

## Research Article

# Applying Big Data Analytics to Monitor Tourist Flow for the Scenic Area Operation Management

Siyang Qin <sup>1</sup>, Jie Man <sup>2</sup>, Xuzhao Wang,<sup>2</sup> Can Li,<sup>2</sup> Honghui Dong,<sup>2</sup> and Xinquan Ge<sup>1,3</sup>

<sup>1</sup>School of Economics and Management, Beijing Jiaotong University, 3 ShangyuanCun, Haidian District, Beijing 100044, China

<sup>2</sup>School of Traffic and Transportation, Beijing Jiaotong University, 3 ShangyuanCun, Haidian District, Beijing 100044, China

<sup>3</sup>School of Economics and Management, Beijing Information Science and Technology University, 12 Qinghe Xiaoying East Road, Haidian District, Beijing 100192, China

Correspondence should be addressed to Siyang Qin; syqin@bjtu.edu.cn and Jie Man; 16120770@bjtu.edu.cn

Received 7 September 2018; Revised 11 November 2018; Accepted 18 November 2018; Published 1 January 2019

Academic Editor: Lu Zhen

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

Considering the rapid development of the tourist leisure industry and the surge of tourist quantity, insufficient information regarding tourists has placed tremendous pressure on traffic in scenic areas. In this paper, the author uses the Big Data technology and Call Detail Record (CDR) data with the mobile phone real-time location information to monitor the tourist flow and analyse the travel behaviour of tourists in scenic areas. By collecting CDR data and implementing a modelling analysis of the data to simultaneously reflect the distribution of tourist hot spots in Beijing, tourist locations, tourist origins, tourist movements, resident information, and other data, the results provide big data support for alleviating traffic pressure at tourist attractions and tourist routes in the city and rationally allocating traffic resources. The analysis shows that the big data analysis method based on the CDR data of mobile phones can provide real-time information about tourist behaviours in a timely and effective manner. This information can be applied for the operation management of scenic areas and can provide real-time big data support for “smart tourism”.

## 1. Introduction

With the rapid development of China's economy, the large and medium-sized cities have entered the Leisure Era. The quality of leisure has become an important evaluation criterion of the living quality of urban residents [1] and an essential part of public life. Witnessing the continuous growth of tourist flow in the scenic spots, it is vital to understand the accurate and real-time travel behaviour information of different types of tourists [2–10]. In recent years, the statistical monitoring methods for the tourist flow mainly focus on video surveillance, entrance gates, and other means [11]. The development of network and Internet technology also has enabled technical means of monitoring based on the WiFi provided by scenic areas and mobile phone application terminals. These flow monitoring methods require cameras, gateways, WiFi base stations, and other equipment to support them. It is difficult to install and implement in all scenic areas. Most mobile applications (such as WeChat) are oriented towards young and middle-aged users and cannot monitor tourists of all types.

The conventional estimation of macroscopic travel demands mostly relies on the empirical judgment on historic data by transportation practitioners. It is costly and the resolution is limited. Therefore, it is practical to develop new and better tools to automate the identification of tourist flows from large-scale mobility data sets. Some research is conducted with smart card fare data [12, 13] and mobile phone data [14].

The quantity of mobile phone users in China is constantly increasing. According to data released by China's Ministry of Industry and Information Technology (MIIT), as of January 2017, the number of mobile phone users in China reached 1.32 billion, and the penetration rate of mobile phones reached 96.2%. The number of mobile phone users in Beijing reached 3.869 million, and the penetration rate of mobile phones reached as high as 178.3%. Nearly everyone who travels carries their cell phone. Based on the continuous development of positioning technology in mobile communication, the tourist flow analysis with Call Detail Record (CDR) data can provide real-time valid data for scenic flow control, tourist diversion, traffic dispersion, safety management, and so on. And these

can support to improve the operation management of the scenic area.

Through the comprehensive analysis of scenic spots and tourists' travel behaviour, researchers have explored tourist flow and tourist behaviours in the corresponding scenic spots. Ferrari [15], through the processing and analysis of the Call Detail Record (CDR) data of mobile phones, analysed the temporal and spatial regulations of personal behaviour and the nature and law of events occurring in the city to provide a theoretical basis for the management of events [15]. Ahas and Silm studied the temporal and spatial distribution of Estonian tourists, offering big data support for tourism planning [16–19]. Based on the CDR data of mobile phones, Dong [20, 21] analysed the spatial and temporal situation of population movement within the Sixth Ring Road area in Beijing at both the regional and road network levels. Etison [22], utilizing the CDR data, established a resident traffic flow monitoring and management system for real-time data monitoring and statistics for mobile phone users in specific areas.

The existing research results have no specific target population for the analysis of mobile phone users and do not cover the research of users' individual behaviour. Therefore, this paper intends to analyse the main tourist attractions in Beijing by employing handset CDR data, focusing on the Forbidden City as the main research object. The author aims to make full use of the advantages of big data to effectively analyse the residence time, the spatial and temporal distribution of tourists, and the behaviour of tourists. The experiment results show that this big data analysis technology can become technical support for the management of urban tourist areas and effectively improve the accuracy of the current tourist flow monitoring.

## 2. Data Description

**2.1. Call Detail Record (CDR) Data.** Positioning with the mobile communication network is a technology that acquires information through mobile Call Detail Records (CDR) in the communication network.

**2.1.1. Location Mode of the Cellular Network.** Each zone covered by the mobile communication network is generally assumed as a regular hexagon, where each mobile user can act as a mobile station (MS), as shown in Figure 1. Regardless of the user's state, whether mobile or stationary, as long as the user calls or accesses the Internet, the user of the mobile phone will exchange data with the nearest mobile phone base station (BS). Therefore, the data containing the base station ID is recorded. Depending on the location of the base station, the current location of the user can be estimated. The specific positioning is shown in Figure 1.

When the user moves from Cell-1 to Cell-7, a total of three location areas (LACs) and seven cells are passed through in sequential order from Cell-1, Cell-2, Cell-3, Cell-4, Cell-5, and Cell-6 to Cell-7. If the users switch on and off, interact with the network, or receive the data regarding user steps, the user's location information is updated since LAC<sub>1</sub>, LAC<sub>2</sub>, and LAC<sub>3</sub> belong to different receiving areas. Taking Cell-2 and Cell-3 of LAC<sub>1</sub> and Cell-4 of LAC<sub>2</sub> as examples, if

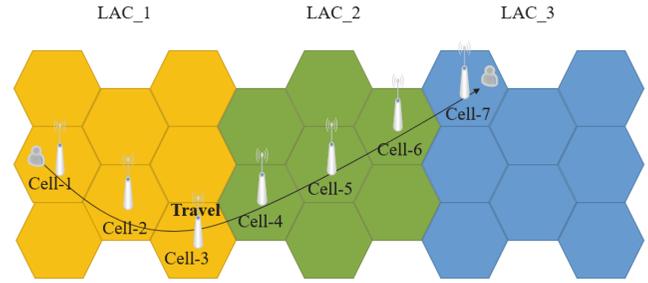


FIGURE 1: Location mode of mobile communication network.

there is no interaction with the network in the process of their moving, such as calling, texting, or surfing the Internet, the user's location will not be updated, since Cell-2 and Cell-3 are from the same location area (LAC<sub>1</sub>). In other words, when two cells belong to the same location area, the two cells have the same location area code. Although the cell area is changed, the location area remains intact, and thus the change in location will not be recorded in this situation. Since Cell-3 and Cell-4 belong to different location areas, when the user moves from Cell-3 to Cell-4, which is equivalent to the user moving from LAC<sub>1</sub> to LAC<sub>2</sub>, the centre will record the user's location changes regardless of whether the user has data interaction with the network during the movement.

This cell phone positioning method is called the Cell-ID positioning method, which is utilized as a web-based cell phone positioning technology without installing additional equipment, acquiring maintenance costs and upgrading the existing mobile network. Although this method cannot obtain the location information accurately, the macroscopic tourist flow analysis used in this paper has already met its accuracy requirement.

**2.1.2. Base Station Data of the Cellular Network.** In general, each base station (BS) has its own fixed location area (LAC). When a cell phone is registered in the communication network, the network will page the LAC location of the cell phone and obtain the corresponding Cell-ID. In this paper, the LAC is intercepted from the location area identification code (LAI) to identify the location area in the GSM network. The MSC area (hereafter called the MSC) is composed of all the location areas under its control. In addition to administering the subordinate location area, the MSC is also responsible for the reception of facsimile data. The geographical location of the base station (BS) mainly consists of information such as its administrative region, latitude, and longitude. The network attributes of the base station (BS) mainly cover information such as the BSC number used to control the communication, the running status, the antenna height, and the base station type.

The Cell-ID and LAC of the selected base station will be taken as the basic characteristic attributes of the base station. The recorded information includes the user's mobile phone ID, the user's phone status, the location area (LAC) of the user, the Cell-ID, the type of event triggered, and the recording position. The record table structure and data samples are shown in Table 1.

TABLE 1: Samples of the mobile phone base station.

ID	Status	BTS	Cell-ID	Community	SITENO	LAC
1	Working	139	45723	Tongzhou DistrictG1	45721	4280
2	Working	138	45722	Tongzhou DistrictG2	45721	4280
...	.....	...	.....	.....	.....	...
ID	Longitude	Latitude	Antenna Height	Antenna Azimuth	Coverage Condition	Station Location
1	116.887742	39.809074	33	120	Residential Area	Exercise Plaza in Yueshang Village, Tongzhou District
2	116.887742	39.809074	33	240	Residential Area	Exercise Plaza in Yueshang Village, Tongzhou District
...	.....	...	.....	.....	.....	...

TABLE 2: Example of CDR.

IMSI	Cell-ID	LAC	CTIME	TIMECHAR
96cbf1288b3ae673f1711c3falea9ac6	4259	11033	1422720000	20150201010000
54895a8b525d586e5de85417e0763eb3	4341	19507	1422720000	20150201010000
.....	.....	.....	.....	.....

Mobile communication networks cover a large amount of the population activity area, especially in urban areas. As shown in Figure 2, more than 70% of the base stations in Beijing are located in the urban area within the Sixth Ring Road area, with a total of 52,759 base stations. Meanwhile, according to statistics from the Ministry of Industry and Information Technology, the number of phone users in Beijing reached 20.62 million in 2015. With the mobile phone base station functioning as a fixed traffic detector and the user mobile phone as a mobile detector, the urban travel holographic information acquisition system is practicable. Therefore, the CDR data can be used in the analysis of extremely complicated travel behaviour.

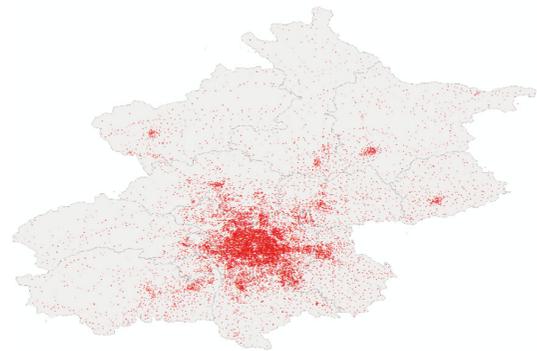


FIGURE 2: Mobile phone base station distribution in Beijing.

2.1.3. *User Data of the Cellular Network.* As with the base station attributes and types, when a user moves, updating the user location data acquired by the GSM network also must be defined by the user attributes. Each location update will generate a user location datum correspondingly. The GSM network encodes the user information not only to monitor the change of its location but also to confirm that users are properly connected to each other when using the network. Hence, it is necessary to correctly address this issue through encoding. Each user has a cell phone number during the use procedure. To protect the privacy of the user, the system will desensitize the user code hiding the user information, in the process of network storage. In this paper, based on the existing data, the author selected IMSI, Cell-ID, LAC, CTIME, and TIMECHAR as the location-updating attributes of the users under the condition of ensuring data quality. The Cell-ID and LAC meaning is consistent with the base station attribute property. The final selected mobile network user locations are updated in the format shown in Table 2.

The user data of the mobile communication adopted in this paper are recorded for one year. The data are collected every two seconds, with a data size of 52,759 pieces. Since there are 86,400 seconds in a day, this collection will amount

to 450 million pieces per day with an average daily data size of 40G. To address large amounts of data quickly, preprocessing the data, cleaning the noise, and converting the format are necessary steps.

2.2. *Processing Platform.* The CDR data are a kind of typically big data. Taking Beijing as an example, the CDR data for one day exceed 400 million pieces, which requires a large amount of calculation for processing. However, the processing capacity and I/O performance of a single machine cannot support such a large data calculation. At the same time, traditional relational databases, such as oracle, can build clusters, but when the amount of data reaches a certain limit, the query processing speed will become very slow, and the performance of the machine is very high. Thus, the Spark big data platform is considered to handle the CDR data in this paper.

The concept of Resilient Distributed Dataset (RDD) is adopted in the Spark framework. Considering that Map Reduce cannot complete effective data sharing at all stages of the parallel computing, RDD in the Spark framework makes up for this defect. Using this efficient data sharing and

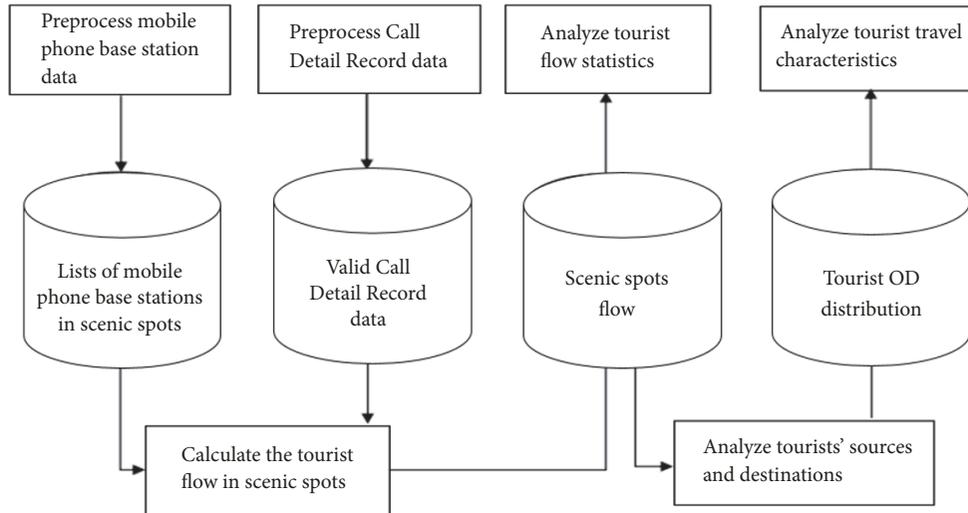


FIGURE 3: The overall structure.

Map Reduce-like operating interfaces, various proprietary types of calculations can be effectively expressed in the Spark framework, and similar performance can be achieved. According to the popular classification of the application area, big data processing can be divided into complex batch data processing, interactive data query based on historical data, and streaming data processing based on real-time data streaming. Because of the abundant expression capability of RDD, the unified large data processing platform capable of simultaneously dealing with the above three situations is derived on the basis of the Spark core. The goal of the Spark ecosystem is to integrate batch processing, interactive processing and streaming processing into the same software stack. In this paper, the Spark SQL interface is used, which provides a distributed SQL engine with a query speed 10 ~ 100 times higher than hive.

### 3. Tourist Flow Statistics and Characteristic Analysis Method

As shown in Figure 3, the overall flow chart depicting the scenic spot flow calculation and tourist travel characteristics analysis, the preprocessing of mobile phone base station data, and the CDR data are completed first to obtain the lists of all the base stations of each surveyed scenic spot and active phone users. Next, through matching the CDR data of the base stations in the scenic spots, the flow of each scenic spot will be obtained, and the tourist flow characteristics will be analysed. Based on the origin and the destination of tourists in the scenic spots, the tourist OD spatial distribution and the travel characteristics of tourists in the relative scenic spots can be obtained.

#### 3.1. Data Preprocessing

**3.1.1. Preprocessing of Mobile Phone Base Station Data.** Because the base station data are continuously improved under the operation of a communication company, there are

many basic data without value. The Spark SQL has processed 52,759 base station raw data into 43,022 valid pieces of data and extracted Cell-ID, LAC, longitude and latitude, base station type, coverage area type, and base station location as new attribute values. The specific details are shown in Table 3.

This paper analyses the main scenic spots in Beijing, including the Forbidden City, the Summer Palace, and the Olympic Forest Park. Working with data from more than 40,000 base stations in Beijing, Python language is adopted to write scripts on the ArcGIS platform, handling 20 scenic spots in a batch. A buffer zone of 100 metres is designed, and then the base stations are screened out in the scenic area by matching with the location of the scenic area, as shown in Figure 4. The Cell-ID and LAC attributes of the base station are selected to store in the lists of base stations in the scenic spots.

**3.1.2. Processing of the CDR Data.** The IMSI number, the only code in the CDR data for identifying the phone user, belongs to the STRING type, which is inconvenient for subsequent operations. Therefore, the IMSI number will be transformed into the LONG type through the HASH code, and its uniqueness and nonnegativity will be verified to ensure that one IMSI number can still determine one phone user. At the same time, due to the ping-pong switching phenomenon in the CDR data, “silent users” and ping-pong switching data have been filtered out by setting a threshold of frequency (PCF), leaving 67% of the data as the source for traffic information collection [9]. After completion of the above two steps, 40G of data per day can be reduced to approximately 16G, which significantly reduces the running time of the subsequent processing. The new CDR data table is shown in Table 4.

**3.2. Tourist Flow Algorithm Based on the CDR Data.** In this paper, the tourist flow statistics of the scenic spots are divided into five parts, namely, the total flow, influx, outflow, stagnant flow, and net increment and variation of the scenic spot

TABLE 3: Processed base station data.

Cell-ID	LAC	Longitude	Latitude	Type	Coverage	Address
435	4310	115.4275	39.9705	1	Residential area	Linchang Road side, Xiaolongmen Village, Qingshui Town, Mentougou District
6495	4148	115.4344	39.96334	1	Road traffic	Xiaolongmen Village, Qingshui Town, Mentougou District, Beijing
.....	.....	.....	.....	.....	.....	.....

TABLE 4: Example of processed CDR data.

HASH_ID	CTIME	LAC	CELL_ID
3304557668498423790	1423328400	4125	30519
5415176000769466540	1423328400	4352	13802
.....	.....	.....	.....

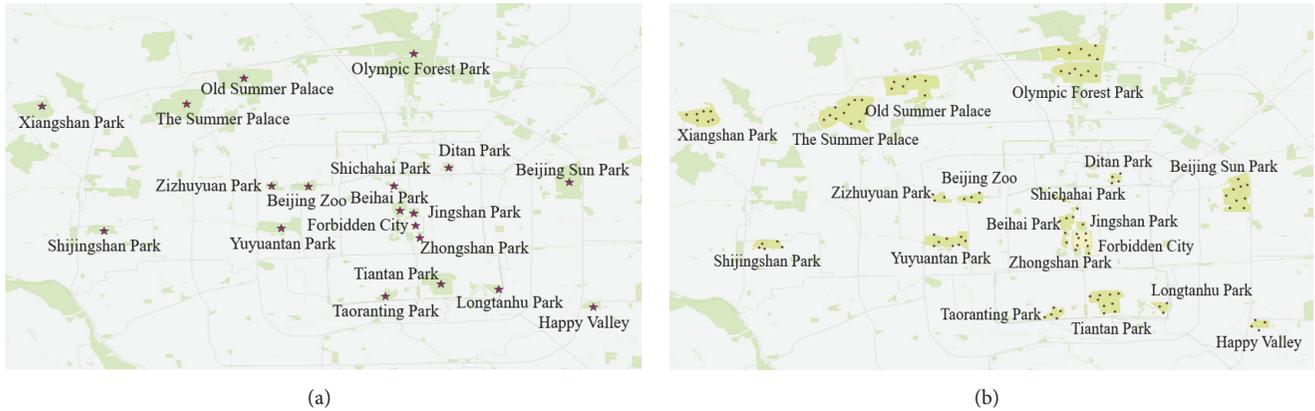


FIGURE 4: Mobile phone base stations in the scenic area. (a) Spatial distribution of typical scenic spots in Beijing. (b) Generating a list of mobile phone base stations in the scenic area.

during a certain period of time. The calculation method is as follows.

The specified study period is  $[a, b]$ , and the time interval from time  $a$  to time  $b$  is  $t$  hours; at the same time, a time period of  $[c, a]$  is set whose time interval is also  $t$  hour(s). Let the phone users of the first time period be collection  $B$ , in which the total number of users is recorded as  $C_b^t$ , and let the phone users of the second time period be collection  $C$ , in which the total number of users is recorded as  $C$ .

(1) The total flow of the scenic spot during a period of time is the total tourist number of the scenic spot in the time period  $[a, b]$ , with the collection expressed as  $B = C_b^t$ .

(2) The stagnant flow of the scenic spot during a certain period of time is the number of tourists in the scenic spot both within the time period  $[c, a]$  and within the next time period  $[a, b]$ , with the collection expressed as  $U_{ab} = |C \cap B| = |C_c^t \cap C_b^t|$ .

(3) The influx of the scenic spot during a certain period of time is the increased tourist number at the scenic spot during a certain period of time, that is, the number of tourists who were not in the scenic spot within the time period  $[c, a]$  but who appeared within the time period  $[a, b]$ , with the collection expressed as  $I_{ab} = B - U_{ab} = C_b^t - |C_c^t \cap C_b^t|$ .

(4) The outflow of the scenic spot during a certain period of time is the number of tourists who appeared within the time period  $[c, a]$  but did not appear within  $[a, b]$ , with the collection expressed as  $O_{ab} = C - U_{ab} = C_c^t - |C_c^t \cap C_b^t|$ .

(5) The net increment of the scenic spot during a certain period of time is the net increase in tourist number at the scenic spot during the time period  $[a, b]$ , calculated by the total influx minus the total outflow over this time period; this item can also be understood as the increased tourist number (which can be negative) of the time period  $[a, b]$  compared with the time period  $[c, a]$ . The collection is expressed as  $R_{ab} = I_{ab} - O_{ab} = B - C = C_b^t - C_c^t$ .

(6) The variation of the scenic spot during a certain period of time is the net increase in tourist number during the time period  $[0, b]$ , that is, the user number calculated by the net increase in tourist number within the time period  $[a, b]$  plus the total outflow within  $[0, a]$ . The collection is expressed as  $S_{0b} = R_{ab} + S_{0a}$ .

The calculation process is shown in Figure 5.

3.3. Tourist OD Analysis. The origin and destination analysis of tourists will reflect the important characteristics of tourist travel. Therefore, this paper considers using O (origin) and D (destination) of the trip to analyse the origin and destination of tourists and travel characteristics. The existing traffic zone division has been referred to the conduct tourist OD analysis.

3.3.1. Traffic Zone. This paper adopts the traffic zone division in the literature [20]. According to the processed mobile phone base station data and the CDR data (see Table 4 for the data format), the mobile phone base station is first defined in

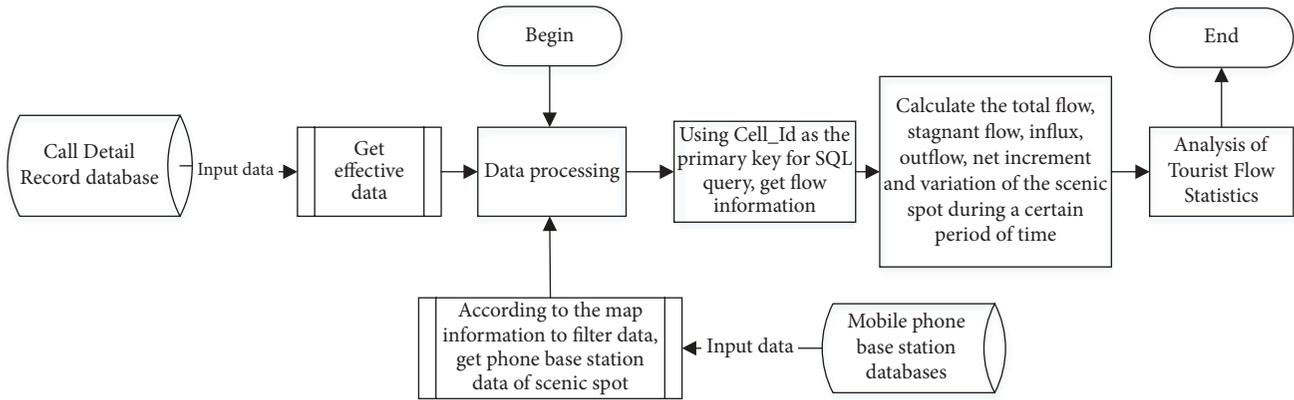


FIGURE 5: Tourist flow algorithm based on the CDR data.

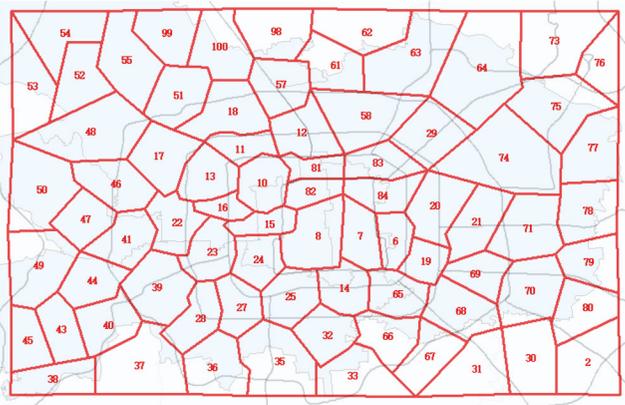


FIGURE 6: Traffic zone division of the area within the sixth ring in Beijing.

terms of traffic semantics as the residential area, work area, and road traffic. Then, it matches the geographic information system and divides these mobile phone base stations into “traffic zones” based on the mobility characteristics of the population directly related to traffic. Figure 6 depicts the division results for traffic zones in the urban area within the sixth ring in Beijing according to the above method.

According to the above method, each divided traffic zone contains a large number of base stations. First, the base station trajectory of each user needs to be extracted by a specific algorithm, through which the continuous location switching of the user at base stations can be obtained. The data in this step provide the database for acquiring the dynamic urban traffic OD, because the data are obtained by using the traffic zone as a unit and converting the user’s base station location information into the traffic zone location information to obtain the user’s zone switching trajectory.

**3.3.2. Tourist OD Analysis Algorithm Based on the CDR Data.** This paper primarily examines the OD flow of tourists in key scenic spots. Since February is a low season for tourism, the operating hours of most scenic spots are from 8:30 am to 16:30 pm. This paper takes the Summer Palace as an example to study the method for the tourist OD analysis.

The algorithm procedure is as follows:

(1) *Select Tourists Visiting the Scenic Spot.* The scenic spot opens at 8:30 am every day. Although it will take approximately 2 to 3 hours to visit the entire area, data should first be screened and should meet the following two conditions: (1) select users who remain in the scenic spot without going to other areas from 8:30 am to 11 am; (2) exclude those who remain in the scenic spot from 8:30 am to 4:30 pm (workers). The results provide user IDs for the tourists in the scenic spot.

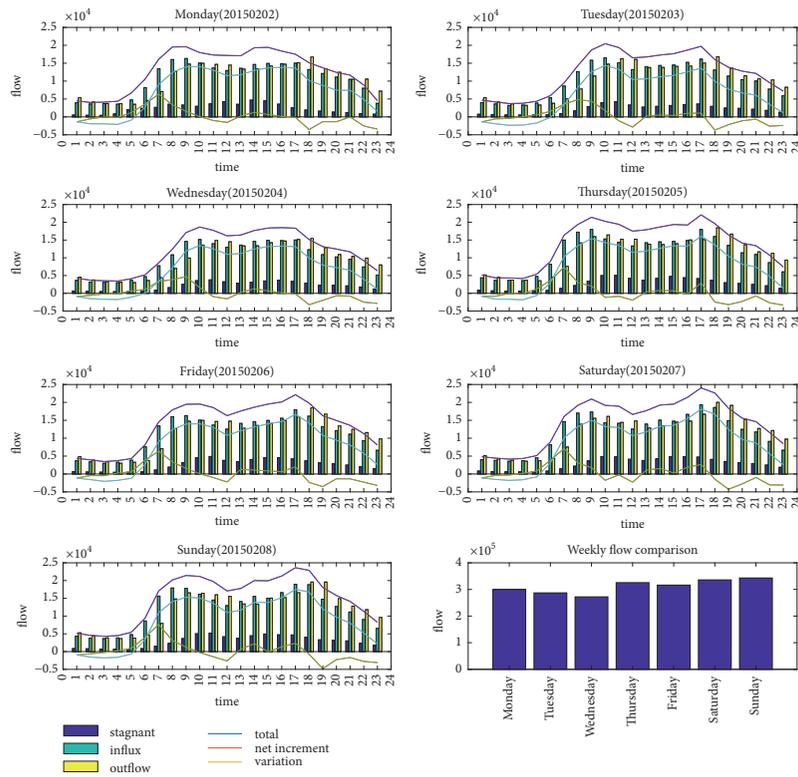
(2) *Obtain Tourists’ OD Information.* For tourists whose IDs have been selected, information on which traffic zone they originate from and where they finally arrive should be determined. This information is obtained by reverse querying, that is, using a user’s ID number to identify where she or he passed from 5 am to 8 am and from 11 am to 2 pm. Therefore, the user’s movement route from 5 am to 2 pm is determined.

(3) *Convert the Trajectory Data into an OD Matrix.* The trajectories obtained for each user may contain a significant amount of location data because the user is constantly moving. However, researchers only need the origin and destination of the users. To improve the fault tolerance, the result position data calculated by the median of the position data in this period multiplied by 0.7 plus the average of the position data in this period multiplied by 0.3. Traffic zones are determined based on the location range of the zones. The number of people in each traffic zone is then counted, and the OD matrix is finally produced.

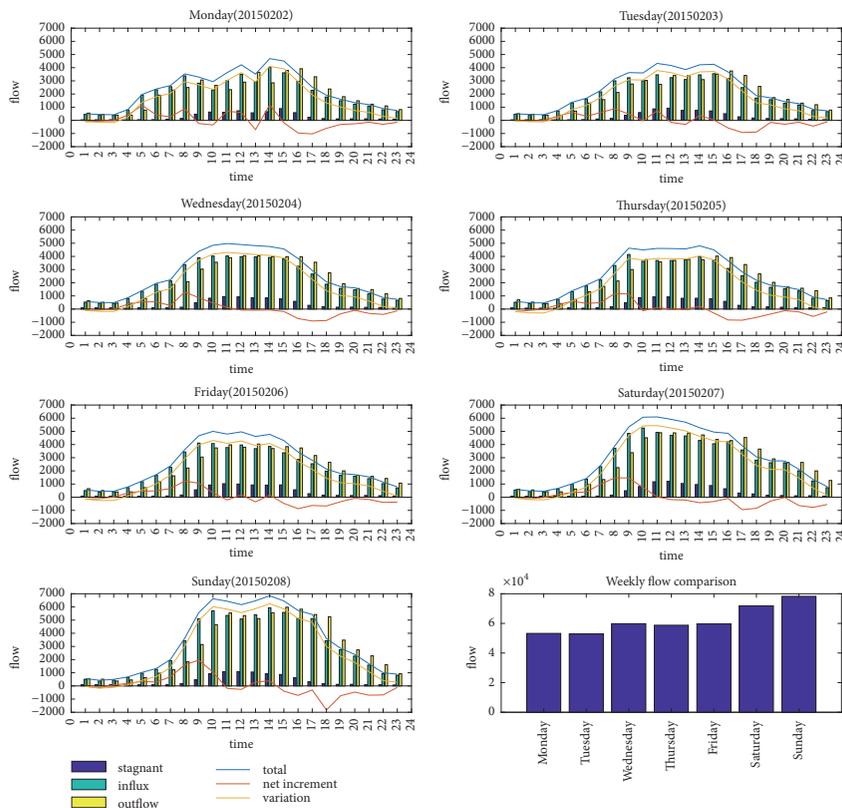
## 4. Experiment and Result Analysis

**4.1. Tourist Flow Statistics.** According to the above algorithm, we collect the statistics of tourist flow in 20 scenic spots and select 2 typical scenic spots including the Olympic Forest Park, Badaling Great Wall, obtaining the flow chart of the scenic spots as shown in Figure 7.

From the flow chart in Figure 7, we observe that the trend in the daily changes over time is similar in the same scenic spot. In general, most tourists begin their visit at 9:00~10:00 am and leave by 17:00~18:00 pm. At the same time, this trend



(a)



(b)

FIGURE 7: Weekly tourist flow chart of the scenic spots. (a) Olympic Forest Park. (b) Badaling Great Wall.

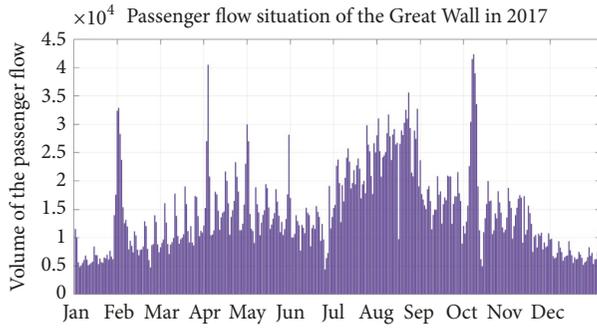


FIGURE 8: Tourist flow of the Badaling Great Wall for one year.

is not the same for traditional scenic areas and general parks; traditional scenic spots such as the Great Wall are closed after 5:30 pm. As a result, there are far more tourists remaining in the afternoon than the tourists who come in, and there are no visitors after 18:00 pm. For a general leisure park such as the Olympic Forest Park, another entering peak occurs after work from 17:00 pm to 18:00 pm, and a departure peak appears from 21:00 pm to 22:00 pm, consistent with daily behaviour.

With regard to another aspect, the number of tourists on the weekends and holidays is much higher than that on working days as shown in Figure 8. Examining the number of tourists in the scenic spots, through time dimensions we find that the number of tourists on weekends far exceeds the number of tourists on working days in the traditional scenic spots. The tourist number in the summer holidays is apparently higher than other seasons. And the tourist flow reached the peaks in holidays such as the Spring Festival, the Qingming Festival, the Dragon Boat Festival and the National day. The reason is that the resident population in Beijing has more time to go out a tour on weekends and holidays.

Comparing the total number of tourists in different scenic spots through spatial dimensions, we obtain the tourist flow order in Beijing shown in Figure 9. Some scenic spots are the hot spots for the tourists.

**4.2. OD Analysis of the Scenic Spot.** To vividly describe the origin and destination of tourists, taking the Summer Palace as an example, we conduct an analysis of tourists (total 1,733 tourists) who visited the palace from 8 am to 11 am on February 8, 2015. Combined with the tourist OD analysis algorithm above, the tourist origin and destination information of the Summer Palace scenic area are obtained together with the GIS platform. Figure 10(a) shows the number of tourists visiting the Summer Palace, whereas Figure 11(b) depicts the number of tourists who move from the Summer Palace to other places.

To better distinguish the different tourist flows, we utilize the thickness and colour of arrows and lines. As the flow rate increases, the width of the arrows and lines increases gradually, and the colour changes from green to red. From the spatial distribution map of the tourists, we can determine whether the origin of tourists (tourist attractions in the

Summer Palace area) or the whereabouts of tourists (Summer Palace area tourists) show “wave” spread, layer by layer, in line with regular traffic rules. Although tourists generally follow the law of “the distance increases and the amount decreases” after one tour, due to the particularity of the scenic spots, tourists choose to visit not only nearby scenic spots such as Summer Palace and Tsinghua, Peking University, but also farther scenic spots such as the Forbidden City and the Beijing Zoo.

To fully and comprehensively analyse the origin and destination of tourists in the Summer Palace, the data not only should be described qualitatively and visually from the space level but also should be described accurately and quantitatively from the statistics level to clarify the relationship between travel and distance.

First, the latitude and longitude of the centre point in the Summer Palace are extracted directly from the GIS platform to calculate the distance between the origin point (the ascent point) of each tourist and the centre of the scenic spot, that is, the geographical distance (spherical distance) between two points on the map. Then, we calculate the number of the above distances that are the same. Third, the arrival and departure probabilities of the tourists in the Summer Palace are calculated. Finally, we fit the distance distribution of the tourists by taking the travel distance as the abscissa and the scenic attraction quantity (occurrence quantity) probability as the ordinate. It is found that the composite exponential function is the best fit. Among the calculations, the fitting formula of the distance distribution of tourist origin is shown as follows:

$$P_O(d) = 0.5065 * e^{(-0.4146d)} + 0.1078 * e^{(-0.05757d)} \quad (1)$$

$$P_D(d) = 0.7927 * e^{(-0.9213d)} + 0.06561 * e^{(-0.1281d)} \quad (2)$$

In the formula,  $d$  is the distance between the scenic spot and the visitor’s location;

$P_O(d)$ , in the case of  $d$ , is the attraction probability of the scenic spots;

$P_D(d)$ , in the case of  $d$ , is the occurrence probability of the scenic spots.

From the formula and Figure 11, it is observed that the origin and destination of tourists are mainly distributed in the close range. That is to say, the majority of tourists still prefer to visit the closer scenic spots.

## 5. Conclusions

This paper presents a method based on the CDR data to analyse the tourist flow of scenic spots, including the collection and processing of the CDR data, tourist flow, travel OD, and other statistical analysis, which is all the helpful information for the operation management of the scenic spots. The conclusion is as follows:

(1) Through an analysis with the CDR data in the scenic spots in Beijing, the results show that the method can effectively analyse the tourist flows and other behaviour information, which can provide big data support to alleviate

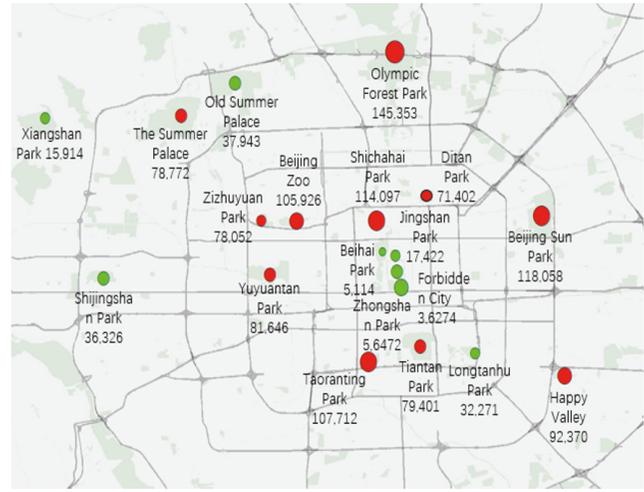
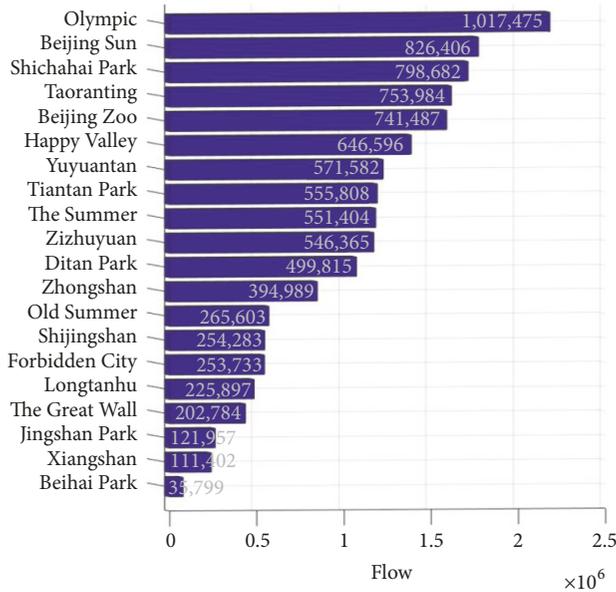


FIGURE 9: Analysis of scenic spot flow. (a) A traffic flow chart of the scenic spots in one week. (b) Scenic spot flow chart of daily average tourists.

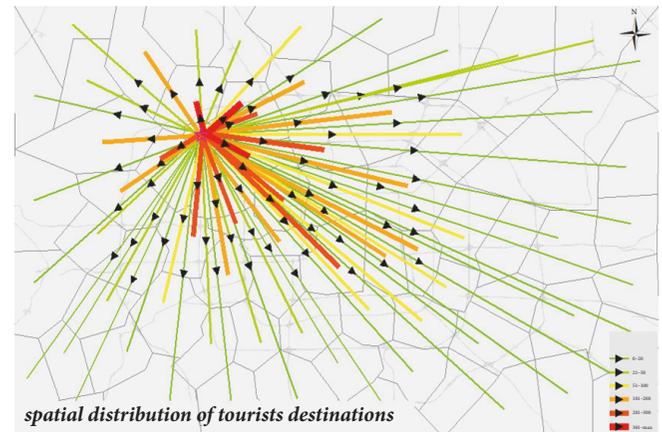
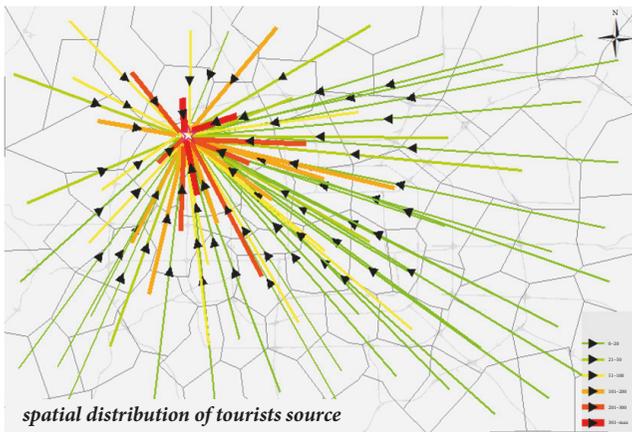


FIGURE 10: Spatial distribution of tourists in the Summer Palace. (a) Spatial distribution of tourist origins. (b) Spatial distribution of tourist destinations.

the traffic pressure of tourism lines, to alleviate the future traffic construction of tourism lines, and to promote the scenic area’s operation management.

(2) Travel OD analysis of the scenic spot can give the spatial origin and destination distribution of tourists, which can be used to help the manager of scenic spot to attract the tourists from different districts in the city.

(3) The analysis shows that the big data analysis method based on the CDR data of mobile phones can provide real-time information about tourist behaviours in a timely and effective manner. This information can be applied in scenic areas and can provide real-time big data support for “smart tourism”.

In the future, researchers can further improve the positioning accuracy and the calculation accuracy for the scenic spots, increase the ability to analyse the individual behaviour of tourists, enhance the application of the statistical analysis of tourist flow in scenic spots to support the operation management, and conduct in-depth research such as traffic monitoring, the tourist source analysis, and the comfort degree evaluation of tourists.

**Data Availability**

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

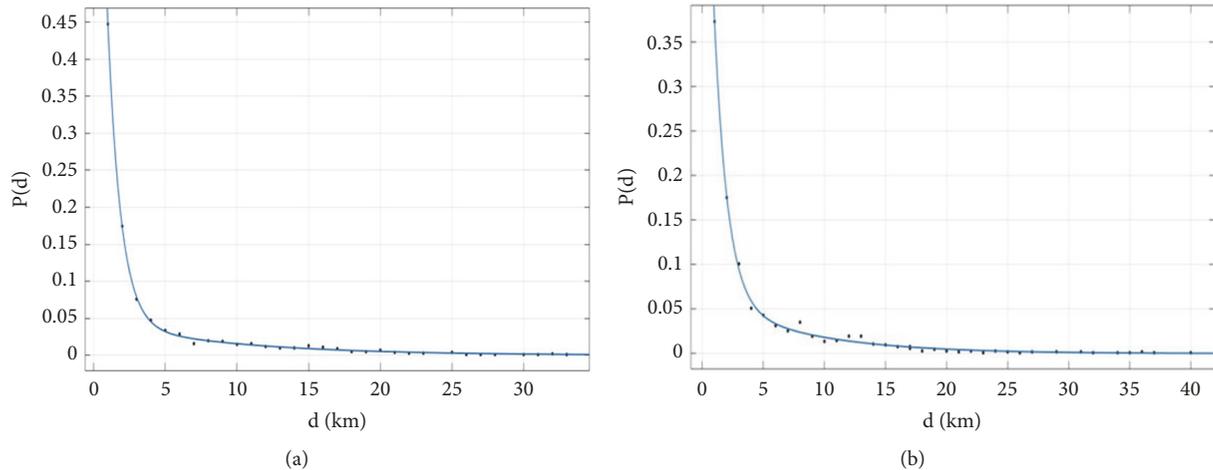


FIGURE 11: Distance distribution of visitors in the Summer Palace. (a) Origin distance distribution of tourists. (b) Destination distance distribution of tourists.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported in part by the National Science and Technology Pillar Program of China (Grant 2014BAG01B02) and Graduate Education Fund (Grant 145260522) and the 2014 National Advanced Training Course (Grant KIL14018530).

## References

- [1] J. Mair and M. Whitford, "An exploration of events research: event topics, themes and emerging trends," *International Journal of Event and Festival Management*, vol. 4, no. 1, pp. 6–30, 2013.
- [2] G. Yao, "A study on the construction framework of intelligent tourism," *Journal of Nanjing University of Posts & Telecommunications*, vol. 12, no. 2, pp. 13–16, 2012.
- [3] Y. B. Yan and H. E. Wen-Juan, "Analysis on the time-space evolution of the domestic tourism flow to China," *Economic Geography*, 2013.
- [4] N. Xianzhong et al., "On the future and the characteristics about domestic tourist flow to jiuzhaigou," *Journal of Mountain Research*, 1998.
- [5] R. Chen, C.-Y. Liang, W.-C. Hong, and D.-X. Gu, "Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm," *Applied Soft Computing*, vol. 26, pp. 435–443, 2015.
- [6] A. Worobiec, L. Samek, P. Karaszkiwicz et al., "A seasonal study of atmospheric conditions influenced by the intensive tourist flow in the Royal Museum of Wawel Castle in Cracow, Poland," *Microchemical Journal*, vol. 90, no. 2, pp. 99–106, 2008.
- [7] X. Wang, H. Dong, Y. Zhou, K. Liu, L. Jia, and Y. Qin, "Travel distance characteristics analysis using call detail record data," in *Proceedings of the 29th Chinese Control and Decision Conference (CCDC '17)*, pp. 3485–3489, May 2017.
- [8] E. C. Higham, "Tourist flow reasoning: the spatial similarities of tourist movements," 1996.
- [9] D. G. Wang, "The influence of Beijing-Shanghai high-speed railway on tourist flow and time-space distribution," *Tourism Tribune*, vol. 29, no. 1, pp. 75–82, 2014.
- [10] S. E. Zhong et al., "Spatial patterns of tourist flow: problems and prospects," *Human Geography*, 2010.
- [11] H. Dong, X. Wang, C. Zhang, R. He, L. Jia, and Y. Qin, "Improved Robust Vehicle Detection and Identification Based on Single Magnetic Sensor," *IEEE Access*, vol. 6, pp. 5247–5255, 2018.
- [12] A. Soltani, M. Tanko, M. I. Burke, and R. Farid, "Travel patterns of urban linear ferry passengers: analysis of smart card fare data for brisbane, queensland, Australia," in *Transportation Research Record Journal of the Transportation Research Board*, no. 2535, pp. 79–87, Transportation Research Board of the National Academies, Washington, DC, USA, 2015.
- [13] S. Rahman, J. Wong, and C. Brakewood, "Use of mobile ticketing data to estimate an origin-destination matrix for new york city ferry service," in *Transportation Research Record Journal of the Transportation Research Board*, no. 2544, pp. 1–9, Transportation Research Board of the National Academies, Washington, DC, USA, 2008.
- [14] F. Calabrese, M. Diao, G. Di Lorenzo, J. Ferreira, and C. Ratti, "Understanding individual mobility patterns from urban sensing data: a mobile phone trace example," *Transportation Research Part C: Emerging Technologies*, vol. 26, pp. 301–313, 2013.
- [15] L. Ferrari, M. Mamei, and M. Colonna, "Discovering events in the city via mobile network analysis," *Journal of Ambient Intelligence and Humanized Computing*, vol. 5, no. 3, pp. 265–277, 2014.
- [16] R. Ahas et al., "Mobile Positioning in Space& Time Behaviour Studies: Social Positioning Method Experiments in Estonia," *American Cartographer*, vol. 34, no. 4, pp. 259–273, 2007.
- [17] R. Ahas, A. Aasa, Ü. Mark, T. Pae, and A. Kull, "Seasonal tourism spaces in Estonia: Case study with mobile positioning data," *Tourism Management*, vol. 28, no. 3, pp. 898–910, 2007.

- [18] R. Ahas, A. Aasa, A. Roose, Ü. Mark, and S. Silm, "Evaluating passive mobile positioning data for tourism surveys: An Estonian case study," *Tourism Management*, vol. 29, no. 3, pp. 469–486, 2008.
- [19] R. Ahas, A. Aasa, S. Silm, and M. Tiru, "Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: case study with mobile positioning data," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 1, pp. 45–54, 2010.
- [20] H. Dong, M. Wu, X. Ding et al., "Traffic zone division based on big data from mobile phone base stations," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 278–291, 2015.
- [21] H. Dong, M. Wu, Q. Shan et al., "Urban residents travel analysis based on mobile communication data," in *Proceedings of the 16th International IEEE Conference on Intelligent Transportation Systems: Intelligent Transportation Systems for All Modes (ITSC '13)*, pp. 1487–1492, October 2013.
- [22] Y. Etison et al., "System and method for real-time monitoring of a contact center using a mobile computer," US20150098561, 2015.



**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

