IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 6, 2020, accepted December 19, 2020, date of publication December 25, 2020, date of current version January 6, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3047394

Congestion Pattern Prediction for a Busy Traffic Zone Based on the Hidden Markov Model

TINGTING SUN[1,2], ZHENGFENG HUANG[1,2], HONGDONG ZHU[1], YANHAO HUANG[1], AND PENGJUN ZHENG[1]

[1]Faculty of Maritime and Transportation, Ningbo University, Ningbo 315832, China
[2]Jiangsu Province Collaborative Innovation Center for Modern Urban Traffic Technologies, Nanjing 211189, China

Corresponding author: Zhengfeng Huang (huangzhengfeng@nbu.edu.cn)

This work was supported in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LY18E080009, in part by the Ningbo Natural Science Foundation of China under Grant 2019A610040, in part by the Project of Ningbo Transportation Technology under Grant HS2020000169, and in part by K. C. Wong Magna Fund in Ningbo University.

ABSTRACT Congestion pattern in a busy traffic zone would not always be a single state. The scenario of congestion occurring in multiple connected road links may be shifted to the scenario of whole zone network paralyzation immediately. It is necessary to dynamically predict the congestion pattern for a busy traffic zone rather than a road link, which could provide information for network system decision making. Considering the close connection between the upstream and downstream traffic flow, this study proposes a congestion pattern prediction model for a busy traffic zone based on the hidden Markov model (HMM). The model establishes a correlation between the external road traffic state (observation state) and internal road traffic state (hidden state) of a busy traffic zone. We acquire these traffic states by cleaning and mining floating vehicle trajectory data. With these data, we calibrate the HMM and predict the zone congestion pattern. This article demonstrates the validity and rationality of the model by taking a hospital area in Ningbo City as an example. The prediction accuracy can reach 83.4%, which is 5.8% higher than that of the autoregressive moving average model. This case result illustrates the feasibility and effectiveness of our approach in the field of congestion pattern prediction for busy traffic zone.

INDEX TERMS Regional congestion, hidden Markov model, congestion pattern.

I. INTRODUCTION

As motorization trend has been rising rapidly in many countries, traffic problems in urban areas of these countries become more serious. Currently, congestion of local network rather than single intersection is more frequently seen in an urban busy traffic zone. The so-called congestion pattern of busy traffic zones in this paper is a cascade congestion pattern of several links. The amount and components of these links may change each time. As seen in Figure 1, the area consisting of 14 directional links is a busy traffic zone. It is enveloped by the dotted line. Each subfigure shows a zone congestion pattern, and their patterns are different. The transition process of congestion pattern, from left to right, indicates that the zone congestion problem is becoming worse. In fact, the traffic congestion phenomenon of busy traffic zones is most frequently the consequence of huge traffic flows on the upstream associated links. Therefore, traffic congestion in a busy traffic zone in the short term can be predicted by using the traffic information of the upstream associated links. From the perspective of an occurring congestion process, when the traffic inflow upstream exceeds the road capacity downstream, a certain link will be correspondingly congested after a period of time. In this case, it may cause immediate congestion backtracking. If the congestion is not effectively alleviated, it will eventually turn into road network congestion in the busy traffic zone. Because the traffic flow is time-related between the upstream and downstream on the road network, the short-term future traffic flow in the busy traffic zone is affected by the current upstream traffic flow. With an increase in traffic flow upstream, the traffic pressure will be transmitted to the interior of a busy traffic zone after a certain time. Therefore, the internal traffic state of a busy traffic zone will change. If the speed of the upstream link of a busy traffic zone is used to predict the future congestion pattern of the area, it can be used to guide traffic evacuation in advance to

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

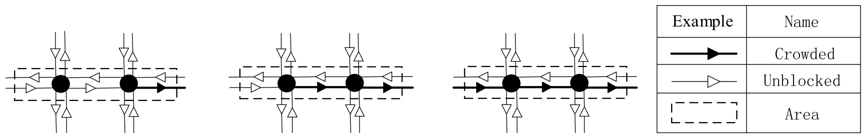


FIGURE 1. Evolution of congestion in an area.

avoid or reduce the occurrence of congestion in a busy traffic zone and to ensure the smooth operation of the urban road network.

At present, many researchers have explored the field of congestion prediction. Most studies build parametric or nonparametric models to predict the link congestion by considering this link's the speed and flow [1]–[5]. Parametric approaches can achieve good performance when traffic shows regular variations. Levin *et al.* [6] applied the Box-Jenkins time-series analyses to predict freeway traffic flow, and they found that the autoregressive integrated moving average (ARIMA) model was most statistically significant. Ahmed *et al.* [7] applied the ARIMA model for traffic flow forecasts in urban arterial roads. Nonparametric approaches can achieve good

performance when the traffic shows irregular variations. There are a lot of researches about nonparametric approaches. Karlaftis *et al.* [8] applied a neural network prediction model to predict traffic flow. Zhang *et al.* [9] applied a support vector machine model to forecast the travel time index (TTI). Okutani *et al.* [10] proposed the Kaman filtering model to predict short-term traffic volume. Innamaa [11] designed a self-adapting short-term prediction model to determine the state of traffic flow based on stationary traffic prediction models. With the advent of the era of big data, many traffic big data prediction models have been created. For example, Zhao *et al.* [12] proposed the long-term memory model (LSTM). Jeon *et al.* [13] proposed a traffic time series forecasting model based on vertical data arrangement. Han [14] proposed a short-term traffic flow prediction model based on multivariable phase space reconstruction and the least squares support vector machine (LSSVM). Cao *et al.* [15] proposed a short-term traffic flow prediction model based on the particle swarm optimization-support vector machine (PSO-SVM). Liu *et al.* [16] proposed a short-term traffic flow forecast model based on a combination of the K nearest neighbor algorithm and support vector regression. Sun *et al.* [17] proposed a short-term traffic state forecasting model with a support vector machine (SVM) considering the proportion of large vehicles. Xing *et al.* [18], through the analysis of the temporal and spatial relevance of vehicle speed, proposed a multilocation prediction model for speed using the radial basis function (RBF) neural network based on the temporal distribution and spatial distribution of vehicle speeds. Chen *et al.* [19] proposed deep convolutional networks for short-term traffic congestion prediction. Tang *et al.* [20] constructed fuzzy neural network to forecast travel speed for multi-step ahead based on 2-min travel speed data collected from three remote traffic microwave sensors located on a southbound segment of a fourth ring road in Beijing City. Zhang *et al.* [21] proposed an end-to-end multi-task learning temporal convolutional neural network (MTL-TCNN) to predict the short-term passenger demand in a multi-zone level. Some approaches conduct the prediction work by considering the impact of external factors (weather, time period, visibility, workday/weekend, etc.) on the road link. For example, Qian *et al.* [22] used a Bayesian network structure to predict traffic state; Wei *et al.* [23] used a nonparametric regression model-K nearest neighbor algorithm to predict the urban road congestion index; and Kumar *et al.* [24] used the ARIMA to predict the traffic volume of main roads in Indian cities. However, lots of aforementioned researches adopt extrapolated prediction methods, which cannot reveal the direct internal causal correlation between these factors and the predicted objects.

Considering the traffic correlation among links, only a few methods achieve a prediction by establishing a correlation between the upstream traffic demand and the current road traffic state. For instance, to construct this relation, Zhang *et al.* [25] and Daganzo [26] used the cell transmission model, and Wang *et al.* [27] used the predictive feedback control model. However, most of these methods are limited to highway scenes. Highway network is simple, and the vehicle direction is fixed, which ensures high prediction accuracy. For an intricate urban road network, it is difficult to determine various route selection options for travelers in advance, and the prediction of causality-related congestion is difficult.

Recently, a few studies have carried out causality-related congestion predictions for urban road networks. For example, Zhu *et al.* [28] utilized the surrounding road traffic state to predict the traffic state of specific road. Liu *et al.* [29], Pang *et al.* [30], and Chu *et al.* [31] proposed finding the causal relationship between space-time congestion and using the causal tree algorithm to predict traffic congestion. However, the traffic network in these studies was modeled by partitioning the urban area into regions or junctions, but real traffic flows are road-based and not region-based. In addition, the recursive implementation of the causal tree algorithm is usually not suitable for large traffic networks with limited time and memory.

In summary, relatively few studies have focused on the prediction of congestion patterns in busy traffic zones. The congestion pattern of a busy traffic zone is related to the traffic state of the upstream links. This paper makes full use of a large number of floating vehicle traffic data to record the historical traffic state of the upstream links and the traffic congestion pattern of a busy traffic zone. Specifically, a dynamic prediction model for the traffic flow state is established based on the hidden Markov model (HMM). The hidden relationships between these state variables of the road, combined with the input current traffic data of the upstream links, are used to predict the short-term congestion patterns for the area in the future.

II. VARIABLE DEFINITIONS

TABLE 1. Variable definitions.

Symbol	Definition
g	GPS sampling interval
t	t is the travel time from the upstream collection point to the congestion point of the busy traffic zone. It is also the smallest time unit used to calculate the average running speed and traffic state variables
T	State sequence length for HMM
V_i	The average speed of the upstream links during the i-th time period
L_{ri}	The traffic state of the r-th ($r \in R$) link in a busy traffic zone at the i-th time period, the value range is { Ⅰ, Ⅱ, and Ⅲ}. R is the set of all links in the area. The traffic state of R-related zone at the i-th time period is set as L_{Ri}
S	Hidden state sequences, in which the hidden state of the i-th time period is s_i
O	Observation state sequences, in which the observation state of the i-th time period is o_i
X	Hidden state set. Suppose there are N hidden states, the i-th $(i \leq N)$ state is denoted as x_i
Y	Observation state set. Suppose there are M observation states, the i-th $(i \leq M)$ states is denoted as y_i

III. THE DETERMINATION OF ZONE BOUNDARY AND THE UPSTREAM DATA COLLECTION POINTS

Regarding the acquisition of a busy traffic zone boundary or the delineation of a busy traffic zone, firstly, we locate the areas of hospitals, schools, and commercial centers that often suffer from congestion. When two or more links exist synchronous congestion each workday, it means there may be a busy traffic zone in this area. Then, observe the number of congestion times in each link of this traffic sensitive area based on historical peak hour data, and thus the congestion probability for each link can be calculated. For those links whose probability is close to 0, we do not consider them as congested links. But for links where congestion occurs frequently, we envelop them to form the busy traffic zone.

The upstream traffic volume can reach the central congestion point of the busy traffic zone after a certain period of time t. Theoretically, it is feasible to use time t to obtain the upstream data collection points. Once the upstream data collection points are determined, the links corresponding to these collection points are used to collect trajectory points. The link speed equals to the average value of instantaneous speeds from these trajectories. It is worth noting that the link speed mentioned in this paper is acquired by considering the trajectory points both in the road link and downstream neighbor intersection during a certain time interval, so that the average speed of all vehicles on the link can be determined more objectively.

The travel time t is a given input value, and it is set to be equal to the state duration of the HMM and the model prediction duration. Since this value is not a resolved one, the core of determining upstream data collection points lies in how to calculate the travel distance according to the duration t. This distance is from the upstream data collection point to the congestion point. Because the travel speed of the vehicle is affected by many factors and becomes unstable, it is very difficult to solve the travel distance or determine the upstream data collection points in real time. In order to improve the practical operability of the engineering problem, we provide the travel speed of each link during a specific large time period of historical dates in advance. This link speed is the average value of instantaneous speeds from all valid trajectory points during that time period. If we obtain the link speed, then the travel time of each link is obtained indirectly according to the formula of distance divided by the speed. Therefore, it is easy to obtain the specific location of the upstream data collection point according to the shortest path, based on the calculated link travel time and prediction duration t.

In addition, we want to mention the relationship between the determination of upstream data collection points and the volatility of vehicle travel time. It is undoubtedly the traffic characteristic values will fluctuate under the influence of downstream dynamic flow, signal timing and other factors. If we want to estimate the traffic characteristic values accurately, it is necessary to reproduce the future short-term vehicle state by dynamic traffic assignment, simulation and other technologies. This kind of prediction issue is difficult, and its prediction accuracy and efficiency require independent analysis through special research. Therefore, this paper does not make an overly detailed analysis of this content in the determination of the upstream collection points. We estimate the travel time in a specific large time period of historical dates. This is actually a quasi-dynamic idea, and the result will be close to the real situation. Moreover, the upstream data collection point only needs to be located on a certain link, and its fault tolerance is very high. A small error of the travel time estimation has little effect on the link search. Even if there is a slight impact, the internal parameter calibration of the HMM will reduce this impact and will not cause major changes to the prediction results of zone congestion pattern.

Once the upstream data collection points are got, we could draw the isochronous boundary by connecting them. Figure 2 is a schematic diagram of the location search for the upstream

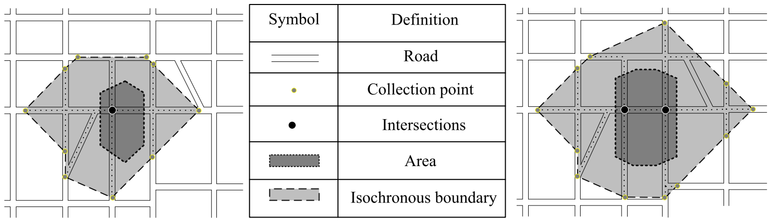


FIGURE 2. Isochronous boundary and upstream traffic state collection point determination.

traffic state collection points corresponding to the two types of busy traffic zones.

IV. MOLDELING

A. DISCRETIZATION OF THE OBSERVATION STATE

After collecting the vehicle speeds of upstream link corresponding to each collection point, we could process them into the average speed V_i by the average aggregation method. However, the average speed is a continuous value, so it is necessary to discretize the average speed to make it be used for the observation state variable of the HMM. This paper uses a more detailed observation state to estimate the hidden state, in which the vehicle speed is divided into 9 discrete values during the discretization operation. According to the traffic analysis report of major cities in China released by Auto Navi Map, the average all-day vehicle speed in Chinese cities is mostly distributed below 40 km/h. Therefore, we conduct the discretization work as follows. If average speed V_i of the upstream links is no more than 40 km/h within the i-th time interval, then the discretization value is $\left[V_i/5\right]$, one exception is that the discrete value is equal to 1 when $V_i = 0$. If the vehicle speed is greater than 40 km/h, then it is uniformly discretized to 9. Therefore, the formula of the observation state value is as follows:

$$\textit{observation state value} = \begin{cases} 1 & V_i = 0 \\ [v_i/5] & 0 < V_i < 40 \\ 5 & V_i \geq 40 \end{cases} \tag{1}$$

B. DISCRETIZATION OF THE HIDDEN STATE

The congestion pattern of a busy traffic zone is represented by the combination form of all the internal links' congestion states. So, the definition of link's congestion state should be the primary consideration.

As to the definition of traffic congestion state for road, it is related to the studied scenario. Indexes such as vehicle delay, queue length, are usually used under the condition of intersection. Travel speed is more often used to judge road link congestion. Even if using the same index, the threshold values for congestion judgement are varied in different countries. They are determined by the limit speed, travel environment, driving behavior, etc. In the following content, we could take the Chinese congestion classification standard as an example to construct the formula of traffic congestion. For main roads: Level I (vehicle speed $\geq$ 30 km/h) indicates smooth travel on the link, Level II (30 km/h $>$ vehicle speed $\geq$ 15 km/h) indicates slow travel on the link, and Level III (15 km/h $>$ vehicle speed) indicates congestion on the link. The congestion state of the r-th link in a busy traffic zone at the i-th period is expressed as L_{ri}. Therefore, the formula of the traffic congestion state of this link is designed as follows:

$$L_{ri} = \begin{cases} \mathrm{I} & V_{ri} \geq 30 \\ \mathrm{II} & 15 \leq V_{ri} < 30 \\ \mathrm{III} & 0 \leq V_{ri} < 15 \end{cases} \tag{2}$$

where the speed of the r-th link at the i-th period is V_{ri}. Then, the hidden state of the traffic busy zone at the i-th period is set as follows:

$$L_{Ri} = 1 + \sum_{r=1}^{|R|} (L_{ri} - 1) \cdot 3^{|R|-r} \tag{3}$$

where R is the set of all links in the area. Take Figure 3 as an example, there are 3 congested links in the busy traffic zone, and thus there are 3^3 (27) hidden states. According to formula (3), various hidden states are labeled in Figure 3.

C. CONSTRUCTION OF THE HMM

The HMM is a time series model based on the Markov model. It establishes the state transition mechanism of two types of related variables to achieve a prediction. The model is widely used to solve scientific problems in the fields of image processing, speech recognition, pattern recognition and signal processing. Recently, it has applications in short-term traffic flow prediction, traffic jam situation recognition, and vehicle trajectory prediction [32]–[37].

Since the upstream traffic state and congestion pattern in a busy traffic zone are both random and correlated, the HMM's state transition matrix and observation probability matrix are used to express these randomly associated features to predict the varying trend in the hidden state of the congestion pattern in the busy traffic zone in a short time. Assuming the HMM $= \lambda$, λ is a tuple with the following five elements: $\lambda = (S, O, \Pi, A, B)$. In this paper, the HMM shown in Figure 4 is constructed based on the information from these five elements.

When the HMM $= \lambda$, Π, A, B are as follows:

1) $\Pi = \{\pi_i\} (1 \leq i \leq N)$ is the initial hidden state probability. π_i is the probability when the hidden state s_2 is equal to x_i at time $\tau = 2$, where $\pi_i = P\{s_2 = x_i\}$.

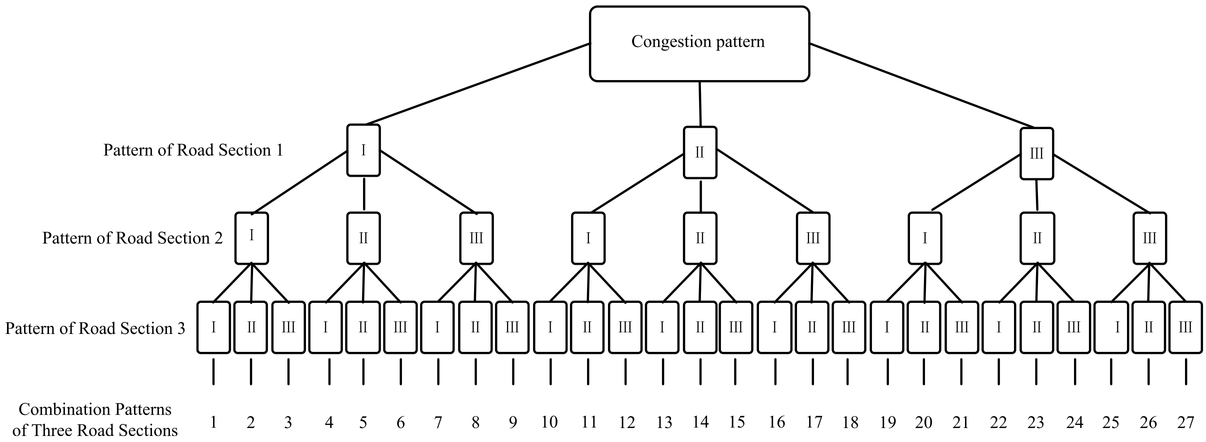


FIGURE 3. Hidden state (congestion pattern) classification.

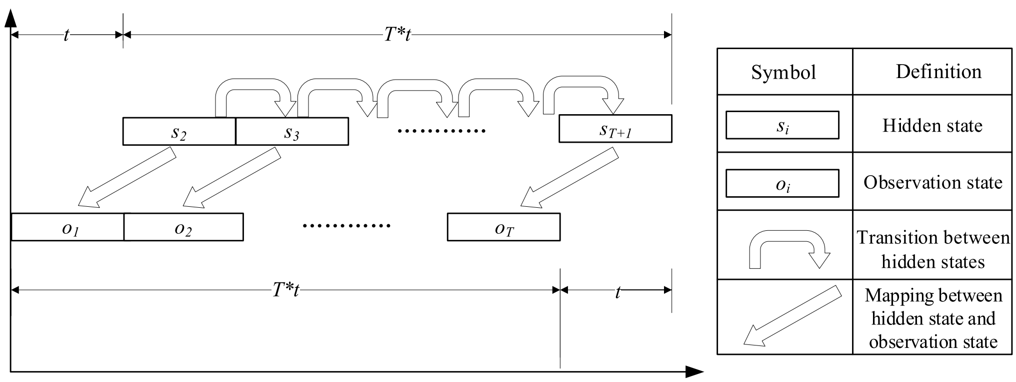


FIGURE 4. Illustration of the state transition.

2) $A = \{a_{ij}\} (1 \leq i, j \leq N)$ is the state transition probability matrix. a_{ij} is the transition probability when the hidden state changes from s_τ (equals x_i at time τ) to $s_{\tau+1}$ (equals x_j at time $\tau + 1$), where $a_{ij} = P(s_{\tau+1} = x_j \,|s_\tau = x_i)$.

3) $B = \{b_j(k)\} (1 \leq j \leq N, 1 \leq k \leq M)$ is the observation probability matrix. $b_j(k)$ is the probability that the observable state o_τ equals y_k at time τ assumes that the hidden state $s_{\tau+1}$ equals to x_j at time $\tau + 1$, where $b_j(k) = P(o_\tau = y_k \,|s_{\tau+1} = x_j)$.

Figure 4 shows the mapping relation between hidden and observation state sequences in terms of temporal dimension. The observation state sequence is expressed as $O = (o_1, o_2, \ldots, o_T)$, and the hidden state sequence is expressed as $S = (s_1, s_2, \ldots, s_T)$. This paper first estimates π_i, a_{ij}, $b_j(k)$ based on a large number of historical samples then combines it with observation sequence $O = (o_1, o_2, \ldots, o_T)$ to calculate the probability of each hidden state in the $T + 1$ period and selects the hidden state with the highest probability as the predicted value of the congestion pattern. The probability of hidden state x_i appearing in period $T + 1$ is $\gamma_{T+1}(i)$.

$$\gamma_{T+1}(i) = P(s_{T+1} = x_i \,|O, \lambda) = \frac{P(s_{T+1} = x_i, O \,|\lambda)}{P(O \,|\lambda)} \quad (4)$$

To simplify the subsequent derivation process, the denominator and numerator of formula (4) are expressed as follows. Here, $\alpha_{T+1}(i)$ is defined as the forward probability. The forward probability is the probability when observation state sequence is $(o_1, o_2, \ldots, o_T)$ and the hidden state is x_i in period $T + 1$.

$$P(s_{T+1} = x_i, O \,|\lambda) = \alpha_{T+1}(i) \quad (5)$$

$$P(O \,|\lambda) = \sum\nolimits_{i=1}^{N} \alpha_{T+1}(i) \quad (6)$$

According to the principle of the forward algorithm, the relationship between $\alpha_{T+1}(i)$ and $\alpha_T(j)$ in the previous period is established as follows:

$$\alpha_{T+1}(i) = \left[\sum\nolimits_{i=1}^{N} \alpha_T(i) a_{ji}\right] b_i(o_T), \; i = 1, 2, \ldots, N \quad (7)$$

Iterate forward step by step according to the above formula to obtain the forward probability of period $\tau = T$, $T - 1, \ldots, 2$ until $\tau = 2$, and the result of $\alpha_2(i)$ is as follows:

$$\alpha_2(i) = \pi_i b_i(o_1), \quad i = 1, 2, \ldots, N \quad (8)$$

Thus far, all intermediate variables $\gamma_{T+1}(i)$ in the calculation can be back calculated based on the results of equation (8).

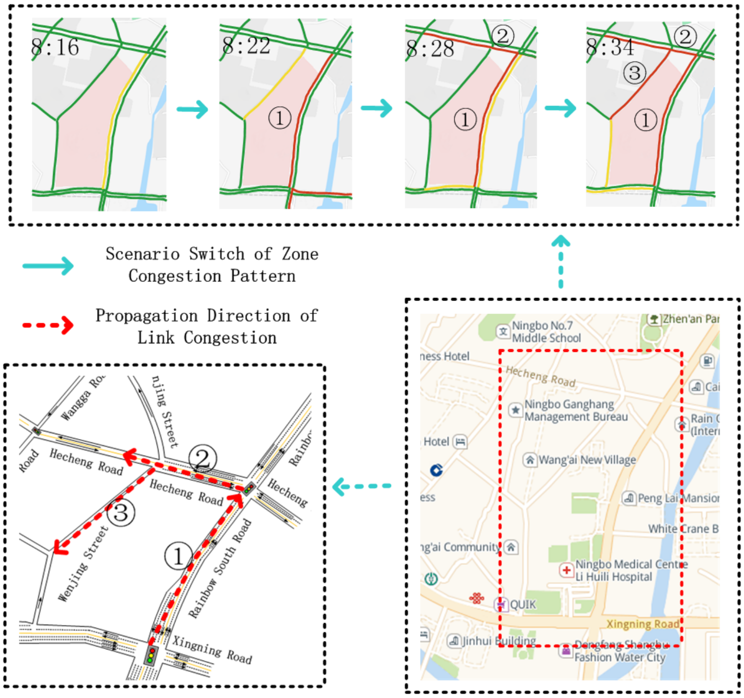


FIGURE 5. Schematic diagram of congestion transfer.

The value ψ_{T+1} corresponds the most likely situation for the hidden state sequence, where x_i is the predicted hidden state in period $T+1$.

$$\psi_{T+1} = \max_{x_i \in X} \gamma_{T+1}(i) \tag{9}$$

V. EMPIRICAL ANALYSIS

A. DATA COLLECTION

The selected busy traffic zone is enclosed by the four roads of Jiangdongnan Road, Rainbow South Road, Xingning Road and Xinhe Road in Ningbo City. This area not only has traffic generating sources such as hospitals, schools, and residential areas but is also an important channel connecting the core functional areas of the city. It undertakes much commuter traffic. According to the statistics, there are more than 20,000 vehicles per day, and congestion often occurs.

In the statistical analysis, the congested links in the study area were mainly located around Li Huili Hospital. The upper part of Figure 5 shows the change in congestion state of the busy traffic zone every 6 mins during the period of 8:10-8:34 on a working day, in which the solid red lines represents the congested links. There are at least three typical patterns of congestion in the hospital area, and there is a mutual evolution process among these patterns. Congestion occurs first in link 1. As congestion spreads upstream, congestion occurs in links 2 and 3, and the area reaches maximum congestion. In the lower left subgraph, the red dotted lines represent the typical congestion propagation direction.

B. DATA SOURCES

The data used in this case is the original taxi trajectory data located in the selected busy traffic zone and the upstream data collection points. The sampling interval of the original taxi trajectory data is 15 s. In addition, due to the frequent congestion during peak hours, the data were collected mainly during the peak hours of working days: 6:00-9:00 am and 3:00-6:00 pm (July and August 2019). The collected data sequence is divided into two categories, in which 80% is used to train the HMM to obtain the model parameters, and the other 20% of the data is used to verify the effect of the model.

There are many anomalies and errors in the original taxi GPS data. Lots of researches have conducted the data cleaning work for this kind of data (such as Sun *et al.* [38]). We eliminated problem data according to the following principles.

1) The distance from the GPS data to the centerline of the nearest link is more than 30 m.

2) The single-point conversion speed of urban roads is more than 100 km/h.

3) Under the influence of high buildings, tunnels, GPS failures, etc., the trajectory data are lost and discontinuous.

4) Although the trajectory data are continuous, when the continuous overlap time of the trajectory data is longer than 2 mins (general signal intersection maximum cycle) and exceeds the normal parking time of an intersection, this indicates that a taxi may be waiting for passengers, turning on or off, or is temporarily out of service.

After the trajectory data cleaning is completed and the road network is matched, a reasonable t value can be designed

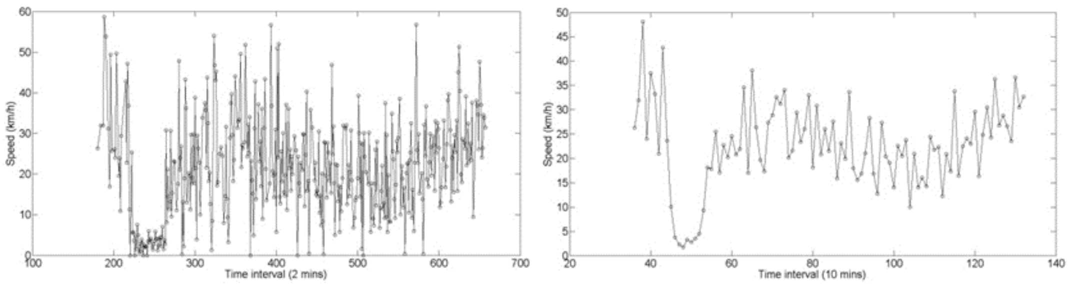


FIGURE 6. Road speed changes by using varied time intervals.

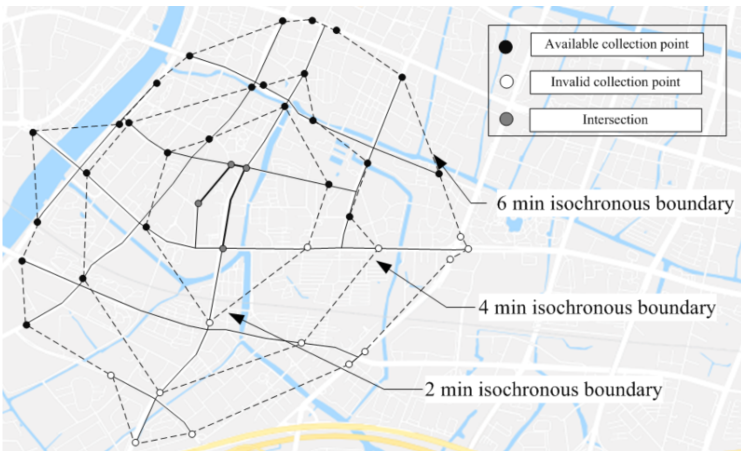


FIGURE 7. Position of the upstream data collection point during morning peak periods.

by enumerating from small to large. Then according to the t value, the average speed of all taxi trajectories on the links can be calculated. If t is too small, many links have no taxi trajectories within certain statistical time periods and thus there is no speed value. For example, when analyzing taxi data, it is found that when t is set to 1 min, more than half of the city links do not have taxi trajectory data during certain periods within the peak period. When t is increased to 2 mins, the speed of most links becomes more regular, but there are still some links with fewer taxis, and the speeds of the vehicles fluctuate greatly and are disordered in various periods (as shown on the left of Figure 6). If t is increased to 10 mins, the speed loss rate on the road is reduced and the speed is more stable, smooth and orderly, as shown on the right of Figure 6. However, since both the prediction time span and the traffic state time interval are set with t, a long t would increase the uncertainty of the upstream and downstream traffic flows and thus increase the prediction error, which is not conducive to the realization of the short-term prediction function. In this case, the upper limit of t is set to 6 mins. In view of the large traffic volume in the studied busy traffic zone and the dense GPS data from taxis, the lower limit of t is set to 2 mins, which allows for the accurate collection of real-time vehicle speed data on most links. In addition, the vehicle speed is supplemented by interpolation for situations in which data are still missing.

C. POSITION DETERMINATION OF THE UPSTREAM DATA COLLECTION POINTS

First determine the data collection points of the upstream links. Then, determine the reachable range boundary. This reachable range is formed by connecting the upstream data collection points. Finally, calculate the speed of the links where the upstream data collection points are located, and determine the speed state sequence.

Figure 7 shows the positions of the upstream data collection points at three different prediction time spans t (2 mins, 4 mins, 6 mins) during the morning peak period. Each intersection in the busy traffic zone is taken as the end point of travel, framing three types of equal-time driving reachable range boundaries (marked by dotted lines in Figure 7). It is necessary to remove the invalid collection points and keep only the valid collection points in the set of all collection points. Among them, the invalid collection points refer to the collection points whose traffic flow direction is opposite to the direction of the road congestion in the zone (marked by white dots in Figure 7). Finally, the remaining valid collection points are marked with black dots. The vehicle speed on the links where the upstream collecting point is located is collected. Then, the average vehicle speed is obtained. Therefore, the vehicle speed state sequence of the upstream links is formed as the observation state input of HMM.

Congestion patterns S

…	10	10	1	10	20	10	10	20	…

$$a_{10,20} = \frac{\#\ of\ pairs\ with\ values (10,20)}{total\ \#\ of\ pairs}$$

Congestion patterns	1	2	3	10	11	12	14	17	18	19	20	21	23	24	25	26	27
1	0.8563	0.0180	0	0.0958	0	0.0060	0	0	0	0.0060	0	0.0120	0	0	0	0	0.0060
2	0.2000	0.3000	0.2000	0	0	0.1000	0	0	0	0	0.2000	0	0	0	0	0	0
3	0	0	0.2857	0.4286	0	0	0	0	0	0	0.1429	0	0	0.1429	0	0	0
10	0.1206	0	0	0.6176	0.0588	0.0147	0	0	0	0.1147	0.0294	0	0.0147	0	0	0.0147	0.0147
11	0.0357	0.0357	0	0.0357	0.6429	0.0357	0	0	0	0.0357	0.0714	0.0714	0	0	0	0	0.0357
12	0	0	0.1333	0	0.1333	0.4000	0	0	0	0.2000	0	0.0667	0	0	0	0	0.0667
14	0	0	0	0	0	0	0.4000	0.2000	0	0	0	0	0	0	0	0.2000	0.2000
17	0.1000	0	0	0	0	0	0.1000	0.5000	0.1000	0	0	0	0	0	0	0.2000	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0000
19	0.0244	0.0244	0	0.0244	0.0244	0.0244	0	0	0	0.6341	0.1463	0.0488	0	0	0.0244	0.0244	0
20	0	0.0339	0.0169	0.0508	0	0.0339	0	0	0	0.0678	0.6780	0.0847	0.0169	0	0	0.0169	0
21	0	0	0	0.0333	0	0.0333	0	0	0	0.1333	0.1333	0.6000	0	0.0667	0	0	0
23	0	0	0	0	0.2500	0	0	0	0	0	0.5000	0	0.2500	0	0	0	0
24	0	0	0	0	0.0667	0	0	0.0667	0	0	0	0	0	0.4000	0	0.0667	0.4000
25	0	0	0	0	0	0.2500	0	0	0	0.2500	0	0	0	0	0.2500	0.2500	0
26	0.0370	0	0	0.0370	0	0	0.0370	0.0370	0	0	0	0	0	0	0	0.3704	0.4815
27	0.0435	0	0	0.0217	0.0217	0	0.0217	0.0435	0.0435	0	0	0	0.0217	0.1304	0.0435	0.1957	0.4130

FIGURE 8. Transition probability matrix for morning peak periods under condition t=2.

D. HIDDEN MARKOV INPUT PARAMETER STATISTICS

In this paper, observation state sequence O and hidden state sequence S of the congestion pattern in the busy traffic zone of Figure 3 are taken as input, and the model parameters are statistically obtained. The obtained parameters mainly include the value of a_{ij} in state transition matrix A and the value of $b_j(k)$ in observation probability matrix B. This article defines a total of 27 congestion patterns in the study area (the busy traffic zone), and thus matrix A should be a $27 * 27$ matrix. It is necessary to eliminate 10 patterns (4, 5, 6, 7, 8, 9, 13, 15, 16, 22) that never statistically appeared. The result of the transition probability among the remaining 17 congestion patterns under a specific prediction scenario is shown in Figure 8.

For a specific scenario where state time unit t is 2 mins, Figure 8 shows the state transition matrix and its element acquisition. The state transition matrix shows that the congestion pattern of the busy traffic zone can evolve to any possible pattern with a certain probability. Under normal circumstances, the probability value on the diagonal is the largest. This indicates that in reality, the congestion pattern in the busy traffic zone will generally remain for a period of time before transferring. At the same time, other transfer phenomena of congestion patterns are consistent with the real-world observations. For example, except for the possibility of maintaining the current pattern, pattern 1 is most likely to evolve to pattern 10 with a probability of 9.58%. Pattern 10 may evolve to pattern 1 with a probability of 12.06% or continue to evolve to pattern 19 with a probability of 11.47%.

Figure 9 shows the observation probability matrix and its element acquisition in a specific scenario. The observation probability matrix represents the transition probability

Congestion pattern

…	1	1	10	1	10	10	19	…

Observation symbol

…	9	8	7	8	7	6	6	…

$$b_{10,7} = \frac{\#\ of\ pairs\ with\ values (10,7)}{total\ \#\ of\ pairs}$$

Congestion pattern	Observation symbol								
	1	2	3	4	5	6	7	8	9
1	0	0.0060	0.0120	0.0120	0	0	0	0.6108	0.3593
2	0	0.2000	0.1000	0	0.2000	0	0.1000	0.3000	0.1000
3	0	0	0.1429	0	0	0.1429	0.7143	0	0
10	0	0	0	0	0.0725	0.3478	0.5217	0.0580	0
11	0	0	0	0	0.1071	0.3929	0.3929	0.1071	0
12	0	0	0	0	0.2000	0.4667	0.3333	0	0
14	0	0	0	0	0.4000	0.2000	0.4000	0	0
17	0	0	0.1000	0	0.2000	0.5000	0.2000	0	0
18	0	0	0	0	0.3333	0.6667	0	0	0
19	0	0	0	0	0.1220	0.6098	0.2439	0.0244	0
20	0	0	0	0	0.0678	0.6271	0.2712	0.0339	0
21	0	0	0	0	0	0.5000	0.4667	0.0333	0
23	0	0	0	0	0	0.7500	0.2500	0	0
24	0	0	0	0	0.0667	0.8000	0	0.1333	0
25	0	0	0	0	0.0712	0.4863	0.3241	0.2324	0
26	0	0	0	0	0.0370	0.3333	0.4815	0.1481	0
27	0	0	0	0	0.1087	0.5217	0.3043	0.0652	0

FIGURE 9. Observation probability matrix and element acquisition.

between the congestion pattern of the busy traffic zone and the speed state of the upstream links. Figure 9 shows that the speed state distribution in the upstream link is concentrated at 5, 6, and 7; that is, the road speed is concentrated at 20 km/h $\sim$ 35 km/h. In addition, congestion pattern 1 of the observation probability matrix (a form in which the travel on all the three links are smooth) corresponds more to higher

TABLE 2. Comparison of the HMM with other method.

Sequence lengths T	Time unit t (min)	Working day					
		Morning peak		Evening peak		Average	
		HMM (%)	ARIMA (%)	HMM (%)	ARIMA (%)	HMM (%)	ARIMA (%)
2	2	75.9	76.1	71.1	74.4	73.5	75.3
	4	69.1	73.3	67.1	72.6	68.1	73.0
	6	62.5	71.7	63.8	70.1	63.1	70.9
3	2	82.1	77.4	78.9	75.6	80.5	76.5
	4	78.2	75.2	74.3	73.3	76.3	74.3
	6	72.1	73.8	67.3	71.2	69.7	72.5
4	2	83.4	77.5	79.3	76.5	81.4	77.0
	4	82.0	75.3	74.9	74.9	78.5	75.1
	6	76.0	74.2	72.6	72.8	74.3	73.5
5	2	83.4	77.5	81.3	76.5	82.4	77.0
	4	82.3	75.5	79.4	74.9	80.9	75.2
	6	77.2	74.3	76.6	73.0	76.9	73.7
6	2	83.4	77.6	81.2	76.6	82.3	77.1
	4	82.2	75.6	79.3	75.0	80.8	75.3
	6	77.2	74.5	76.5	73.1	76.9	73.8

speed states 8 and 9. This situation occurs more frequently in the period before 7 o'clock. It is the time very close to the beginning of congestion. The vehicles are fewer and vehicle speed is relatively fast, and the traffic is smooth in the busy traffic zone.

E. RESULTS AND DISCUSSIONS

To evaluate the effect of the congestion pattern prediction in the busy traffic zone, this article uses accuracy as the evaluation index as follows:

$$precision = U_c / U_q \tag{10}$$

In formula (10), U_c is the number of accurate predictions, and U_q is the number of total predictions.

Figure 10 compares the accuracy of the HMM with three time units t (2 mins, 4 mins, 6 mins) combined with different predicted sequence lengths T (2-6).

In Figure 10, as the length of the prediction sequence increases, the prediction accuracy increases continuously, while the degree of increase decreases and gradually stabilizes. Longer prediction sequence length contains more information about the evolution of the traffic state. So the prediction of traffic state will be more accurate. But the earlier information in the sequence has very little impact on the subsequent traffic state. When the prediction sequence length reaches 4, the accuracy rate is already high, and there is very little change of accuracy if it still increases. In addition,

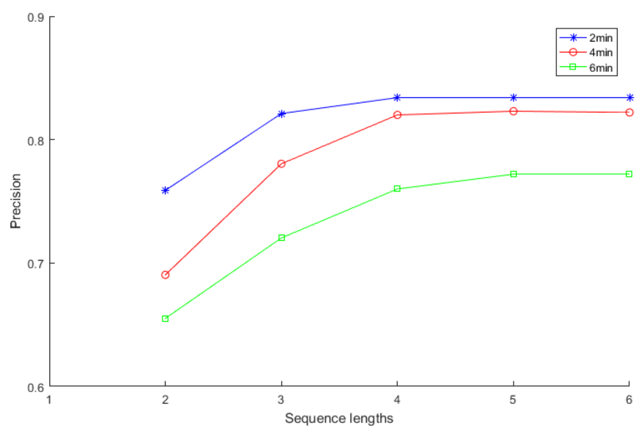


FIGURE 10. Prediction accuracy of the various prediction sequence lengths.

a small time unit t indicates that the upstream boundary is close to the core point of the congested area and the prediction accuracy will be high. In particular, when the time unit t decreases from 6 mins to 4 mins, the prediction accuracy rate increases sharply.

Table 2 shows the prediction accuracy of two time series models (HMM and ARIMA) as a function of T and t. Among them, the ARIMA model is a traditional regression prediction model and only uses the historical data of the target link to predict the speed.

According to the prediction results of the two types of models, they show a common feature: as the length of the prediction sequence increases or the time unit decreases, the prediction accuracy rate would increase. However, their change speeds are different. When the length of the prediction sequence and the time unit change, the prediction accuracy of the ARIMA model basically does not change significantly, but the prediction accuracy of the HMM is very sensitive. When T is 2, the ARIMA model has higher prediction accuracy than the HMM, because the prediction sequence length is relatively short and contains less historical information. However, when T increases to 3, the prediction accuracy of the HMM in condition of short time unit performs better. When T is larger than 3, the prediction accuracy of the HMM model is 4.9% higher than the ARIMA model in average. In general, the HMM has the best prediction accuracy of 83.4% and 80% during morning and afternoon peaks respectively, which is 5.8% and 4% better than the best prediction accuracy rate of ARIMA model in these two cases.

According to the above data analysis, we know the HMM has a higher prediction accuracy rate, compared with the ARIMA model. The reason is that the HMM could consider the temporal and spatial characteristics, and take full advantage of the historical traffic information which containing these characteristics. In order to use more historical traffic information, a note is that we should set the sequence length (T) as a number larger than 3. As the length of the prediction sequence increases, the recorded historical information also increases. In these conditions, the HMM can obtain a stable and high accuracy rate.

VI. SUMMARY AND CONCLUSION

This paper uses the taxi trajectory data from Ningbo City to propose a method based on the hidden Markov model and uses the speed of the upstream links to predict the congestion pattern in the busy traffic zone. First, the trajectory data of taxis are processed and trained. Then establish the state transition matrix to simulate the random change of congestion patterns in the busy traffic zone. At the same time, consider the influence of the upstream traffic state to create the observation probability matrix. At last, utilize the above two matrixes and HMM model to achieve the prediction of congestion pattern in the busy traffic zone.

It can be seen from the case analysis results that the prediction accuracy of this method is better than that of the traditional ARIMA model. Although the prediction accuracy rate is slightly lower than the ARIMA model when the sequence length is small (less traffic information), with an increase in the sequence length, the prediction accuracy rate increases rapidly. In this case, the prediction accuracy rate of the HMM outperforms that of the ARIMA model and maintains a high prediction accuracy rate. With the extension of the state time unit, even though the prediction accuracy of both models has decreased, the prediction accuracy of the HMM remains higher than that of the ARIMA model.

The proposed HMM can help traffic management agencies adopt an active traffic management strategy. For example, to predict the future congestion pattern in a short period of time by collecting the dynamic speed of the upstream links of the congestion area, traffic management and dredging can be carried out in advance to prevent urban congestion.

Regarding the case in this paper, there are many phenomena of state transitions to itself. We would try to adopt an improved method named S-HMM in the future, which may be able to effectively avoid degradation problem. Once obtaining more sufficient data, we could further try some deep learning methods such as LSTM to predict the traffic state of a busy traffic zone.

REFERENCES

[1] C. Antoniou, H. N. Koutsopoulos, and G. Yannis, "Dynamic data-driven local traffic state estimation and prediction," *Transp. Res. C, Emerg. Technol.*, vol. 34, pp. 89–107, Sep. 2013.

[2] R. Noroozi and B. Hellinga, "Real-time prediction of near-future traffic states on freeways using a Markov model," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2421, no. 1, pp. 115–124, Jan. 2014.

[3] B. L. Smith, B. M. Williams, and R. Keith Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transp. Res. C, Emerg. Technol.*, vol. 10, no. 4, pp. 303–321, Aug. 2002.

[4] M. G. Karlaftis and E. I. Vlahogianni, "Memory properties and fractional integration in transportation time-series," *Transp. Res. C, Emerg. Technol.*, vol. 17, no. 4, pp. 444–453, Aug. 2009.

[5] S. R. Chandra and H. Al-Deek, "Predictions of freeway traffic speeds and volumes using vector autoregressive models," *J. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 53–72, May 2009.

[6] M. Levin and Y. D. Tsao, "On forecasting freeway occupancies and volumes," *Transp. Res. Rec., J. Transp. Res. Board*, no. 773, pp. 47–49, Jan. 1980.

[7] M. S. Ahmed and A. R. Cook, "Analysis of freeway traffic time-series data by using box-jenkins techniques," *Transp. Res. Rec.*, vol. 722, no. 1, pp. 1–9, 1979.

[8] M. G. Karlaftis and E. I. Vlahogianni, "Statistical methods versus neural networks in transportation research: Differences, similarities and some insights," *Transp. Res. C, Emerg. Technol.*, vol. 19, no. 3, pp. 387–399, Jun. 2011.

[9] Y. Zhang and Y. Liu, "Traffic forecasting using least squares support vector machines," *Transportmetrica*, vol. 5, no. 3, pp. 193–213, Sep. 2009.

[10] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Transp. Res. B, Methodol.*, vol. 18, no. 1, pp. 1–11, Feb. 1984.

[11] S. Innamaa, "Self-adapting traffic flow status forecasts using clustering," *IET Intell. Transp. Syst.*, vol. 3, no. 1, pp. 67–76, Mar. 2009.

[12] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: A deep learning approach for short-term traffic forecast," *IET Intell. Transp. Syst.*, vol. 11, no. 2, pp. 68–75, Mar. 2017.

[13] S. Jeon, B. Hong, and B. Kim, "Big data processing for prediction of traffic time based on vertical data arrangement," in *Proc. IEEE 6th Int. Conf. Cloud Comput. Technol. Sci.*, Dec. 2014, pp. 326–333.

[14] F. Han, "Short-term traffic flow prediction based on multivariable phase space reconstruction and LSSVM," *J. Inf. Comput.al Sci.*, vol. 11, no. 9, pp. 3209–3217, Jun. 2014.

[15] C. Cao and J. Xu, "Short-term traffic flow predication based on PSO-SVM," in *Proc. Int. Conf. Transp. Eng.*, Jul. 2007, pp. 167–172.

[16] Z. Liu, W. Du, D. M. Yan, G. Chai, and J. H. Guo, "Short-term traffic flow forecast based on combination of K nearest neighbor algorithm and support vector regression," *J. Highway Transp. Res. Develop.*, vol. 34, no. 5, pp. 122–128, 2017.

[17] J. Y. Sun, R. J. Mou, and Y. H. Liu, "Study on short-term traffic state forecasting model of SVM considering proportion of large vehicles," *J. Highway Transp. Res. Develop.*, vol. 35, no. 10, pp. 126–132, 2018.

[18] S. S. Xing, Y. L. Gu, L. J. Shen, X. L. Yang, and G. X. Zhuang, "A traffic congestion prediction model based on vehicle speed of urban expressways," *J. Transp. Inf. Saf.*, vol. 197, no. 2, pp. 54–60, 2016.

[19] M. Chen, X. Yu, and Y. Liu, "PCNN: Deep convolutional networks for short-term traffic congestion prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 11, pp. 3550–3559, Nov. 2018.

[20] J. Tang, F. Liu, Y. Zou, W. Zhang, and Y. Wang, "An improved fuzzy neural network for traffic speed prediction considering periodic characteristic," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 9, pp. 2340–2350, Sep. 2017.

[21] K. P. Zhang, Z. J. Liu, and L. Zheng, "Short-term prediction of passenger demand in multi-zone level: Temporal convolutional neural network with multi-task learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 29, no. 4, pp. 1480–1490, Apr. 2020, doi: 10.1109/tits.2019.2909571.

[22] C. Qiao, L. Dai, S. Lin, and X. Li, "Traffic status prediction based on random restart hill-climbing," *J. Transp. Syst. Eng. Inf. Technol.*, vol. 16, no. 1, pp. 162–168, 2016.

[23] Q. B. Wei, Z. C. He, X. S. Zheng, C. J. Chen, and J. F. Yang, "Prediction of urban traffic performance index considering multiple factors," *J. Transp. Syst. Eng. Inf. Technol.*, vol. 17, no. 1, pp. 74–81, 2017.

[24] S. V. Kumar and L. Vanajakshi, "Short-term traffic flow prediction using seasonal ARIMA model with limited input data," *Eur. Transp. Res. Rev.*, vol. 7, no. 3, pp. 1–9, Sep. 2015.

[25] A. M. H. Zhang, P. Zhang, and J. D. Cao, "Prediction of incident influence scope on expressway based on demand forecasting and CTM," *Transp. Res.*, vol. 1, no. 1, pp. 48–53, 2015.

[26] C. F. Daganzo, "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transp. Res. B, Methodol.*, vol. 28, no. 4, pp. 269–287, Aug. 1994.

[27] Y. Wang, M. Papageorgiou, and A. Messmer, "Predictive feedback routing control strategy for freeway network traffic," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1856, no. 1, pp. 62–73, Jan. 2003.

[28] G. Zhu, K. Song, P. Zhang, and L. Wang, "A traffic flow state transition model for urban road network based on hidden Markov model," *Neurocomputing*, vol. 214, pp. 567–574, Nov. 2016.

[29] W. Liu, Y. Zheng, S. Chawla, J. Yuan, and X. Xing, "Discovering spatio-temporal causal interactions in traffic data streams," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining KDD*, 2011, pp. 1010–1018.

[30] L. X. Pang, S. Chawla, W. Liu, and Y. Zheng, "On detection of emerging anomalous traffic patterns using GPS data," *Data Knowl. Eng.*, vol. 87, pp. 357–373, Sep. 2013.

[31] V. W. Chu, R. K. Wong, W. Liu, and F. Chen, "Causal structure discovery for spatio-temporal data," in *Database Systems for Advanced Applications* (Lecture Notes in Computer Science). Cham, Switzerland: Springer, 2014, pp. 236–250.

[32] Y. Qi and S. Ishak, "A hidden Markov model for short term prediction of traffic conditions on freeways," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 95–111, Jun. 2014.

[33] X. Xiong, L. Chen, and J. Liang, "A new framework of vehicle collision prediction by combining SVM and HMM," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 699–710, Mar. 2018.

[34] A. Bezuglov and G. Comert, "Short-term freeway traffic parameter prediction: Application of grey system theory models," *Expert Syst. Appl.*, vol. 62, pp. 284–292, Nov. 2016.

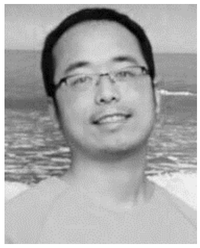
[35] J. Barros, M. Araujo, and R. J. F. Rossetti, "Short-term real-time traffic prediction methods: A survey," in *Proc. Int. Conf. Models Technol. Intell. Transp. Syst. (MT-ITS)*, Jun. 2015, pp. 132–139.

[36] L. Rapant, K. Slaninová, J. Martinovič, and T. Martinovič, "Traffic speed prediction using hidden Markov models for Czech Republic highways," in *Agent and Multi-Agent Systems, Technology and Applications*. Cham, Switzerland: Springer, 2016, pp. 187–196.

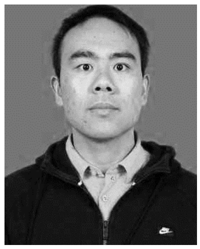
[37] N. Miller, M. A. Thomas, J. A. Eichel, and A. Mishra, "A hidden Markov model for vehicle detection and counting," in *Proc. 12th Conf. Comput. Robot Vis.*, Jun. 2015, pp. 269–276.

[38] D. Sun, K. Zhang, and S. Shen, "Analyzing spatiotemporal traffic line source emissions based on massive didi online car-hailing service data," *Transp. Res. D, Transp. Environ.*, vol. 62, pp. 699–714, Jul. 2018.

TINGTING SUN was born in Dezhou, Shandong, China. She received the engineering degree in traffic from Shandong Jiaotong University, in 2017. She is currently pursuing the master's degree with the Faculty of Maritime and Transportation, Ningbo University. Her main research interest includes traffic transportation.



ZHENGFENG HUANG received the Ph.D. degree in transportation planning and management from Southeast University, Nanjing, China, in 2014. He spent one year as a joint doctoral student with the University of Wisconsin, Madison, from 2012 to 2013. Since 2014, he has been with the Faculty of Maritime and Transportation, Ningbo University, China, where he is currently an Associate Professor. His research interests include proactive traffic control and emergency evacuation.



HONGDONG ZHU was born in Xuzhou, Jiangsu, China. He received the engineering degree in naval architecture and marine from Chongqing Jiaotong University, in 2015. He is currently pursuing the master's degree with the Faculty of Maritime and Transportation, Ningbo University. His main research interest includes traffic transportation.

YANHAO HUANG was born in Chengdu, Sichuan, China, in 1997. He received the bachelor's degree in electrical engineering and automation from Southwest Jiaotong University, in 2019. He is currently pursuing the master's degree with the Faculty of Maritime and Transportation, Ningbo University. His main research interest includes traffic transportation.

PENGJUN ZHENG received the Ph.D. degree in transportation planning and management from the University of Southampton. He was with the University of Southampton for many years. He is currently a Guest Professor with the University of Southampton and a Professor with the Faculty of Maritime and Transportation, Ningbo University, China. His research interests include intelligent transportation engineering, and transportation information and control.