Citation: LI Xingyue, LIU Fengjia, YU Wei, ZENG Liming. Analysis of Crowd Spatial Activities Based on Software Development Kit (SDK) Data [J]. Urban Transport of China, 2020 (4): 38–46, 10.

Analysis of Crowd Spatial Activities Based on Software Development Kit (SDK) Data

LI Xingyue[1], LIU Fengjia[1], YU Wei[1], ZENG Liming[2]

1. Zhejiang Urban & Rural Planning Design Institute, Hangzhou 310030, Zhejiang Province, China;
2. Zhejiang Merit Interactive Network Technology Co., Ltd., Hangzhou 310000, Zhejiang Province, China

Abstract: A spatial activity analysis method for specific urban spaces and specific populations at the mesoscale was constructed based on Software Development Kit (SDK) to explore the applicability of emerging data sources in the analysis of crowd spatial activities. With Zhoushan as an example, the paper analyzes the characteristics of residents' and tourists' activities in different locations respectively in four aspects: original starting point, travel path, travel intensity and spatial distribution. The regularity and difference of the two types of travelers are compared. The results show that resident activities are always near the residential area with small daily and hourly variations, while the tourist activities are concentrated on roads, docks and scenic spots with large overall daily and hourly changes. With more precise positioning and rich information in users' attributes, SDK data can serve as the basis for temporal and spatial optimization and customized travel services of travel for specific groups, public facilities, emergency management, etc. **DOI:** 10.13813/j.cn11-5141/u.2020.0403-en

Keywords: spatial activities; Software Development Kit (SDK) data; distribution characteristics; workplace and residence; residents; tourists; Zhoushan

0 Introduction

Urban spatial structure is closely related to the pattern of crowd activities. Understanding spatial activities of people is helpful to reveal the temporal and spatial regularity of urban space, to explore the potential drive of crowd activities in the city and to evaluate the rationality of urban infrastructure construction [1]. This is also important for urban traffic management, travel services, and optimal allocation of public resources.

The big data technology has been under rapid development and gradually penetrated in all fields of society; at the same time, it has provided a new basis and enabled the understanding of spatial activity pattern of urban population. With people's spatial activities becoming more complex, the concept of integrating complexity theory and big data analysis technology in urban transportation has been increasingly popular [1]. Currently, a large amount of research has been conducted on spatial activities of residents or tourists with big data, such as location-based services (LBS) data and Baidu point of interest (POI) data. 1) As a typical type of LBS data, mobile phone signaling data are widely used. These data can be used to study the characteristics of urban crowd activities and further identify their spatial activity patterns [2], obtain the activity characteristics of specific groups to support special urban regulation [3], and analyze activity rules of residents and guide their behavior to improve the quality of community and urban space [4–5]. 2) Compared with mobile phone signaling data, Weibo check-in LBS data originated from social media can provide data for gender, region, and other population attributes. 3) By using Weibo check-in LBS data, people can be identified according to user attributes and more detailed research can be performed with respect to time, space, activities, and users. Differences and patterns are discovered by analyzing the spatiotemporal behavior characteristics of different types of people [6–8]. 4) LBS data based on mobile APPs are real-time location data generated when users are using location request of mobile Internet through APP login, search, sending, receiving, push and other events. 5) Talking data have been used to quantify the vitality of the block and measure the impact of built environment on the vitality of streets [9]. In addition, through multi-source data fusion, commonality and difference of residents' regional characteristics from different communities can be examined [10]; spatial and temporal characteristics of urban work-residence relationship can be measured; spatial distribution characteristics of urban entertainment vitality can be assessed; the service scope of urban business districts can be estimated [11–12].

Existing studies have explored the application of various

Received: 2020-03-05
First author: LI Xingyue (1988–), male, from Hangzhou of Zhejiang Province, master's degree, engineer, Deputy Director of Creative Research Center and Deputy Director of Smart City Research Center, is mainly engaged in the research on big data application in urban transportation and transportation planning. E-mail: antli8899@163.com

LBS data to urban population spatial activities, spatial distribution, spatial vitality quantification, and other aspects. Most of the data sources are mobile phone signaling data and studies show similarities among methods mostly based on statistical analysis and clustering algorithm [3–5,7–8]. On the basis of completing a classification of population and space, the distribution characteristics of different populations and spaces are analyzed. However, due to the limitation of mobile phone signaling data, data locations can be accurate only at the base station and the accuracy of various analyses is generally at the block level. In addition, when the sample size is large, it is difficult to obtain a reliable clustering conclusion.

Different from previous studies, this paper is based on the mobile APP data collected through the Software Development Kit (SDK). SDK data can provide more accurate positioning information. The analysis of crowd activities can be accurate at the building level and activity spots can be located in the community, hospitals, shopping malls, roads, etc. SDK data can also provide insight into users' online and offline behavior and user characteristics to form multi-dimensional, rich and accurate crowd portraits, such as genders, ages, hometowns, hobbies, consumption levels, and other user behavior.

This paper takes Zhoushan City in Zhejiang Province, China, as an example. Zhoushan is a tourist city with a small number of permanent residents and a large tourist population. In addition to focusing on urban residents, this study also targets on tourists to analyze the characteristics of tourism spatial activities.

1 Data description

1.1 Data source

The SDK data were provided by the "Ge Tui" product of Zhejiang Daily Interactive Network Technology Co., Ltd. A wide variety of APPs have been created when the "Internet +" concept becomes increasingly popular as the rapid development of mobile Internet. The SDK data used in this paper are anonymous geographic location data of APP users. The data acquisition methods include event triggering and passive interaction. The data also include time and spatial locations of users. Event-triggered acquisition is initiated by a user's location request using the mobile Internet; it comes from user's instant location data generated when APP login, search, sending, receiving, push, and other events occur. Passive interactive acquisition comes from the background report of self-starting APP, which can also trigger and return data when users are not using the APP.

SDK has disadvantage and advantage for location data collection. The disadvantage is that, because the status of each user's equipment and the network environment are different, the reporting frequency among different devices is not completely consistent despite the consistent active acquisition frequency; the acquisition frequency will directly affect power consumption of the user's equipment, so it can't be adjusted at will. This limitation may result in quality differences among samples. Observing the distribution of samples before calculating the total quantity is needed in order to improve the efficiency and credibility of data collection. The advantage of SDK is that the location information directly comes from GPS module of the device, which has high precision and small deviation and can meet the needs of high-resolution crowd statistics, travel OD, and other scenarios. In addition to locations, the SDK also collects other anonymous data with considerable data dimensions.

1.2 Data characteristics

According to statistics, the average number of reports per day for a single user of SDK data is 40; the average number of reports per hour is 1.67. The hourly distribution of data reporting (see Fig. 1) suggests the peak periods of SDK data acquisition at 12:00–13:00 and 19:00–22:00, with more than two reports per hour; while the low-activity period is 2:00–4:00 with less than one report per hour on average. Based on the spatiotemporal correlation of data, travel trajectories of users can be constructed to reflect the changes of population mobility and agglomeration.

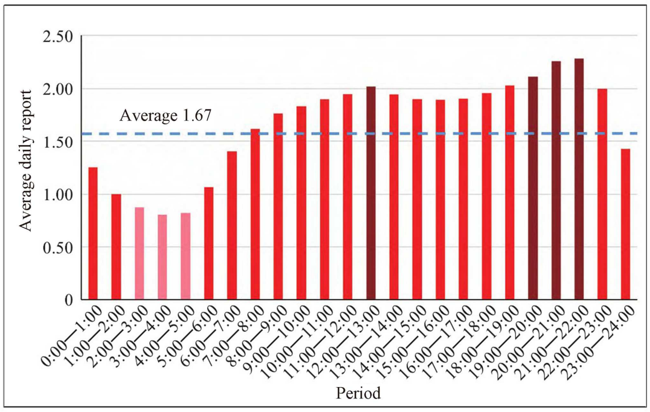


Fig. 1 Hourly distribution in SDK data reports

Compared with conventional big data such as mobile phone signaling data, SDK data have the advantage of offering more accurate geographical locations and powerful crowd portrait (see Table 1). The SDK data positioning can be accurate at the building level with geographic hash up to 9 string length (Geohash9; the theory of Geohash is to view the earth as a two-dimensional plane, which is decomposed recursively into smaller sub blocks or grids with the same code in a certain latitude and longitude range); it can customize the research scope according to different planning requirements. In addition, the SDK of mainstream APPs can carry out multi-dimensional and high-precision crowd portraits of users and obtain various attribute information of users.

Tab. 1 SDK data and LBS data

Item	Mobile phone signaling data	Other LBS data	SDK data
Data source	China Mobile, China Unicom and China Telecom	Baidu, Ali, Tencent and other Internet companies	Mobile App
Collection method	Active reporting and passive interaction	Active triggering at login	Active triggering and partial passive interaction
Openness	Low	Not open to others, only to their own business	High
Continuity	High	Low	High
Positioning accuracy	Block level 200–1 000 m	GPS precision, big fluctuation	Positioning accuracy of Geohash9 reaching 4.8 m × 4.8 m at the building level of 20–40 m scale
User portrait	Need to integrate with other data to obtain user attributes	Genders, regions and other simple attributes	Genders, ages, regions, hobbies, consumption levels and other multidimensional portraits
Influence factor	Low positioning accuracy in suburban under the influence by the distribution of base stations	Only active access, limited by users and background	Affected by user behavior and usage habit

1.3 Study area

This study uses 140 days of continuous SDK data from April 1 to August 18, 2018, in the Zhoushan city area. The data include over 3 billion data records. After data cleaning, 25.92 million daily data records produced by 480 thousand users and approximately 54 data record for each user are identified. Specifically, the number of valid location points is 8 762, under Geohash with string length of eight (i.e., Geohash8) for a study area of 19.0 m × 38.2 m.

2 Analysis method of crowd spatial activity

This paper focuses on residents and tourists. First, the crowd is classified as residents and tourists to allow for analysis of their individuality and generality. The research mainly analyzes workplace and residence locations, paths, space-time distribution characteristics, and tourism characteristics. The decision threshold used in the algorithm is a optimal choice based on the comparison among multiple thresholds.

2.1 Crowd classification

The analysis data are based on source datasets after signal deduplication and other preprocessing methods. The following rule is applied to distinguish residents and tourists: Users who have reported their locations in the target city and have appeared for more than 70 days are identified as local residents; users who stay for more than 4 h and less than 7 d are identified as tourists.

2.2 Workplace and residence analysis

Residence and employment are two important activities of a city. The distribution of residents' workplaces and residences has a great impact on urban transportation, planning, and management. With the urban expansion and space reconstruction, a series of changes have taken place in the relationship between workplaces and residences in many cities. Traffic congestion resulted from the separation of workplaces and residences has been widely recognized. SDK data can be used to provide an accurate analysis of workplaces and residences of residents to obtain their distributions in a city.

The algorithm of workplace and residence analysis is as follows: 1) Multiple daily data records per resident are scored based on the reporting time. The scoring system takes user's life routine as reference; the time periods when users are more likely motionless are assigned high scores, while the opposite are assigned low scores. The periods 7:00–10:00, 12:00–13:00 and 18:00–21:00 are not included, because users are likely commuting or out of home during these time periods. Therefore, it is necessary to remove them to avoid biased calculation and improve the credibility of the results. 2) According to the score assignment, a Geohash8 level workplace and residence with the highest score is generated every day. 3) All the recognition results are combined for historical summary and iterative update, with the most frequently appeared places identified as the user's workplaces or residences (see Fig. 2).

2.3 Path analysis

Amid the rapid growth of urban travel demand, traffic congestion is increasingly serious and becomes a significant factor restricting urban development. The generation of urban passenger transport results from the movement of people. The route selection of travelers, especially those travelers using passenger cars, is critical to urban traffic management and organization. Previous research based on travel behavior modeling of traveler's route choices usually lacks practical application or intuitive observation.

With high-precision positioning coordinates, users' travel path can be directly located in the urban roadway network. Through the analysis of spatiotemporal correlation of users' feedback data, the use of urban roads or the distribution of commuter travel paths in a urban roadway network during specific periods (e.g., morning rush hours) can be assessed.

For the analysis of tourist routes, the utilization of urban roadway network and the dependence of tourists flow on main roads can also be evaluated to clarify roadway positioning. Through comparison with the travel paths of residents, the impact of tourists flow on the urban roadway network can be qualitatively analyzed and the organization and management of urban tourism traffic flow can be further optimized.

Path analysis needs to capture the characteristics of user movement. The specific algorithm is as follows: according to the time sequence of user report records, calculate the average speed between each user's positions and record the data with speed larger than 1 m/s; then iterate the location of the moving records to obtain the distribution of crowd paths.

The path analysis in this paper is a superposition analysis of the user's overall activity path. Its aggregation point may

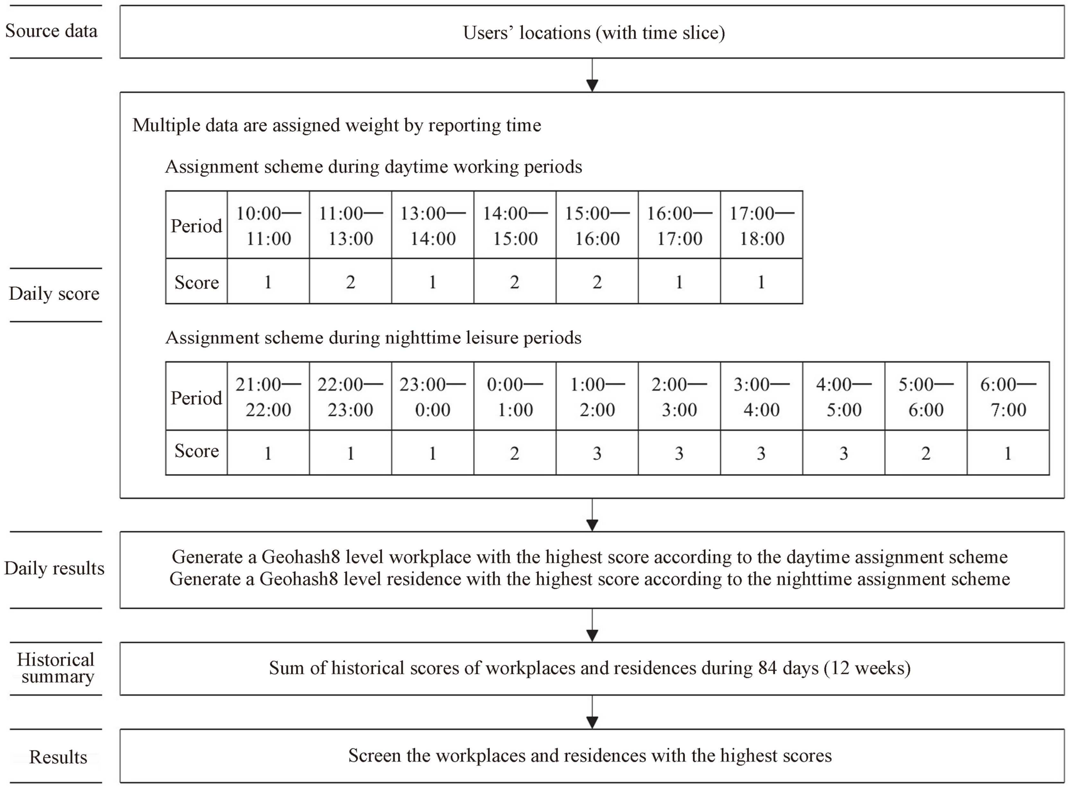


Fig. 2 Calculation process of residences and workplaces

not only locate on the road, but also in a shopping mall, a park, a scenic spot, and other possible activity spots or route points. Path analysis takes the user as a unit and focuses on capturing users' dynamic movement instead of a static record. User path is not limited to high-speed movement of motor vehicles but covers the activity paths of non-motor vehicles and walking.

2.4 Analysis of temporal and spatial characteristics

The activity space of residents can directly reflect the use of urban space and quality of life; at the same time, it can indirectly reflect the vitality and layout of a city. Exploring spatial and temporal characteristics of urban residents' activities has always been a popular topic in studies of urban spatial activities with big data. The distribution of tourist activities is closely related to the time and the tourist attractions, which directly reflects the popularity of each scenic spot and tourists' inclination to visit. It is of practical significance to guiding the development of tourism resources and reasonably allocate the service facilities of scenic spots.

The user aggregation algorithm is introduced to analyze the temporal and spatial distribution of crowd activities: filter out user records on specific dates, such as working days, weekends, and holidays; calculate the duration of stay at each location and determine the threshold value of the duration for different groups; for the records with residence time longer than the threshold, if a user's positions suggest an A—B—A pattern, treat Position A as two stays and calculate their duration; count the number of people after deduplication at each position.

In addition, residents' weekday travel largely focuses on commuting and official business; however, shopping, recreation, medical treatment, entertainment, and other activities mainly occur on non-working days. With the clarification of user's residence location, travel destinations of residents during non-working days can be analyzed to investigate their dependence on urban public service facilities.

The analysis of travel destinations on non-working days needs to involve travel distance of residents. The algorithm is as follows: select data from typical non-working days and calculate the distance between the location and residence on the basis of the user aggregation algorithm; keep the records with distance greater than 1 km and count the number of people who have been deduplicated at each location.

2.5 Analysis of tourism characteristics

Where tourists come from and how they arrive have been the key information for urban tourism industry. Statistics of the number and origins of tourists are critical to improving the management, service and marketing of the tourism industry. It also serves as a valuable reference for planning and construction of large-scale external transportation facilities in the city.

An entrance gateway of tourists refers to the way in which tourists enter the city, which generally includes toll stations, wharfs, stations and airports. The demand and proportion of sea, land, and air passenger transport can be obtained by analyzing tourists entering the gateway, which can guide adjustment of passenger transport structure, collect and distribute tourists flow, and plan new passenger transport hubs.

Different from previous approaches of obtaining user attributes through signaling interaction between the operator's base station and mobile phone and combining with the mobile phone's ownership location information, SDK data can be directly used to identify the tourist's origin and access portal information through crowd portrait and further generate statistical results.

In addition, tourists often travel with a variety of visiting and entertainment activities. Given the excellent crowd portrait ability of SDK data, certain behavior characteristics of tourists can be observed through various APPs during the tour, which is difficult to achieve by mobile phone signaling data and Internet check-in data. With the analysis of tourists' photo-taking behavior as an example, the SDK data in certain time span can be selected to filter and retain the camera APP report records of tourist users. The average number of photographers at each location can be obtained by deduplication and statistical analysis.

3 Case study

Acoording to the above analysis method of crowd spatial activities, the feasibility and rationality of each algorithm are verified with the SDK data from Zhoushan of Zhejiang Province for 140 consecutive days as an example. Zhoushan, located in the northeast of Zhejiang Province, is a typical island city with a land area of 1 440 km^2. By the end of 2018, the permanent population of Zhoushan was 1.17 million, and the urbanization rate was 68.1%. Zhoushan has a long history and rich landscape resources. There are more than 1 000 Buddhist cultural landscapes, natural landscapes of mountain and sea, and island fishing custom landscapes. Among these resources, Putuo Mountain is a national scenic spot and the only Grade 5A tourist scenic spot in Zhoushan. As a typical tourist city, Zhoushan has a substantially high proportion of tourist population. In 2018, the number of tourists was 63.21 million and the number of overnight visitors was 177 thousand.

3.1 Analysis of differences between residents and tourists

Based on population classification, the time density characteristics of residents and tourists are analyzed. The average daily reported data volume of all users is 54. The average daily report of each resident is 42, with an average of 1.75 items per hour; the average daily report of each tourist is 63, with an average of 2.63 items per hour.

As shown in Fig. 3, the fluctuation in hourly data for residents is small, while that for tourists is relatively large. The overall data volume of tourists is larger than that of residents, indicating that tourists use mobile APPs more frequently. The two peak time windows in the distribution of residents' average daily reports are 12:00–13:00 in the afternoon and 20:00–21:00 in the evening. In addition to the above peak periods, the tourists' average daily reports suggest another peak time window of 10:00–11:00.

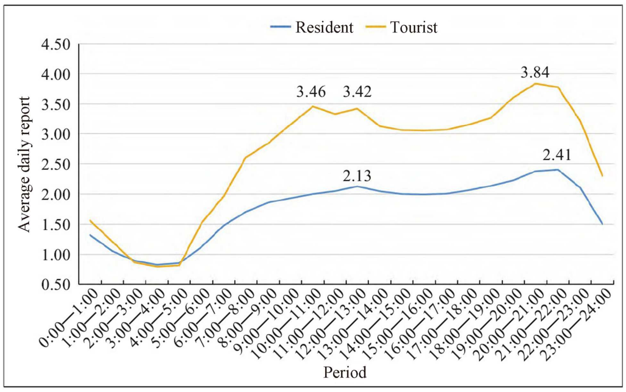


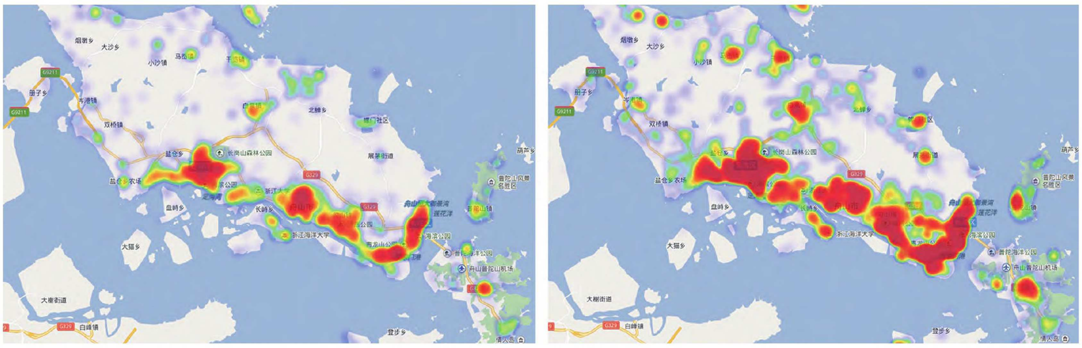
Fig. 3 Hourly distribution of data volume from resident and tourist reports

3.2 Analysis of workplace and residence of residents

The workplace and residence of residents are analyzed with the reported records of 140 days. The identification results (see Fig. 4) suggest that residential areas of Zhoushan are mainly distributed in the south coast of the Zhoushan Island, including Dinghai District and Putuo District; residential areas also include the town centers and scenic spots. The spatial distribution characteristics of workplaces and residences are similar; however, workplaces are more concentrated than residences in the downtown area in the southeast of the island.

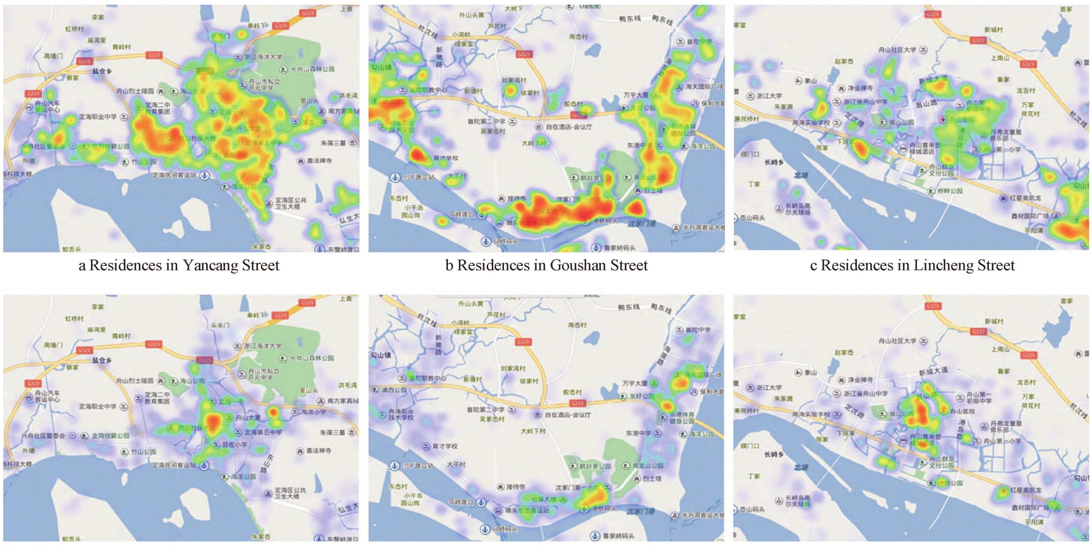
Compared with the current distribution of community and commercial areas in Zhoushan, it can be found that (as shown in Fig. 4): for residence, the red gathering areas on the Zhoushan Island include the Xishan Community and the Dongshan Community of Dinghai District in the south as well as the Hedong Community of Putuo District in the southeast; workplaces largely locate in the commercial center of Dinghai District in the west, Zhoushan Municipal Government area in the middle, and the Binhai commercial district of Putuo District in the southeast. These distribution patterns from the analysis are generally consistent with the actual workplace and residence distribution in Zhoushan.

With respect to the distribution density, residential areas are scattered, while workplaces are relatively concentrated. This distribution is more significant at the street level. Yancang Street, Goushan Street, and Lincheng Street have concentrated workplaces, while residences are scattered (see Fig. 5).



a. Residences b. Workplaces

Fig. 4 Distribution of residences and workplaces



a Residences in Yancang Street b Residences in Goushan Street c Residences in Lincheng Street

d Workplaces in Yancang Street e Workplaces in Goushan Street f Workplaces in Lincheng Street

Fig. 5 Distribution of residences and workplaces along streets

3.3 Path analysis

The analysis period of residents' travel paths is during morning crush hours from 7:00 to 9:00; for tourists, the analysis period is from 7:00 to 20:00. The data of 30 days in April, 2018 are selected to see the path distribution (see Fig. 6 and Fig. 7).

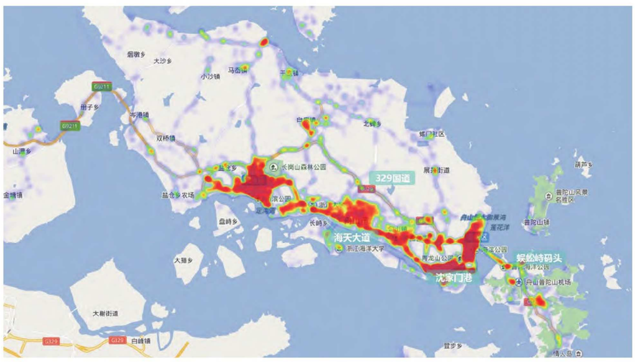


Fig. 6 Resident travel path analysis

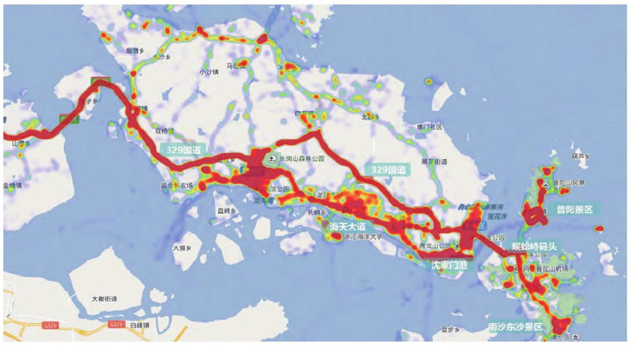


Fig. 7 Tourist travel path analysis

The analysis suggests that most residents' paths concentrate in residential areas, mainly including Yancang Street, Lincheng Street, and Goushan Street. Haitian Avenue, as a primary arterial connecting the three streets, carries the majority of commuting traffic flow and serves as the main path for residents. In contrast, the eastbound and westbound of National Highway 329 is an important corridor in Zhoushan, but the residents' paths do not cover it much. On the airport connecting roadway, Yongzhou Expressway, and other roadways going out of the city, there are also fewer residents' paths. The distribution of residents' paths during morning crush hours indicates Haitian Avenue as the primary longitudinal commuter path of Zhoushan; the latitudinal commuter paths mainly include Linchang Road, Provincial Highway 231, and Dingma Roadway.

Different from the characteristics of residents' paths, tourists' routes are mainly distributed on urban arterial roads, docks, and scenic spots, given their destination-oriented travel. Some major routes include Yongzhou Expressway, National Highway 329, Haitian Avenue, Shenjiamen Port, Wugongzhi Wharf, Putuo Mountaina, and Nansha Dongsha Scenic Spot. The National Highway 329 and Haitian Avenue serve as the primary roadways for tourists' access to the scenic spots. As the connector between the highway and Zhujiajian Scenic Spot of Putuo Mountain, the National Highway 329 is much more demanded for tourists' instead of residents. Haitian Avenue is the major commuter route for residents in Zhoushan and also an important corridor for tourists. Therefore, traffic management is much needed for this avenue; vehicles with non-local licenses could be restricted during morning and evening rush hours as needed. In addition, tourists can be directed to National Highway 329 and the northern ring road to enter and leave the scenic area to separate commuters and tourists and mitigate the impact of through traffic.

3.4 Analysis of spatial and temporal characteristics

The spatial and temporal characteristics of residents' and tourists' activities are analyzed based on the records on April 30, 2018 (the Labor Day holiday), August 17, 2018 (Friday), and August 18, 2018 (Saturday).

Based on the distribution of residential users' dwelling time, 30 min that corresponds to the 95 percentile in the distribution is selected as the threshold for assessing residents' activities (see Fig. 8). The temporal and spatial distributions of residents' activities are calculated with the user aggregation algorithm of residents (see Fig. 9).

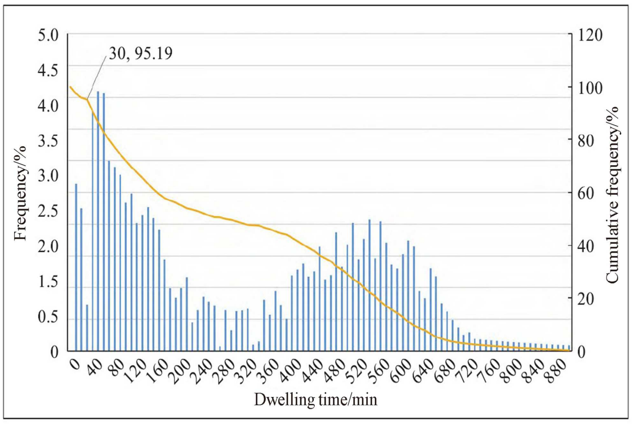
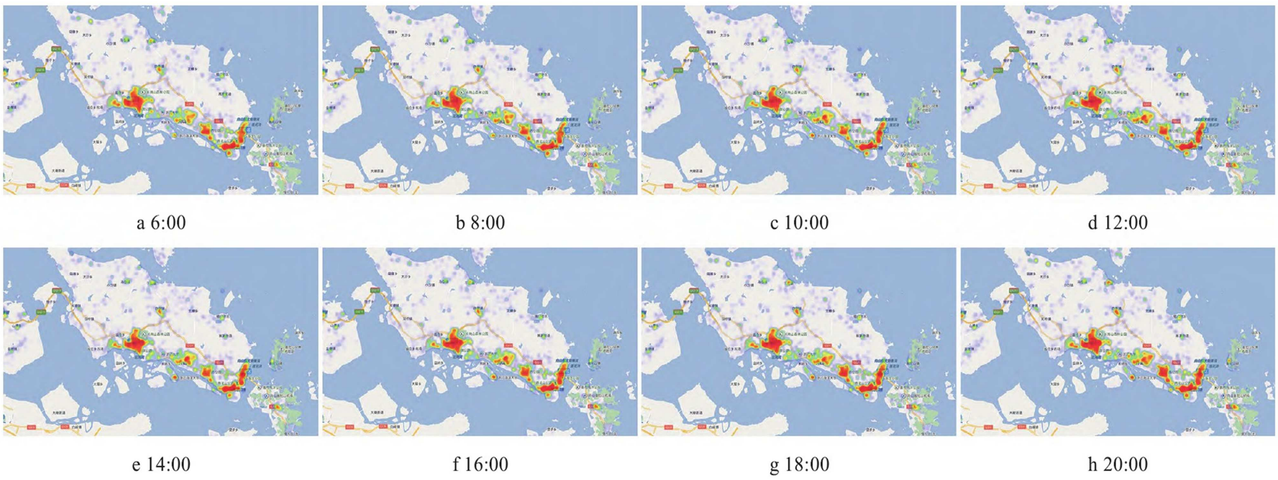


Fig. 8 Distribution of resident dwelling time

For tourist activities, the assessment threshold of 20 min is selected (see Fig. 10). The temporal and spatial distributions of tourist activities are also calculated with the user aggregation algorithm of tourists (see Fig. 11).

The temporal and spatial distributions show that activities of residents concentrate in the southern and southeast coastal areas. Activities are most distributed in Yancang, Lincheng, and Goushan streets. Similar distribution characteristics are found in all time periods throughout the day with small temporal variations, which reflects the general pattern of residents' activities. These characteristics suggest that residents of Zhoushan consistently have activities near their residence locations; the activity space of residents is typically restricted by their residences with relatively short average travel distance.



a 6:00 b 8:00 c 10:00 d 12:00

e 14:00 f 16:00 g 18:00 h 20:00

Fig. 9 Temporal and spatial distributions of resident activities

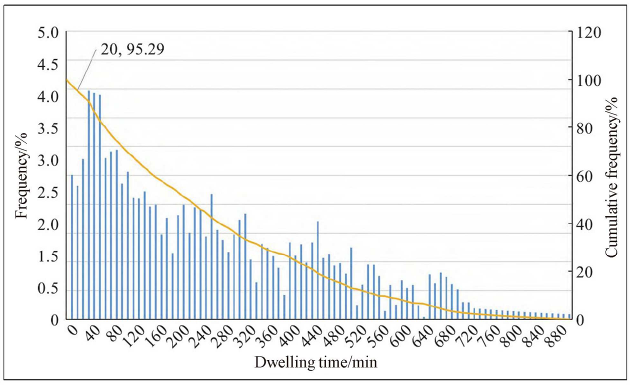


Fig. 10 Distribution of tourist dwelling time

The activity space of tourists has larger temporal variation than that of residents. No activity distribution is found in the early morning; activities during the daytime mainly occur in the Putuo Mountain, Zhujiajian, and other scenic spots, as well as National Highway 329 and Haitian Avenue. The activities at night decrease with a growth of distribution out of scenic spots and on roads out of Zhoushan, which is expected given the travel characteristics of tourists.

The temporal and spatial distributions of activities have also shown that on non-working days, residents' travel mostly concentrates near the residence area and little involves scenic spots. Analysis is conducted to examine travel destinations of residents in Zhoushan on non-working days.

On the macroscale, travel destinations of residents on non-working days largely remain close to the residence area; at the street scale, destinations of residents mainly locate in Kaihong Square, Haizhongzhou Commercial Square, Donggang Ganghui Square, Zhoushan Library, Zhoushan Gymnasium, Zhoushan Hospital, Putuo Hospital, and other shopping malls and public service facilities. In contrast, urban parks and scenic spots are less attractive to residents. In the future, the environmental improvement and quality enhancement of urban parks should be emphasized by creating specific features and enriching activities to improve the attractiveness and vitality of parks and reduce idle rate of public space.

3.5 Analysis of tourist characteristics

1) Tourist origin

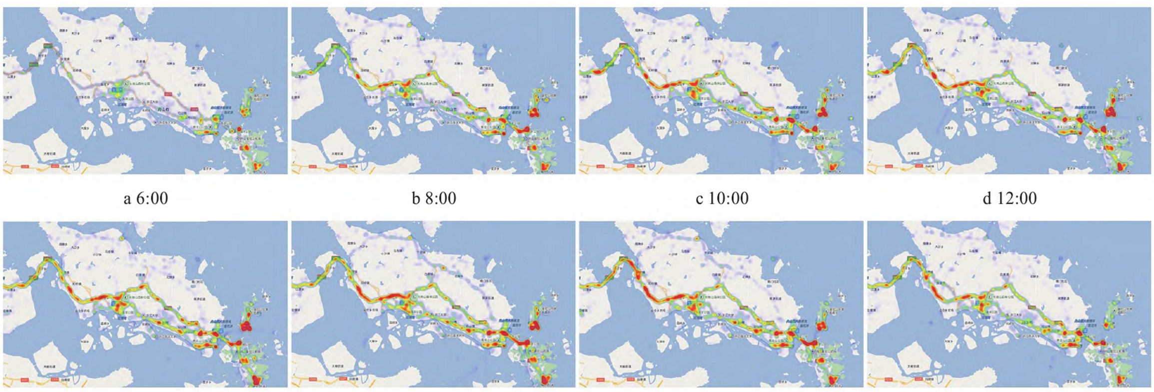
The origins of tourists and associated statistics can be directly obtained from the SDK data: tourists in Zhoushan mainly come from the Yangtze River Delta, especially the area in Zhejiang Province. Jiangsu, Zhejiang, and Shanghai tourists account for about 70% of all tourists in Zhoushan. The top three origin cities are Ningbo, Shanghai, and Hangzhou, accounting for 14.8%, 13.7%, and 9.8%, respectively, followed by Suzhou, Fuyang, Jinhua, and other cities (see Table 2).

Table 2 Distribution of origin cities of tourists

No.	City	Province	Proportion/%
1	Ningbo	Zhejiang	14.8
2	Shanghai	Shanghai	13.7
3	Hangzhou	Zhejiang	9.8
4	Suzhou	Jiangsu	3.7
5	Fuyang	Anhui	3.3
6	Jinhua	Zhejiang	2.3
7	Shaoxing	Zhejiang	2.2
8	Jiaxing	Zhejiang	2.2
9	Taizhou	Zhejiang	1.7
10	Wenzhou	Zhejiang	1.7

2) Tourists' entry

Zhoushan is an island city and has three ways of entry for tourists: highway toll stations, sea ports, and airport. With the data on April 30, 2018 as an example, the location where each visitor first appeared is counted as the visitor's port of entry. Highway entry accounts for 82.7%, which is the main way for tourists to enter Zhoushan. The above analysis of origins has shown that most tourists in Zhoushan are from Zhejiang Province without long distance travel; as expected, this correlates with higher proportion of tourists entering Zhoushan through highways.



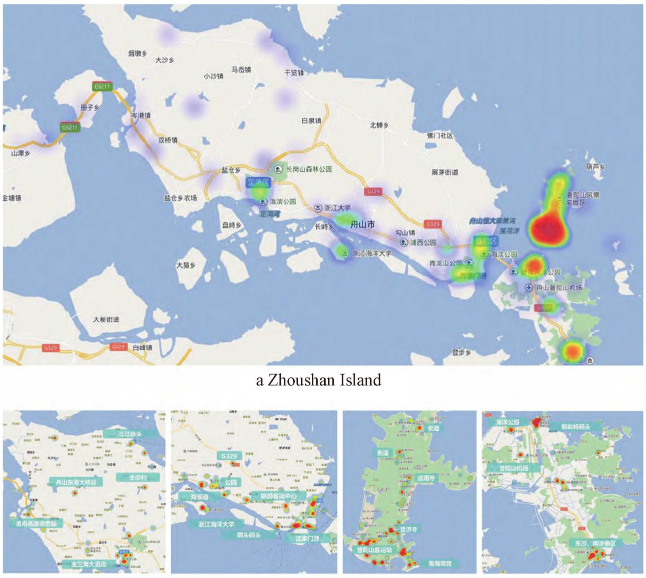
a 6:00 b 8:00 c 10:00 d 12:00

e 14:00 f 16:00 g 18:00 h 20:00

Fig. 11 Temporal and spatial distributions of tourist activities

3) Tourist photo-taking locations

With three months of data from April 1 to June 30, 2018, the photo-taking locations of tourists in Zhoushan are analyzed to summarize the distribution results (see Fig. 12). Major scenic spots are typically the places where tourists take photos. As a famous scenic spot, Putuo Mountain attracts a large number of tourists to visit and take photos. Nanhai Guanyin, as a landmark and popular scenic spot, has the highest density of photo-taking activities.



a Zhoushan Island

b Dinghai District c Shenjiamen Port d Putuo Mountain e Nansha Scenic Spot

Fig. 12 Distribution of sites where tourists take photos

The distribution analysis based on SDK data in this paper has suggested more detailed results than other studies. The distribution of workplaces and residences can be presented with the resolution of urban blocks; the distribution of travel paths, destinations, and activity hotspots can be located to specific shopping malls, squares, hospitals, roadways, scenic spots, etc. In addition, the distribution analysis of photo-taking locations, as presented in this paper, cannot be performed with other LBS data. High-precision positioning and multi-dimensional portrait are the advantages of SDK data.

4 Conclusion

In this paper, SDK data were applied to analyze the spatial activities of urban population. With Zhoushan as an example, the spatial activities of urban residents and tourists were studied and compared. The spatial activity characteristics of multiple groups of population were summarized with respect to workplace and residence distribution, travel path distribution, and temporal and spatial distribution of activities. Analysis methods based on high-precision positioning data were also developed to assess crowd activities and urban space.

As a new source in the current big data era, SDK data provide strong support for studies of crowd activities and urban space. The application of the SDK system with a large amount of data requires substantial data processing capability, as well as investment of high-cost cluster development and daily maintenance. With the high-performance infrastructure, relevant personnel need to master a variety of programming skills, such as HiveQL, Spark, Python, and Java. Qualified personnel and high-performance infrastructure are the bases for the maximum use of SDK data to benefit traffic analysis and other valuable applications.

References

[1] Yang Dongyuan. The Integration of Big Data and Complexity Theory in the Study of Residents' Activity Space[J]. Urban Planning Forum, 2017 (2): 31–36 (in Chinese).

[2] Zhong Weijing, Wang De, Xie Dongcan, et al. Dynamic Characteristics of Shanghai's Population Distribution Using Cell Phone Signaling Data [J]. Geographical Research, 2017, 36 (5): 972–984 (in Chinese).

[3] Zhaomengni, Jiao Hongzan, Sun Zhihao. 基于多源数据的城中村居民时空间活动特征分析：以武汉市为例 [C]//中国城市规划学会. 共享与品质：2018 中国城市规划年会论文集 (05 城市规划新技术应用). Beijing: China Architecture & Building Press, 2018: 1083–1094 (in Chinese).

[4] Yan Qing, Li Chenggu, Chen Cai, et al. Characteristics of Activity Space and Community Differentiation in Changchun: A Study Using Mobile Phone Signaling Data [J]. Human Geography, 2018, 33 (6): 35–43 (in Chinese).

[5] Zhong Weijing, Wang De. Urban Space Study Based on the Temporal Characteristics of Residents' Behavior [J]. Progress in Geography, 2018, 37 (8): 1106–1118 (in Chinese).

[6] Zhang Ziang, Huang Zhenfang, Jin Cheng, et al. Research on Spatial-Temporal Characteristics of Scenic Tourist Activity Based on Sina Microblog: A Case Study of Nanjing Zhongshan Mountain National Park[J]. Geography and Geo-Information Science, 2015, 31 (4): 121–126 (in Chinese).

[7] Wang Bo, Zhen Feng, Zhang Hao. The Dynamic Changes of Urban Space-Time Activity and Activity Zoning Based on Check-in Data in Sina Web [J]. Scientia Geographica Sinica, 2015, 35 (2): 151–160 (in Chinese).

[8] Zhou Yan, Li Yanxi, Huang Yueying, et al. Analysis of Classification Methods and Activity Characteristics of Urban Population Based on Social Media Data [J]. Journal of Geo-Information Science, 2017, 19 (9): 1238–1244 (in Chinese).

[9] Niu Xinyi, Wu Wanshu, Li Meng. Influence of Built Environment on Street Vitality and Its Spatiotemporal Characteristics Based on LBSPositioning Data [J]. Urban Planning International, 2019, 34 (1): 28–37 (in Chinese).

[10] Li Weifeng, Cheng Xiaoyun, Yu Qing, et al. 基于多源数据的居民活动区域特征分析 [C]//中国城市规划学会. 规划 60 年：成就与挑战——2016 中国城市规划年会论文集 (04 城市规划新技术应用). Beijing: China Architecture & Building Press, 2016: 957–975 (in Chinese).

[11] Zhao Pengjun, Cao Yushu, Wan Hairong. 基于 LBS 数据的职住平衡对比研究：以北京城区为例 [C]//中国城市规划学会. 持续发展理性规划：2017 中国城市规划年会论文集 (05 城市规划新技术应用). Beijing: China Architecture & Building Press, 2017: 97–115 (in Chinese).

[12] Jiang Jinliang. 多源大数据融合的城市商圈识别及空间活动特征研究 [C]//中国城市科学研究会. 2017 城市发展与规划论文集. Beijing: China City Press, 2017: 505–512 (in Chinese).