrated trip specific optimal power management for plug-in hybrid electric vehicles," in *Proceedings of the American Control Conference, ACC '09*, pp. 4607–4612, 2009.
- [16] S.-I. Jeon, S.-T. Jo, Y.-I. Park, and J.-M. Lee, "Multimode driving control of a parallel hybrid electric vehicle using driving pattern recognition," *Journal of Dynamic Systems Measurement and Control*, vol. 124, no. 1, pp. 489–494, 2002.
- [17] M. Schulze and J. Z. Riveros, "Impact of electrical vehicles on strategic planning of energy infrastructure," in *Proceedings of the International Conference on Power System Technology*, pp. 1–6, 2010.
- [18] M. J. Lighthill and G. B. Whitham, "On kinematic waves. ii. a theory of traffic flow on long crowded roads," *Proceedings of the Royal Society A Mathematical Physical & Engineering Sciences*, vol. 229, no. 1178, pp. 317–345, 1955.
- [19] R. M. Flynn, R. A. Kasimov, J. Nave, and R. R. Seibold, "Self-sustained nonlinear waves in traffic flow," *Physical Review E Statistical Nonlinear & Soft Matter Physics*, vol. 79, no. 2, Article ID 056113, 2008.
- [20] H. Tang, Y. Liang, Z. Huang et al., "Key technology of real-time road navigation method based on intelligent data research," *Computational Intelligence & Neuroscience*, vol. 2016, Article ID 1874945, p. 4, 2016.
- [21] M. Behrisch, D. Krajzewicz, and M. Weber, *Simulation of Urban Mobility*, Springer Berlin Heidelberg, 2014.
- [22] A. Lozano, G. Manfredi, and L. Nieddu, "An algorithm for the recognition of levels of congestion in road traffic problems,"

- Mathematics and Computers in Simulation*, vol. 79, no. 6, pp. 1926–1934, 2009.
- [23] W. Zhang, G. Tan, N. Ding, and G. Wang, “Traffic congestion evaluation and signal control optimization based on wireless sensor networks: model and algorithms,” *Mathematical Problems in Engineering*, vol. 2012, Article ID 573171, 18 pages, 2012.
- [24] M. Liu, L. Yu, X. Zhang, and S. Guo, “Cumulative logistic regression-based measurement models of road traffic congestion intensity,” *Journal of Beijing Jiaotong University*, vol. 32, no. 6, pp. 52–56, 2008.
- [25] S. Maerivoet and B. De Moor, “Traffic flow theory,” *Physics*, vol. 1, no. 1-2, pp. 5–7, 2005.

