IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 7, 2019, accepted September 3, 2019, date of publication September 6, 2019, date of current version September 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2939787

A Novel Approach for Online Car-Hailing Monitoring Using Spatiotemporal Big Data

TONG ZHOU[1,2,3], WENZHONG SHI[3], XINTAO LIU[3], FEI TAO[1,2], ZHEN QIAN[1], AND RUIJIA ZHANG[4]

[1]School of Geographical Sciences, Nantong University, Nantong 226007, China
[2]Key Laboratory of Virtual Geographical Environment, MOE, Nanjing Normal University, Nanjing 210046, China
[3]Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong
[4]Institute of College of Foreign Studies, Nanjing Agriculture University, Nanjing 210095, China

Corresponding author: Wenzhong Shi (lswzshi@polyu.edu.hk)

This work was supported in part by the National Natural Science Foundation of China under Grant 41301514 and Grant 41401456, in part by the Nantong Key Laboratory Project under Grant CP12016005, in part by the National College Students Innovation and Entrepreneurship Training Program under Grant 201910304036Z, and in part by the Graduate Research and Innovation Projects of Jiangsu Province under Grant 201910304113Y.

ABSTRACT Car-hailing service has increasingly become popular and fundamentally changed the way people travel in the era of sharing economy. Although such service brings convenience to people's lives, it also causes safety and property concerns. Many studies have been conducted to access the efficiency and effectiveness of car-hailing, but little has been done on its safety monitoring. However, with the rapid development of information technologies such as Internet of Things (IoT), Geographical Information Science (GIS) and automatic monitoring, a more advantageous approach than the current simple drivers screening and testing is feasible. A new model including five indexes i.e. region dangerous index, offset distance of the origin-destination, real-time speed under traffic conditions, vehicle travel time and passenger information, is therefore proposed in this paper based on big data mining of the historical vehicle GPS trajectory data. Experiments were conducted to validate the model in the Gangzha District of Nantong City, China. Several other types of data were used in the experiments, e.g. points of interest (POI), road network data and urban image. The results showed that the proposed model effectively monitored the vehicle when it was driving in a ''potentially dangerous area''. In addition, the model could accurately identify the driver's abnormal driving behaviors, such as bypass and abnormal stop. The prediction accuracy of the experiments was 92.06%, among which the discrimination accuracy of the abnormal stop was 100% and that of the detour was 90.57%. All these validate the applicability of the model for future management systems for car-hailing services.

INDEX TERMS Car-hailing, big data, trajectory data, Internet of Things (IoT), geographical information science (GIS), points of interest (POI), potentially dangerous area.

I. INTRODUCTION

The online car-hailing industry has been a popular commercial travel mode that combines the Internet and the sharing economy [1]. The online car-hailing integrates the traditional operation mode of Cruise and Carpooling, making it easy for people to call driving service through mobile devices at any time or even in remote areas of the city. It greatly facilitates the travel of production and also solves the imbalance between the demand and supply of vehicles. Many existing studies have focused on the factors that affect sharing efficiency, e.g. [2]–[4]. At present, it has become one of the main modes of urban public transport travel. However, due to the low entry threshold and fast development speed of the ''online car-hailing'' industry, especially the imbalanced development in medium and small cities, the comprehensive quality of the workers in the industry is at different levels. The lack of high worker quality makes vehicle safety

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

management more difficult, for example, a series of cases have been reported in which passengers and property safety was significantly infringed [3].

Government departments and management agencies have strengthened the management of laws and regulations, such as via the central-local two-level management mechanism. To a certain extent, it helps solve the above problems [5], [6]. However, it is limited to regulatory supervision, and only raises the threshold for vehicles to join the online ride-hailing platform. As a result, while it works in the verification stage of drivers, it is not effective in strengthening the supervision and management of the whole transportation process. Meanwhile, the rapid development in mobile communications, sensor networks, and geospatial analysis technology allows sufficient hardware and software support for real-time monitoring [7]–[9]. This makes real-time network car supervision possible. However, currently available supervision platforms only focus on real-time location query and historical track playback [10], [11], owing to a large number of net cars managed by the platform [12], [13]. There is a lack of technology to automatically find abnormal vehicles from a large number of vehicles. In fact, the conventional monitoring platform is only semi-automatic, and it requires a lot of manual assistance to deal with special situations. This is not conducive to timely prevention of possible security issues, nor to the rapid emergency work in the event of problems. There is an urgent need for automatic supervision technology for car-hailing safety.

Existing studies have shown that there is a certain correlation between the crime rate and the geographical location [14], [15]. In addition, criminal behavior is highly correlated with spatiotemporal factors, often revealing high incidence periods and some hotspot spatial areas. Moreover, population density, economic status, and regional categories can have significant impacts on such areas [16], [17]. Therefore, it is of great significance to predict criminal behavior based on the spatiotemporal distribution principal. In the present study, we intended to select the sparsely populated and relatively remote areas based on the temporal and spatial characteristics of the crime committed by the driver of online car-hailing [18], [19], To a certain degree, the selected areas are also related to the built-up environment of the city. In the past, there have been researching cases based on Landsat and nocturnal light data. Landsat 8 has a relatively high spatial resolution [20], [21], and the night light data can reflect on the more dynamic part of the building [22], [23]. Combined with the POI (points of interest) data provided by third-party social media websites such as Dazhong Dianping, it can further indicate the prosperous areas of the city [24], [25]. Through the integration of these data, we can extract the areas in the built region that are more prone to criminal behaviors to a certain extent, so as to provide a geographical reference for real-time supervision.

At present, most existing research has been dedicated to online car-hailing issues, for instance, the prediction of possible congestion using vehicle speed and other factors, the prediction of possible hot spots using historical empirical data [28], [29], the discovery of driver bypass fraud using trajectory data [30], [31], and other studies on outlier detection of urban traffic vehicles [32], [33]. Nevertheless, little has been done to monitor the safety of car-hailing considering the multiple driving indexes comprehensively as well as the geographic environment.

A novel approach is proposed in this study, which includes several innovations as following. First, it is a quantitative model that integrates five indicators into a score, namely ''risk score'', to evaluate the driving state of the vehicle. The higher the score, the more likely the vehicle is in an abnormal state. Base on the method, we can easily find several suspect vehicles from a massive set, and this makes our current management objects more streamlined, thus our work efficiency can be improved. Second, the geographical environment is considered in our model, this is also a highlight of our work. According to the high correlation between driver crime and the geographical environment, the practice of quantifying the geographical environment as an evaluation factor will have a huge contribution to the safety detection of vehicles. Finally, the option setting of this model is flexible, in case of different monitoring purposes, different kinds of indexes combination can be selected. The number of factors and the range of the parameters can be adjusted according to different objects. The greater the amount of data accumulated in the future, the more simulations will be performed, and the settings of these indexes and parameters will become more reasonable.

The rest of this paper is organized as follows: Section 2 and 3 introduces the main ideas and design course of the model; study area, data source, data processing are given in Section 4; Section 5 describes the results and discussion, followed by a summary in Section 6.

II. METHODOLOGY AND MODEL

A. MOTIVATION AND CONCEPTS

Criminal behaviors are closely related to temporal factors and tend to concentrate in a high-incidence time period and some hotspot areas. Based on this, a concept of ''potentially dangerous zone'' is proposed and implemented in the present study. A potentially dangerous zone can be an area with sparsely populated (low-vitality area), remotely distributed (non-built-up areas), surrounding hollow-ness (mountains and forests, construction land) or an area where passengers have high rate of criminal record, among which specific areas such as sparsely populated, remote and open zones can be regulated according to regulatory needs and actual conditions. The data show that the proportions of crimes committed by online car-hailing drivers differ in different time periods. Accordingly, one day can be divided into different periods: high-incidence section, potential danger section, and safety section [3]. Based on this concept, this paper assigns corresponding risk factors to the dangerous areas in different time periods to achieve reasonable supervision of urban areas.

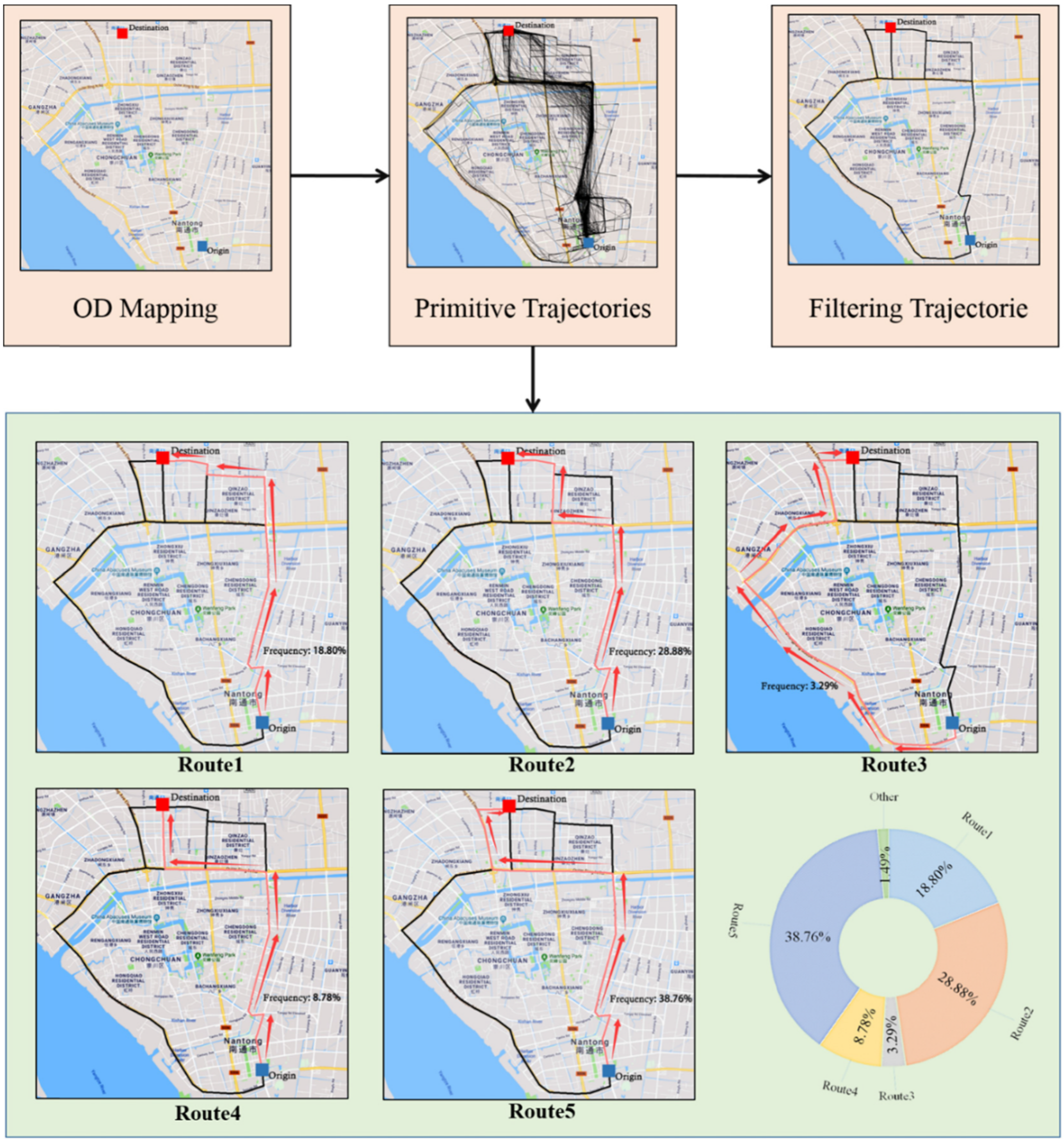


FIGURE 1. **Analysis process of origin-destination frequency (After the origin trajectory data filtering, the main routes can be found, and the frequency of each route can be counted).**

The vehicle GPS trajectory describes the current driving state and can be used not only for traffic monitoring/forecasting [18], personalized route recommendation [37], driving route selection [20] but also for monitoring the driver's abnormal driving behavior [38]. In the view of the monitoring of passenger safety status, this study will use the big data analytical technology to select appropriate model factors to be combined with the potential danger areas for a comprehensive analysis. When a vehicle is driving in a potential danger zone, the model will calculate the risk score. Combined with the particularity of cases of online ride-hailing drivers of high risk to infringe on passengers, this model contains the processing of passenger gender and quantity information, that is, when the passengers are female and the number is small, the level of ride safety will be reduced. In doing so, the model can effectively provide security assurance.

Based on big data analysis technology, this study aims to mine the origin and destination (OD) frequency law of different orders of vehicles and then selects the model superiority factors. The OD frequency law refers to the frequency law of different order routes of vehicles based on the same starting position and ending position. The route with high frequency is the normal driving route, and the route with low frequency may have abnormal driving behavior e.g. detour of the driver [36], [37].

In this paper, OD frequency analysis of 15.36 million taxi order data of Nantong, China in 2018 is carried out. Figure 1 demonstrates how the space-time law of order trajectory with low frequency is applied in this analytical process

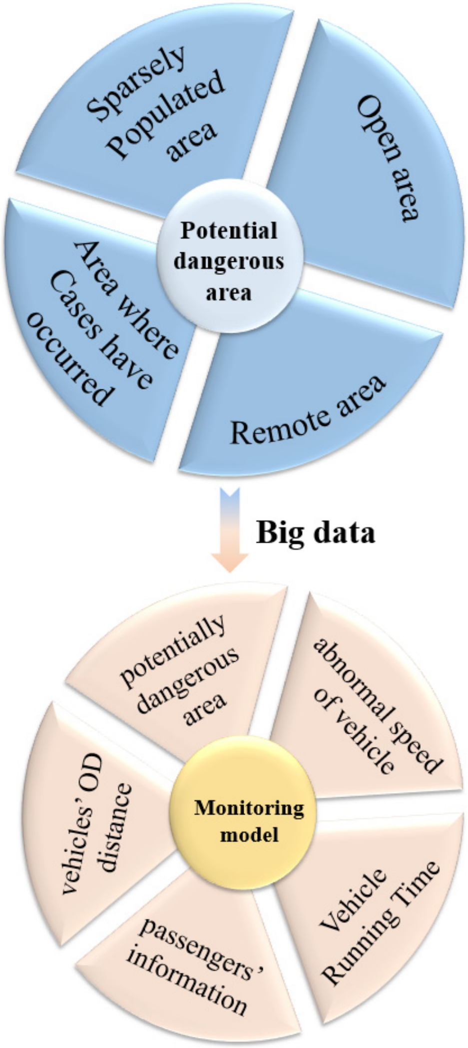


FIGURE 2. Modeling structural diagram.

of origin-destination frequency. Through analysis, the time and distance factors of this kind of orders are higher than those of normal driving orders and have certain regularity. Considering the particularity of online car-hailing cases, such as parking crime, this paper takes vehicle OD distance, abnormal speed and driving time as model factors.

B. CONSTRUCTION OF MODEL

A multi-parameter safety state monitoring model is constructed based on a series of factors. The current driving safety can be judged by calculating the real-time risk score P_{total} of the vehicle:

$$P_{total} = W * D_i * \sum_{i=1}^{L} \theta_i * P_i \tag{1}$$

In the formula, D_i is the risk index of the potential danger zone, P_i is the risk score of each factor, W is the passenger safety factor, P_{total} is the total risk score, and θ_i is the weight of P_i. The structure of the model is shown in Figure 2.

III. INDICATORS OF MODEL

In this model, the travel trajectory and passenger information of vehicles in a potentially dangerous zone are analyzed, and the corresponding risk score is given. The safety condition of passengers in vehicles is monitored according to the total risk score P_{total}. The detailed analytical steps are as follows.

A. POTENTIALLY DANGEROUS ZONE

The steps for dividing the potentially dangerous zone are as follows:

1) According to the safety conditions, time period of the city is divided into high-incidence period T_1, potential danger period T_2, and safety period T_3.
2) Based on the POI data as well as the pick-up and drop-off points of urban vehicles at different time periods, make low-thermal regions are classified and established, from which dense areas are assigned with the risk index D_1.
3) Extract the urban non-built areas and assign the risk index D_2.
4) Subject the result data of step 2 and step 3 to the overlay operation. The overlay result is taken as the remote region *Sp1*, and the danger index of the corresponding area is superimposed.
5) Extracting the construction site and taking the forest area as the open zone *Sp2*, and assigning it the risk index D_3.
6) For areas where relevant cases have occurred or have been complained about, the point set is used to describe *Pi*, and a buffer zone with a radius of r is built around it, which is the high-incidence zone *Sp3*, and the hazard index *D4* is assigned.
7) Combine the risk index of *Sp1*, *Sp2,* and *Sp3* and superimpose the risk index accordingly. The area where the risk index is not 0 is the potential danger zone.
8) The risk degree of potential danger zone will change in different time periods. In the high incidence period of the case, the risk index will be increased by I_1, while the risk index will be increased by I_2 in the potential danger zone. The risk index will remain unchanged in the safety period. The formula is as follows:

$$D_i|T_j = I_j + D_i \tag{2}$$

In which, $D_i|T_j$ means that the risk index of the zone i in the period of T_j isD_i, where j is 1 or 2.

"Dangerous degree index", as shown in Table 1, can be assigned to the corresponding index according to the regional characteristics [24], in order to differentiate degrees of dangerous zones in the remote areas, open zones and high-incidence zones.

B. OD DISTANCE OF VEHICLES

For a long time, there have been widespread cases in which passengers are detoured by some drivers and charged more. Especially when driving in unfamiliar areas, where passengers may not be able to recognize such anomalous behavior.

TABLE 1. Example of potential danger zone scoring.

Type	Element	Subdivision of risk index（1-20）
Remote zone	Non-built-up area	3
	low-vitality area	5
Open zone	Mountain and forest area	7
	Construction areas	13
High-incidence zone	The area where have been complained	19
	The area where cases have occurred	20

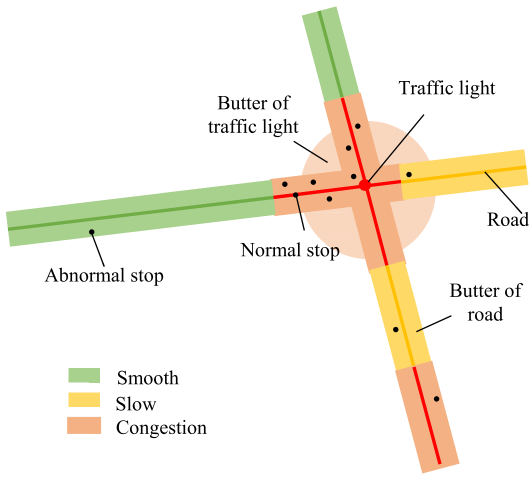


FIGURE 3. Road congestion diagram.

For this phenomenon, the real-time vehicle trajectory can be studied to analyze the current route and compare it with the route with high OD frequency [22], [23], so as to judge whether the driver has detour behavior.

The above is carried out by analyzing the potentially dangerous zones in conjunction with the characteristics of specific cases.

When the vehicle drives in a dangerous area, the distance between the current position and the destination is not reduced but is larger than the distance between the previous track point and the destination. Moreover, when the distance difference exceeds a certain threshold value, the danger assessment is carried out, and the specific method is as follows:

1) Initialize the danger score P_1 of the current vehicle order as 0 and the flag *dis_flag* in the "distance increasing state" as 0.
2) When the vehicle drives into a potential danger zone, calculate the real-time distance $S(Pos_{\mathrm{i}}^{(\mathrm{n})}, Des^{(\mathrm{n})})$ between the vehicle's current location $Pos_{\mathrm{i}}^{(\mathrm{n})}$ and the destination $Des^{(\mathrm{n})}$:

$$\mathrm{S}\left(Pos_i^{(n)}, Des^{(n)}\right) = \sqrt{(Des_x^{(n)} - Pos_{ix}^{(n)})^2 + (Des_y^{(n)} - Pos_{iy}^{(n)})^2} \quad (3)$$

In the formula, n is the number of the order, i is the trajectory point number of order n.

3) Judge whether S' is greater than 0. If greater than 0, set *dis_flag* as 1, and judge distance difference ΔS; when $\Delta \mathrm{S}$ exceeds the threshold value X, the risk score P_1 is updated to P_1+*1* for every n meters. Among them, S' and ΔS are calculated as follows:

$$S' = \frac{\mathrm{d}\mathrm{S}(Pos_i^{(n)}, Des^{(n)})}{dt} \quad (4)$$

$$S = S(Pos_i^{(n)}, Des^{(n)}) - S(Pos_{i-1}^{(n)}, Des^{(n)}) \quad (5)$$

In the formulas, when S' is greater than 0, S(Posi(n), Des(n)) will increase.

C. ABNORMAL SPEED OF VEHICLES

The speed information in the GPS data can directly reflect the current state of the vehicle. Through the speed variation and speed magnitude, the driving state of the vehicle can be directly identified. When the vehicle speed is high or the speed variation is large, the driver will have abnormal driving behavior [21], [23]. In the past cases of online car-hailing drivers infringing on passengers, drivers would choose to stop to commit crimes [17]. At this time, GPS would record the current vehicle speed state, so it is crucial to judge abnormal parking behavior in the order process.

If the parking time is not too long, it may be because of waiting for the red light or traffic jam. If it exceeds a certain threshold, the behavior can be judged as abnormal behavior, and the risk assessment is carried out. The specific methods are as follows:

1) Initialize the risk score P_2 of the current order as 0 and the sign *stop_flag* which is in "abnormal parking state" as 0.
2) When a vehicle is driving in a dangerous area, if the speed is 0, the real-time traffic congestion and traffic light waiting condition are analyzed to judge whether the vehicle is in a congested or red light state, that means, then the vehicle data either falls into the congested section or in the traffic light buffer zone, *stop_flag* will be 1.
3) When the *stop_flag* is 1, and the parking time exceeds the threshold Y, the risk score P_2 is updated to "P_2+1" for every n minutes.

In practical applications, considering that the driver tends to change shifts and take short breaks when they are in "passenger-carrying mode", the order behavior of "stop state" may result in the excessive risk score. To correct this, the data source will be processed. When the number of passengers is 0 (at which state the APP does not transfer passenger data to the server), the current order will not be counted.

D. DRIVING TIME OF VEHICLES

Abnormal driving behaviors of drivers are not only manifested in OD path but also can be perceived in time.

TABLE 2. Partially processed trajectory data record.

License plate number	Time	Speed	Lon	Lat	Order	Type
Su FB****	2018-12-23 02:11:46	0	120.8499	32.01197	1	Pick-up point
Su FB****	2018-12-23 02:12:51	7.6	120.8498	32.01182	1	
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Su FB****	2018-12-23 02:18:44	24.9	120.8685	32.01253	1	
Su FB****	2018-12-23 02:19:10	18.3	120.8684	32.01493	1	Drop-off point

Combining the current position and time of vehicle trajectory, space-time analysis is conducted, and the abnormal behaviors of vehicles can be identified by comparing the travel time of actual path and the travel time of shortest path [25].

In the model, when the vehicle has been driving for longer than the estimated OD time, and the time difference exceeds a certain threshold, the behavior is judged to be a potential abnormal behavior, and the risk is assessed. The specific method is as follows:

1) Initialize the risk score P_3 of the current order as 0 and the sign *time_flag* which is in "abnormal time state" as 0.
2) Calculate the estimated OD time T_p through Route Matrix API provided by Baidu Map.
3) Calculate the driving time T_e and ΔT. When ΔT is greater than 0, *time_flag* is 1. T_e is calculated as follows:
$$T_e = time_e - time_o \tag{6}$$
In the formula, $time_e$ is the current time, $time_o$ is initial departure time.
4) When *time_flag* is 1, calculate ΔT. If ΔT is greater than the threshold T_s, the score P_3 is updated to P_3+1 for every m seconds. Where ΔT is calculated as follows:
$$\Delta T = T_e - T_p \tag{7}$$
In which, T_p is the estimated driving time.

E. INFORMATION OF THE PASSENGERS

In the past, most of the victims of online car-hailing cases have been associated with female passengers, and when there were few passengers in the car [17]. There was a strong correlation between the gender and number of passengers and the probability of the crime case [24]. Therefore, it is very important to analyze the case based on the information of passengers. In this model, by analyzing the gender and number of passengers, the appropriate risk assessment coefficient is given. The specific methods are as follows:

1) Initialize the passenger risk coefficient Wc as 0 and the sign of "gender of passenger" *sex_flag* as 0.
2) Use the self-developed mobile APP to get the number of passengers *count* ($count \geq 1$) and gender. When all the passengers are female, *sex_flag* is 1 and W_c is 1.
3) When *sex_flag* is 1, the risk coefficient of each additional passenger is reduced by wc_i ($1 \leq i \leq 3$), and W_c is W_c-wc_i.

The above factors reflect the real-time comprehensive situation of the order, and the detection efficiency can be improved by adjusting the parameters to meet the regulatory needs and urban characteristics during use.

IV. DATA AND PROCESSING

The Gangzha District of Nantong city has a flat terrain, and it is adjacent to Shanghai in space, thus it an excellent geographical location with idea traffic conditions for modeling experiments. The functional divisions in this area are mostly a rural-urban fringe area. Many "potential danger zones" proposed in this paper can be found in this area, making it highly useful and appropriate for experiments.

A. VEHICLE TRAJECTORY DATA

The experimental data were the trajectory data obtained from the Beidou /GPS Monitoring Service Center of Jiangsu Pacific Communication Technology Co., Ltd, based on the taxi operation data of Nantong from 23rd December to 24th December 2018, information of these data include license plate number, track time, coordinates, direction, instantaneous speed, operation status, and other information.

However, the above obtained raw taxi trajectory data contained some abnormal conditions, such as time crossing, positioning error, data format error. Thus, data preprocessing was carried out such as by eliminating error information, correcting position deviation, and screening all order data in vehicle data. The partially processed trajectory data are shown in Table 2, in which *Lon* is longitude, *Lat* is latitude, the *order* is the order number and *type* is the type of trajectory point.

B. POI DATA

The POI data was downloaded from the Baidu Map API by the web crawler technology, which includes 1052 records with property fields such as residence, school, food, hospital, scenic spot, and community, and other keywords such as coordinates, and names et al, Table 3 shows different parts recorded for of Nantong City.

C. ROAD NETWORK DATA AND ITS CONGESTION CONDITION

The experimental data was the OSM open-source road network first-level road data [34], whose attributes include road name, geometric type and other characteristics. The real-time road condition query interface provided by the web crawler and Baidu Map and the Ali Cloud server are combined

TABLE 3. Partially processed trajectory data record.

X	Y	Name	Type
120.802885	32.094889	Xirui Hotel	Food
120.846762	32.096578	Happy Middle School	School
120.812354	32.108432	Wuli New Village	Resident
120.816256	32.065897	The Second People's Hospital	Hospital
120.864820	32.046191	Nantong Adventure Kingdom	Scenic spot
120.844770	32.051645	Wanhao Villa	Community

TABLE 4. Partially processed congestion data record.

Time	Road_name	Status	Status_desc
2018/12/04 15:30	Gongnong Road	2	Slow
2018/12/04 16:00	Gongnong Road	1	Unblocked
2018/12/04 16:30	Gongnong Road	1	Unblocked
2018/12/04 17:00	Gongnong Road	1	Unblocked
2018/12/04 17:30	Gongnong Road	1	Congested

to obtain the real-time road condition from 0 o 'clock on the 1st of December 2018 to 0 o 'clock on the 1st of January, 2019 [35]. As is referred to by the Baidu Map standard, the status of 1 stand for unblocked, which means that the vehicles can run fast; 2 stands for slow, that means the vehicles slow down but do not stop; 3 stands for congestion, which means vehicles are congested. 4 represents serious congestion, in other words, the vehicles are difficult to drive. Because of the delay in the HTTP request and data return, the time in congestion data was not in full hours, so it needs to be corrected. The data is processed in batches, and Table 4 is the road congestion status data after time processing.

D. REMOTE SENSING IMAGE DATA

The remote sensing image data of Lujia-01 night-light data of the 29th of October, 2018 was downloaded from the Hubei Data and Application Network of the High-Resolution Earth Observation System. The Landsat-8 OLI image data was the image of December 2017 downloaded from the Geospatial Data Cloud. The original data contained 12 bands, with the cloud volume of 0.46% and the spatial resolution of 30 meters, and Figure 4 shows the sample data within the study area.

V. RESULT AND DISCUSSION

A. PARAMETERS SETTING

The vehicle trajectory data and the POI data respectively describe the dynamic and static urban vitality characteristics. In order to distinguish the vitality of different areas of the city, the nuclear density analysis was carried out on pick-up and drop-off point and POI data [25], [26], from which high and low-density areas were identified, and sparsely populated areas were extracted.

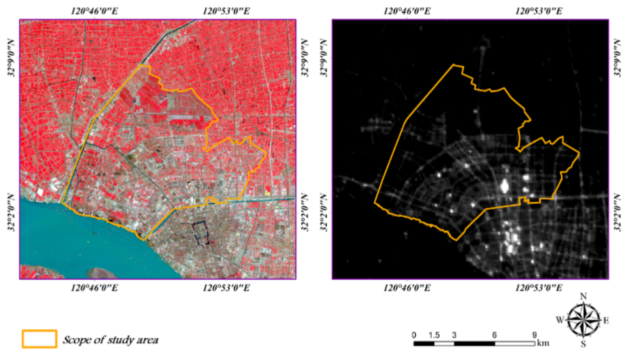


FIGURE 4. Remote sensing image data of the study area (Left is the standard false-color composite images (made up of information from bands 3–5) of Landsat8 OLI, and right is the night light image of the Loujia1-01).

In recent years, many scholars have extracted urban built-up areas by means of night-light remote sensing data and conducted studies on urban expansion, GDP development level [27]. Based on the hypothesis that greater light intensity, results in higher possibility of urban built-up of the area, the present paper uses a bisection method to set the threshold which can be continuously adjusted by comparing with the statistical data, and to extract the non-built area [28] and set as a remote zone in the city. Through the digitization of Landsat-8 OLI image, extract mountain forests and building land were also extracted as open areas in the city. The areas extracted above are assigned the risk index and merged. Then, by superimposing the risk index of the corresponding location, the potential danger zone in T_1, T_2 and T_3 periods (in this paper, T_1 is 0:00∼6:00, $T2$ is 15:01∼24:00, $T3$ is 6:01∼15:00) were extracted (Figure 5).

According to the parameter selection of OD distance, speed anomaly and driving time, we selected the orders that may have abnormal behavior (driver detour and stop time is too long) as the experimental data, and compared the effect of different parameters on selected factors among different groups. These orders involving the passenger complaint were extracted from the Nantong taxi management system.

Figure 6 illustrates how orders with possible detour and abnormal stop behavior were displayed. These orders may correspond with the possible detour and abnormal stop behaviors, and then, we can easily find the strange graph after its trajectory data been displayed on the map.

From these data, risk scores of each factor were calculated for different parameters, and the results are shown in Figure 7.

For orders with abnormal behavior, the calculated results of the risk score of each factor in the model better reflected the real-time state of vehicles. To more accurately assess the real-time safety state of passengers, on the premise that the danger integral value is moderate and the real-time trajectory status can be accurately reflected. Taking the P_1 risk score as an example, as shown in Fig. 7(a), when x is one-fifth of the *OD distance* (OD_DIS), abnormal fluctuations in behavior can be captured more easily. Under this condition, n_1 is selected as

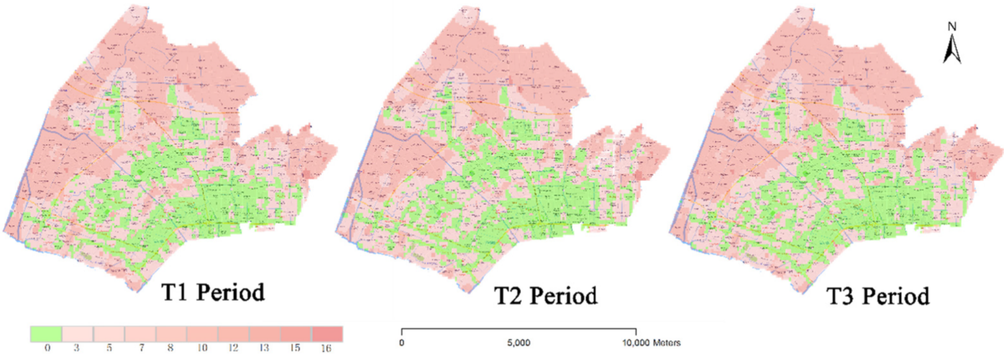
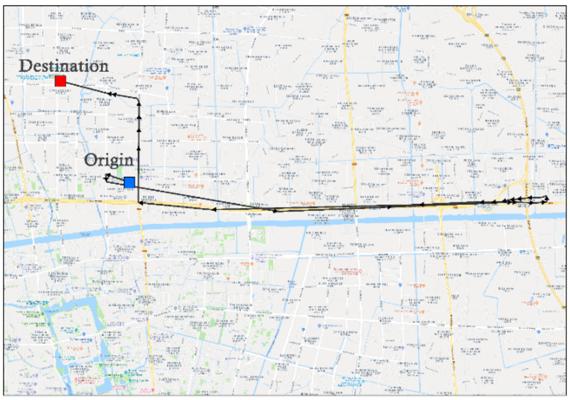


FIGURE 5. Potential danger zone map in each time period.



(a)

(b)

FIGURE 6. Order trajectory with abnormal behavior (a) detour behavior (b) abnormal stop behaviour.

TABLE 5. Optimal parameters of safety analysis.

Factors	Parameter	The optimal value
OD distance	X	OD_DIS/15
	n1	X/6
Abnormal speed	Y	60
	n2	9
Driving time	Ts	Tp*2
	n3	Tp/10

TABLE 6. Parts records of score.

License plate number	Time	Type	P_{total}
Su FB****	2018-12-23 02:02:48	Pick-up point	0
Su FB****	2018-12-23 02:05:03		0
Su FB****	2018-12-23 02:05:19		0
Su FB****	2018-12-23 02:05:49		0
Su FB****	2018-12-23 02:06:19		0
Su FB****	2018-12-23 02:06:50		0.117
Su FB****	2018-12-23 02:07:20		5.512
Su FB****	2018-12-23 02:07:49		7.956
Su FB****	2018-12-23 02:08:20	Drop-off point	9.945

one-sixth of x, then the result is more reasonable. Because the simulation curve of P_2 and P_3 risk score is similar to P_1, so the parameter selection is similar as above. The optimal parameters were selected and listed in Table 5, in which *OD_DIS* is the required distance for OD.

We first extracted all the order data in the Gangzha district and took different orders as experimental objects to select the optimal parameters. Then, used the model script to perform a batch processing to obtain real-time risk score of the vehicle. The Analytic Hierarchy Process (AHP) was used to calculate the weight, and the results were 0.3, 0.44 and 0.26 respectively, the weights were used to calculate the comprehensive risk score P_{total}. When the trajectory points were in dangerous areas, the potential danger index of the current trajectory area was assigned. Since the App is still in the developmental stage, t the total risk score P_{total} was calculated based on the case that the passenger is a single woman (*sex_flag* is 1) in the present study. Some results are shown in Table 6.

B. ANALYSIS AND DISCUSSION

Analysis of the study results showed that there were more orders in the T_2 period, and the risk scores in this period were generally higher than those in periods T_1 and T_3. The risk score of certain individual orders in the T2 period could be dozens of times higher than other orders in other time periods. By tracking orders with higher risk scores, it was found that the P_1 and P_3 points were high in most moments of this particular high-risk order. The distance and time of the order in the model were found to be abnormal thus the vehicle was judged as having a detour behavior [38]. The same vehicle was found at some time travel in potentially dangerous areas, which ultimately leads to a higher calculated total score in the model (Figure 7).In addition, the risk score during the holiday was found to be generally lower than that of the working days. Due to the increased number of travelers, the number of orders also increased. As a result, the probability of illegal behavior of the driver was actually decreased and the safety of the passengers was significantly improved. This suggests that the monitoring model proposed in the present study to be reliable.

By comparing the risk scores of the orders on the 23rd of December and the 24th of December, it was found that latter were generally lower, and the number of orders in the T1 on the 24th of December was about twice as that of the same period on the 23rd of December. The number of orders increased as a direct result of being near Christmas Eve. Such increase led to a reduced probability of illegal behavior of the drivers, significantly improved the safety of driving, which proved again the reliability of the proposed detection model.

By analyzing the safety situation, it was found that the risk score P_{total} was generally higher in the T_2 on the 23rd and 24th. Further analyses of the trajectory of the order and the P_1 and P_3 scores of the model, it could be concluded that detours in individual vehicles might have occurred. In all periods, when P_{total} of the vehicle is not zero, the P_2 score of most vehicles is low, meaning that the number of abnormal stops of vehicles is less. Because of the larger weight of P_2 in the model, the order is deemed less risky. In the present study, interviews were conducted with drivers with a high P_{total} and a high P_2. It was found that when the red light time was too long, the score tends to be higher. The present model could, therefore, change parameters accordingly. In addition, by comparing data between holidays and workdays, it was concluded that it would be best for female passengers to choose to travel when the traffic volume is large and to avoid choosing the time when the number of passengers is small, so as to improve the safety of driving.

Since there are no examples of passengers being infringed in the experimental area, the "potential danger zone" could not be established by referencing the real crime locations, and the abnormal behaviors of vehicles will be evaluated by comprehensive use of the other factors. The orders were sorted in a descending order based on the total value of P_{total}, and the top 0.5%, which is 63, of the total 12600 orders were selected.

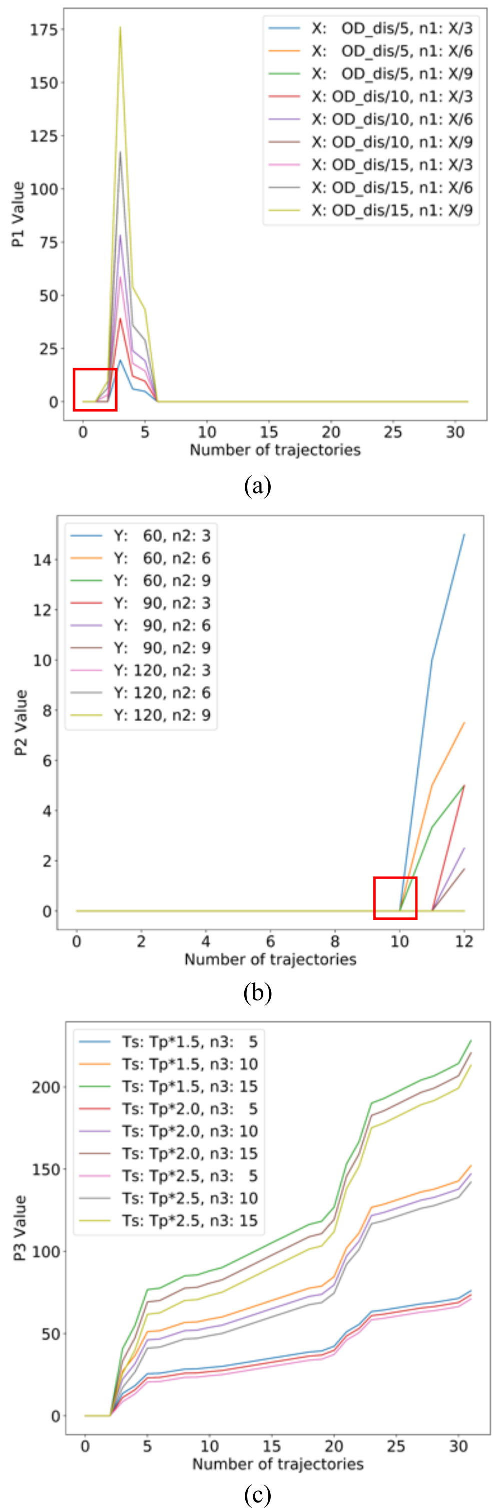


FIGURE 7. **Comparison of various factors' risk scores (a) P_1 risk score (b) P_2 risk score (c) P_3 risk score.**

The trajectory data of the selected orders were identified using the taxi management system and OD frequency analysis method based on big data analysis. As a result, 58 orders were identified to have abnormal behaviors, such as detour, abnormal stop, and the experimental accuracy reached 92.06%,

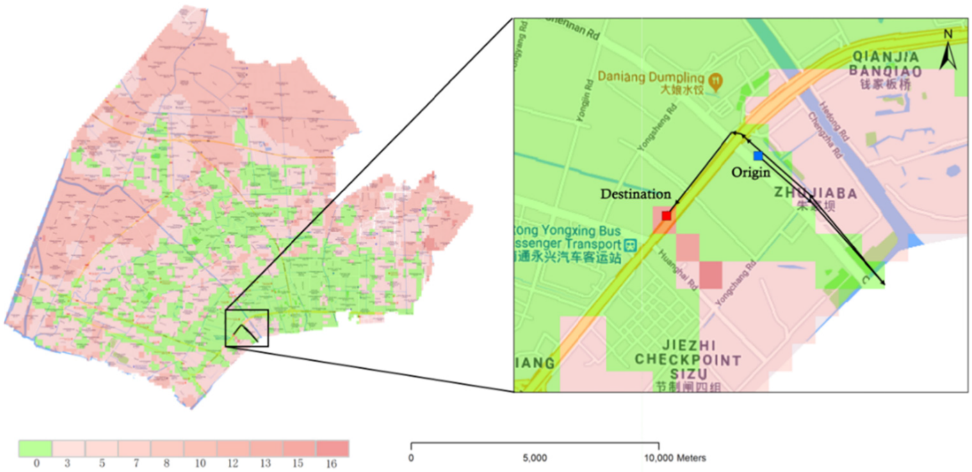


FIGURE 8. Maximum score order trajectories in T2.

TABLE 7. Result of model checking.

license plate number	The start time of order	The end time of order	The maximum value of P_{total}	Type of behavior
Su FB****	2018-12-02 16:16:33	2018-12-02 16:22:05	43.794	normal
Su FB****	2018-12-03 15:13:35	2018-12-03 15:27:14	137.482	detour
Su FB****	2018-12-05 15:59:57	2018-12-05 16:17:20	89.685	detour
Su FB****	2018-12-09 07:11:59	2018-12-09 07:26:13	104.843	detour
⋮	⋮	⋮	⋮	⋮
Su FB****	2018-12-23 17:00:18	2018-12-23 17:37:34	91.702	abnormal stop
Su FB****	2018-12-23 07:50:30	2018-12-23 08:03:30	228.634	detour

among which the discrimination accuracy of the abnormal stop was 100% and that of the detour was 90.57% Partial results are shown in Table 7. Among them, P_2 of most vehicles was lower, which suggest that the number of abnormal stops of vehicles was less, although some orders remained in higher scores. Some of this be due to the abnormally long red light time, therefore by interviewing drivers with higher P_{total} and higher P_2 orders, the proper adjustment was performed for the model parameters to better reflect the characteristics of urban traffic.

Due to the lack of considering the actual situation of urban traffic, such as road congestion and road construction status, the estimated OD time of the current model was calculated based on Baidu Map API under ideal conditions. As a result, some calculated results may be obviously less than the actual required time, which leads to certain errors in the model results.

VI. CONCLUSION

In summary, the present study examined the safety monitoring of ‘‘online car-hailing’’ passengers from the perspectives of geographical time and space, and proposed a multi-factor safety state monitoring model based on ‘‘potentially dangerous zone’’. By analyzing different factors and their combined effects, potentially dangerous zones in cities were extracted. Model factors were selected based on big data analysis, and better parameters were selected based on multi-group experimental analysis. The real-time trajectory behavior of vehicle orders was analyzed by using space-time behavior. The experimental results showed that the model has a better monitoring effect than exist APP. However, in practical application, due to the location accuracy of vehicle GPS trajectory points and the calculation accuracy of the Baidu Map API, there is still some inaccuracy in this proposed monitoring model. In the future work, we will focus more on solving the impact of data and will take into consideration the actual complex traffic conditions to reduce the scoring error.

ACKNOWLEDGMENT

The authors would like to thank the editor and the anonymous reviewers who provided insightful comments on improving this article and thank The Hong Kong Polytechnic University for providing the opportunity of academic exchange. We also thank Mr. Rong Tsao, Research Scientist of Agriculture and Agriculture-Food Canada, for his valuable advices on the improvement of English writing.

REFERENCES

[1] Q. Sun, Y. He, Y. Wang, and F. Ma, ‘‘Evolutionary game between government and ride-hailing platform: Evidence from China,’’ *Discrete Dyn. Nature Soc.*, vol. 2019, Jan. 2019, Art. no. 9545102.

[2] J. Hamari, M. Sjöklint, and A. Ukkonen, ‘‘The sharing economy: Why people participate in collaborative consumption,’’ *J. Assoc. Inf. Sci. Technol.*, vol. 67, no. 9, pp. 2047–2059, 2016.

[3] K. Li, ‘‘Investigating the effect crime has on Uber and Yellow Taxi pickups in NYC,’’ M.S. thesis, Dept. Econ., Colby College, Waterville, ME, USA, 2019.

[4] A. A. Damaini, G. S. Nugroho, and S. Suyoto, ‘‘Fraud crime mitigation of mobile application users for online transportation,’’ *Int. J. Interact. Mobile Technol.*, vol. 12, no. 3, pp. 153–167, 2018.

[5] H. Jiang and X. Zhang, ‘‘An experimental model of regulating the sharing economy in China: The case of online car hailing,’’ *Comput. Law Secur. Rev.*, vol. 35, no. 2, pp. 145–156, 2019.

[6] Y. Wu, "On legal supervision of car-hailing services," in *Proc. 5th Int. Conf. Humanities Social Sci. Res. (ICHSSR)*. Guilin, China: Atlantis Press, 2019, pp. 1–4.

[7] R. Almeida, R. Oliveira, M. Luís, C. Senna, and S. Sargento, "A multi-technology communication platform for urban mobile sensing," *Sensors*, vol. 18, no. 4, p. 1184, 2018.

[8] Z. Wang, S. Dang, S. Shaham, Z. Zhang, and Z. Lv, "Basic research methodology in wireless communications: The first course for research-based graduate students," *IEEE Access*, vol. 7, pp. 86678–86696, 2019.

[9] Q. Liu, T. Ozcelebi, L. Cheng, F. Kuipers, and J. Lukkien, "CluFlow: Cluster-based flow management in software-defined wireless sensor networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2019, pp. 1–8.

[10] X. Li and Z. Zhang, "Research on local government governance and enterprise social responsibility behaviors under the perspective of cournot duopoly competition: Analyzing taxi companies and online car-hailing service companies," *Math. Problems Eng.*, vol. 2018, Aug. 2018, Art. no. 5794232.

[11] X. Kong, M. Li, K. Ma, K. Tian, M. Wang, Z. Ning, and F. Xia, "Big trajectory data: A survey of applications and services," *IEEE Access*, vol. 6, pp. 58295–58306, 2018.

[12] C. Ding, "Application of GIS technology in the construction of urban traffic sharing multimedia information platform," in *Multimedia Tools and Applications*. New York, NY, USA: Springer, 2019, pp. 1–13. doi: 10.1007/s11042-019-7500-0.

[13] Y. Gao and J. Chen, "The risk reduction and sustainable development of shared transportation: The Chinese online car-hailing policy evaluation in the digitalization era," *Sustainability*, vol. 11, no. 9, p. 2596, 2019.

[14] M. Pak, S. Gülci, and A. Okumuş, "A study on the use and modeling of geographical information system for combating forest crimes: An assessment of crimes in the eastern Mediterranean forests," *Environ. Monit. Assess.*, vol. 190, no. 2, p. 62, 2018.

[15] C. Catlett, E. Cesario, D. Talia, and A. Vinci, "Spatio-temporal crime predictions in smart cities: A data-driven approach and experiments," *Pervas. Mobile Comput.*, vol. 53, pp. 62–74, Feb. 2019.

[16] M. P. J. Ashby and K. J. Bowers, "A comparison of methods for temporal analysis of aoristic crime," *Crime Sci.*, vol. 2, no. 1, p. 1, 2013.

[17] M. Boldt and A. Borg, "Evaluating temporal analysis methods using residential burglary data," *ISPRS Int. J. Geo-Inf.*, vol. 5, no. 9, p. 148, 2016.

[18] S. Ko, *Didi Chuxing: Expansion and Risk Management*. London, U.K.: SAGE Publications, 2019.

[19] A. Keler and J. D. Mazimpaka, "Safety-aware routing for motorised tourists based on open data and VGI," *J. Location Based Services*, vol. 10, no. 1, pp. 64–77, 2016.

[20] D. D. Ngoc, H. Loisel, C. Jamet, V. Vantrepotte, L. Duforêt-Gaurier, C. D. Minh, and A. Mangin, "Coastal and inland water pixels extraction algorithm (WiPE) from spectral shape analysis and HSV transformation applied to Landsat 8 OLI and Sentinel-2 MSI," *Remote Sens. Environ.*, vol. 223, pp. 208–228, Mar. 2019.

[21] S. Hasan, W. Shi, X. Zhu, and S. Abbas, "Monitoring of land use/land cover and socioeconomic changes in south china over the last three decades using landsat and nighttime light data," *Remote Sens.*, vol. 11, no. 14, p. 1658, 2019.

[22] X. Li, X. Li, D. Li, X. He, and M. Jendryke, "A preliminary investigation of Luojia-1 night-time light imagery," *Remote Sens. Lett.*, vol. 10, no. 6, pp. 526–535, 2019.

[23] Z. Huang, Y. Zhang, Q. Li, T. Zhang, N. Sang, and H. Hong, "Progressive dual-domain filter for enhancing and denoising optical remote-sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 759–763, May 2018.

[24] Y. Yue, Y. Zhuang, A. G. O. Yeh, J.-Y. Xie, C.-L. Ma, and Q.-Q. Li, "Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy," *Int. J. Geograph. Inf. Syst.*, vol. 31, no. 4, pp. 658–675, 2017.

[25] T. Jia and Z. Ji, "Understanding the functionality of human activity hotspots from their scaling pattern using trajectory data," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 11, p. 341, 2017.

[26] S. An, H. Yang, and J. Wang, "Revealing recurrent urban congestion evolution patterns with taxi trajectories," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 4, p. 128, 2018.

[27] S. Sun, J. Chen, and J. Sun, "Traffic congestion prediction based on GPS trajectory data," *Int. J. Distrib. Sensor Netw.*, vol. 15, no. 5, 2019. doi: 10.1177/1550147719847440.

[28] M. Lu, J. Liang, Z. Wang, and X. Yuan, "Exploring OD patterns of interested region based on taxi trajectories," *J. Vis.*, vol. 19, no. 4, pp. 811–821, 2016.

[29] L. Cai, F. Jiang, W. Zhou, and K. Li, "Design and application of an attractiveness index for urban hotspots based on GPS trajectory data," *IEEE Access*, vol. 6, pp. 55976–55985, 2018.

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

[31] Y. Wang, K. Qin, Y. Chen, and P. Zhao, "Detecting anomalous trajectories and behavior patterns using hierarchical clustering from taxi GPS data," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 1, p. 25, 2018.

[32] J. Mao, W. Tao, C. Jin, and A. Zhou, "Feature grouping-based outlier detection upon streaming trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 12, pp. 2696–2709, Dec. 2017.

[33] Y. Djenouri, A. Belhadi, J. C.-W. Lin, D. Djenouri, and A. Cano, "A survey on urban traffic anomalies detection algorithms," *IEEE Access*, vol. 7, pp. 12192–12205, 2019.

[34] T. Novack, Z. Wang, and A. Zipf, "A system for generating customized pleasant pedestrian routes based on OpenStreetMap data," *Sensors*, vol. 18, no. 11, p. 3794, 2018.

[35] G. Gao, Z. Wang, X. Liu, Q. Li, W. Wang, and J. Zhang, "Travel behavior analysis using 2016 Qingdao's household traffic surveys and Baidu electric map API data," *J. Adv. Transp.*, vol. 2019, Mar. 2019, Art. no. 6383097.

[36] Z. Zhou, W. Dou, G. Jia, C. Hu, X. Xu, X. Wu, and J. Pan, "A method for real-time trajectory monitoring to improve taxi service using GPS big data," *Inf. Manage.*, vol. 53, no. 8, pp. 964–977, 2016.

[37] C. Chen, D. Zhang, P. S. Castro, N. Li, L. Sun, S. Li, and Z. Wang, "iBOAT: Isolation-based online anomalous trajectory detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 806–818, Jun. 2013.

[38] S. Liu, L. M. Ni, and R. Krishnan, "Fraud detection from taxis' driving behaviors," *IEEE Trans. Veh. Technol.*, vol. 63, no. 1, pp. 464–472, Jan. 2014.

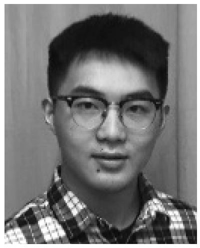
TONG ZHOU received the bachelor's degree from Jiangsu Normal University, in 2000, and the M.S. degree from Nanjing Normal University, in 2007, where he is currently pursuing the Ph.D. degree. He is also an Associate Professor with the School of Geographical Sciences and the Professional Head of geographic information science with Nantong University. He is also a Visiting Scholar of land surveying and geo-informatics with The Hong Kong Polytechnic University. His research interests include geographic information science, urban informatics, human mobility, spatiotemporal data analysis and mining, remote sensing, and GIS algorithm.

WENZHONG SHI received the Ph.D. degree from the University of Osnabrück, Vechta, Germany, in 1994. He is currently the Head of the Department of Land Surveying and Geo-Informatics, a Otto Poon Charitable Foundation Professor in urban informatics, a Chair Professor in GISci and remote sensing, and the Director of the Laboratory for Smart City and Spatial Big Data Analytics, The Hong Kong Polytechnic University. He has published more than 400 scientific articles (with more than 200 SCI articles) and 15 books. His current research interests include the areas of urban informatics for smart cities, GISci, and remote sensing with focusing on analytics and quality control for spatial big data, object extraction, and change detection from satellite images and LiDAR data, integrated mobile mapping technology, and 3D and dynamic GISci modeling. He also serves as an Editorial Board Member for a number of international journals. He received a number of prestige awards, including an award from the International Society of Photogrammetry and Remote Sensing and Natural Science Award from the State Council, China. He served as the President for the Commission II for International Society for Photogrammetry and Remote Sensing, from 2008 to 2012 and the President for the Hong Kong Geographic Information System Association, from 2001 to 2003.

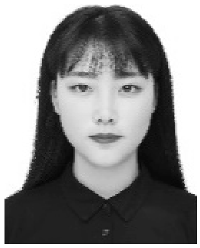
XINTAO LIU received the B.Eng. degree in survey from Hohai University, China, in 1998, the M.Sc. degree in cartography and GIS from Nanjing Normal University, China, in 2003, and the Ph.D. degree in geo-informatics from the Royal Institute of Technology, Sweden, in 2012. He received the Ph.D. Scholarship from Lars Erik Lundebergs. In 2012, he joined the Department of Civil Engineering, Ryerson University, Canada, where he was a Postdoctoral Fellow in GIS and transportation until 2016. He has been a Sessional Lecturer with Ryerson University, since 2015. He is currently an Assistant Professor with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University. He is also a PI and a Co-PI of several national projects funded by Sweden, Canada, and Hong Kong. His research interests include GI services and science, urban computing, and GIS in transportation, especially the state-of-the-art technologies to advance smart city for a better urban life.

FEI TAO received the B.S. degree in geography teacher and the M.S. degree in marine geography from Nanjing Normal University, Nanjing, China, in 2004 and 2007, respectively, where she is currently pursuing the Ph.D. degree.

From 2007 to 2019, she was a Lecturer with the School of Geographical Sciences, Nantong University, Jiangsu, China. She has presided over a project supported by the Young Scientists Fund of the National Natural Science Foundation of China and published more than ten articles. Her research interests include remote sensing applications and geographic big data analysis.



ZHEN QIAN was born in Zhenjiang, Jiangsu, in 1997. He entered in geographic information science with the School of Geographical Sciences, Nantong University, Jiangsu, in 2016. He is mainly engaged in spatiotemporal data mining research, published six articles on spatiotemporal data mining, applied for four national patents of China, and two software copyrights. He received many national awards and scholarships, including the M Prize of the American Mathematical Modeling Competition and the National Scholarship.



RUIJIA ZHANG was born in Lianyungang, Jiangsu, China, in 1997. She received the bachelor's degree from Nantong University, in 2019. She has been admitted to Nanjing Agricultural University as a Graduate Student with the major in English translation. From 2018 to 2019, she was a member of the Intelligent Analytical Teacher-Student Team for Geographic Big Data, Nantong University, Jiangsu. Her research interest includes the development of Chinese and Western translation theories and geographic information technology. Her awards include the Second Prize in National English Competition for College Students and Preliminary Round of the 2018 ''FLTRP Cup'' English Writing Contest, and the Meritorious Winner Award in 2018 Mathematical Contest in Modeling, which she was a Translator.