Analysis of Crowd Movement Based on Mobile Phone Signaling

Yanhao Luo, Lu Zhang *

School of Electrical and Information Engineering, Quzhou University, Quzhou, 324000, China

*Corresponding Author: zhanglu@qzc.edu.cn

Abstract

This study is only some attempts to analyze and manage the application of mobile phone signaling data in urban events. In the distribution law of large passenger flow time, this study focuses on the changes of peak and peak duration of passenger flow. However, if multi-source measured data is used for calculation, using the refined scenic spot survey data and scenic spot purchase data to assist, it will further improve the credibility of the data and show the analysis of individual time and space behavior. The underlying causes of this phenomenon, such as the decision-making factors such as the action machine and the choice of preferences, rely on questionnaires, interviews and other survey methods, combined with small data for in-depth research. The preliminary analysis in this study selects the representative areas with the most significant problems and the most prominent phenomena, further explores the internal motivations of the phenomena, and obtains a more in-depth analysis, and meticulously describes and verifies the established spatial distribution rules. The planning and management of event activities are linked to the actual needs and are properly applied.

Keywords

Crowd flow; Regional density distribution; Highway network density; Urban spatial structure.

1. Introduction

National Day Golden Week, the user consumption data released by many online travel platforms shows that during the National Day, the country's vast travel force will exceed 700 million person-times, and Hangzhou's tourism enthusiasm ranks eighth in the national city list. At the same time, with 157 domestic scenic spots in the "11 before the announcement of price cuts or free tickets, the National Day travel army must be the first batch of tourists to harvest ticket benefits, however, if visitors do not make online bookings in advance, or because of scenic spot restrictions And was blocked at the door. The study found that the spatial distribution of festivals and passenger origins is different from the traditional concentric pattern. The overall pattern is “conical, and the areas with the largest increase in passenger flow are mostly convenient for subways and densely populated. The spatial statistical units with increased travel rates in the province account for the total. 88.43%. Through tracking, 90.38% of the types of tourists' recreation behavior have changed. On the first day of the National Day, the behavior of the above tourists is transformed from the indoor passive recreation behavior represented by home relaxation and shopping to the proactive nature. Outdoor recreation.

This paper clarifies the relationship between the growth rate of the mobile phone signaling data (festival than usual) and the increase rate of the hotspot area in the scenic spot by predicting the site and time period of the scenic spot; Three-level early warning program. The research proves that mobile phone signaling data has strong practical significance in human flow early warning and tracking identification, which can realize the transition from passive monitoring to active prediction and monitoring to early warning.

2. Data collection

The analysis and prediction of crowd flow needs to grasp the law of urban form evolution, internal and external contact law, population post distribution characteristics, and residents' travel mobility

characteristics, so as to scientifically and reasonably predict the destination of crowd flow, in order to set up layout infrastructure and provide support services. Traditional sampling surveys consume a lot of manpower and material resources, low sampling rate, low precision, long data update period, and can only obtain data of specific time segments, and it is increasingly difficult to adapt to the planning management needs of the new era. The widespread use of mobile phones and massive mobile phone signaling data analysis technology provide an efficient and convenient means of data acquisition for urban planning. Compared with traditional resident travel surveys, mobile phone signaling data analysis has the advantages of wide coverage, large analysis samples, low implementation cost, and long-term continuous monitoring.

Cell phone signaling refers to signal data of auxiliary communication other than user data (such as call voice, short message content, and internet data packet). The mobile phone signaling data is directly derived from the mobile communication systems of operators (China Mobile, China Unicom, China Telecom). The mobile communication system is composed of a network subsystem (NSS), a base station subsystem (BSS), a mobile station (MS), and an operation support system (0SS). The NSS mainly performs the switching function, mobility management, and security management. The database function includes functional entities such as the Mobile Switching Center (MSC). The BSS is mainly responsible for performing functions such as wireless transmission and reception and radio resource management, including a base station controller (BSC) and a base transceiver station (BTS) controlled by it. The MS is a mobile client device part, including a mobile terminal (MS) and a customer identification card (SIM). 0SS is used for mobile user management, mobile device management, and network operation and maintenance. Each component of the mobile communication system uses different interfaces for communication. For example, the Um interface is used between the MS and the BTS, the A-bis interface is used between the BTS and the BSC, the A interface is used between the BSC and the MSC, and the E interface is used between different MSCs. Monitoring different interfaces can collect mobile phone signaling of different content. At present, Chongqing Mobile and Chongqing Unicom have realized the signal acquisition based on the A interface and the E interface. The main contents are shown in Table 1. The collected signaling events mainly include the user making a call (calling/called), sending and receiving short messages, hanging up, location update, and BSC switching.

Mobile phone signaling main field show as Table 1.

Table 1 Main fields of mobile phone signaling

Numbering	Field Name	Field meaning
1	***MSID***	**Encrypted SIM card number**
2	***Timestamp***	**Signaling acquisition time**
3	***LAC***	**Location area number**
4	***CellID***	**Area number**
5	***Cause***	**Signaling event type**
6	***MSCID***	**MSC number BSC**
7	***BSCID***	**Numbering**

3. Research methods

3.1 Stroke recognition

The task of the itinerary recognition is to convert a track point that cannot be directly understood into a recognizable dwell position and a movement between each dwell position. As shown in Figure 1, the spatially discrete track points are divided into two categories: the stay point and the move point.

The stop point is used to obtain the position or position range when the traveler performs the stay activity, and the travel path of the traveler is generated by moving the point. Further, the information such as the staying time and the purpose of staying activities can be mined and analyzed by the staying

position information, and information such as the number of trips, travel time, travel mode, travel distance, etc. can be extracted from the travel path.

The mobile phone location data used was obtained by installing the location record software on the recruited volunteer mobile phone. This time, the M0biletrack software was used. The software supports WinCE, SymbianS60 and Andr0id and other smart phone systems, using the Cell-ID positioning method. The acquisition time interval is 1 min, and the text data is obtained after the base station information processing by the website, as shown in FIG. 2 . It mainly includes the time, longitude and latitude information of the location record. It also needs to add information such as record keywords and volunteer numbers to distinguish the location records.

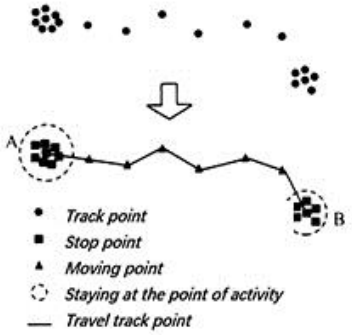

Figure 1 shows a schematic diagram of the identification

```
<Placemark>
<name>(1)19 day 0:03 a.m</name><ExtemdedData>
<Data name="location">
<value>39.934216,116.333984</value>
</Data>
<Data name="starting time">
<calue>2011-09-19 00:03:06</value>
</Data>
<Data name="End time">
<value>2011-09-19 00:03:06</value>
</Data>
```

Figure 2 Mobile phone positioning data text format

At present, there are few research methods for the trajectory data obtained by the locating method and the related recognition algorithms for the locating data of the mobile phone. Based on the full study of the characteristics of mobile phone positioning data, this paper proposes a speed-based track point combined stroke recognition algorithm, and verifies the accuracy of the algorithm through programming experiments. The main flow of the algorithm is divided into three parts: (1) speed calculation; (2) candidate stop position generation; and (3) stop point recognition. 1.1.2 The speed data obtained from the original trajectory data does not contain velocity information. The first step of the algorithm needs to be recorded according to the track point.

The longitude, latitude and time information calculates the speed of the pedestrian at each track point. Strictly meaningful instantaneous velocity calculations are difficult and complex, so consider using the average velocity over a trajectory where the trajectory points are located instead. For mobile phone positioning trajectory data, its positioning accuracy and drift characteristics depend on the distribution density of mobile base stations. The distance drifting away from the city center is large, and sometimes 1 or 2 km of positioning drift and jitter may occur. In response to the above problems, try to use the straight line distance between the track points instead of the track path distance to participate in the speed calculation. When calculating the speed of the track point, the sum of the adjacent distances between the track points P1, P2, P3, P4, and Ps is no longer calculated, and the track points P1 and Ps are directly calculated. The straight line distance between them is the length of the path that the traveler passes between t1 and ts. The formula is as follows.

$$p_3 . v = \frac{D_{(1,5)}}{\Delta t_{(1,2)} + \Delta t_{(2,3)} + \Delta t_{(3,4)} + \Delta t_{(4,5)}}$$

3.2 Time distribution of the number of people in festivals and weekdays

The trial and error based on data selection threshold selection, and the reliability judgment described by the actual phenomenon. The number of samples in the scenic spot, the scenic spot, and the scenic spot in the first two weeks of the festival can be extracted. The time of the first record of the day is the time of the visitor. The last recorded time is taken as the time of the scenic spot, and the middle time is in The state of the scenic spot. The trend of the sample changes with time is: the holiday has an increase of 66.03% than the weekday passenger flow, and the increase of passenger flow after the opening day is not large. It is consistent with the trend of the increase and decrease of the actual passenger flow, in line with the overall law of the three cherry blossom f Interests: about 375 visitors on working days. Holidays after the opening. The passenger flow can reach more than 1,875 person-times; the rainy weather has a greater impact on the number of tourists. According to the above qualitative judgment. The extracts of the increase and decrease of the total sample size in the two weeks is consistent with the increase and decrease of the actual passenger flow: the holiday is larger than the passenger flow on weekdays The passenger flow on the rainy day is small. Less than the non-rainy days of the weekdays. According to the sample and the actual number of passenger flows increase or decrease the regularity of the display. The representativeness of the sample data can be Qualitatively determined.

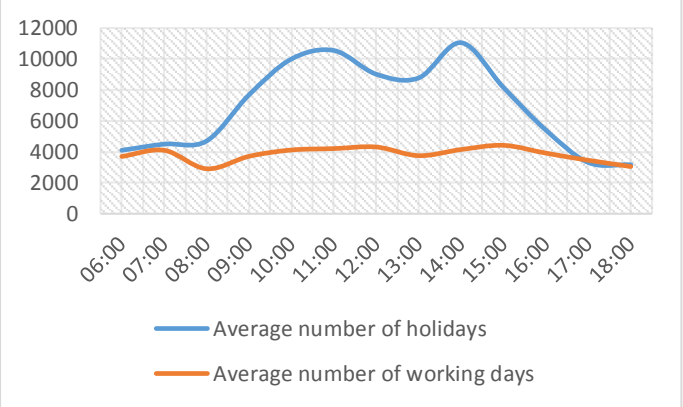

Figure 3 Number of mobile phone signaling records in scenic spots during holidays and non-holidays

Based on the time attribute of the mobile phone signaling data, the law of increasing and decreasing the number of people entering the scenic spot and the number of people in the scenic spot during the time period of the opening scenic spot (6:00-18:00) can be read hour by day: the peak hours of rapid increase of people flow on weekdays and festivals . It is 8:00-11:00 on weekdays. The festival is 8:00-10:00: the peak duration of the arrival of the big passengers on weekdays and festivals is different. The former was extended 1 hour later than the latter. It may be related to the increase in the number of people playing and the increase in the time required for long-distance travel. Among the relevant standards, there are tourists who produce a population density index value from the crowded to the crowded feeling in person/hm2. As a "never crowded to crowded" person 1:3 density limit. Multiply the value of the population density indicator by the area of Yaowangshan. It is possible to calculate the number of visitors in the scenic spot that may give visitors a feeling of crowding. Refer to the population density index 1 of the relevant crowding experience, according to 122 people/hm2 as the “delimitation criterion from no crowding to crowding feeling. The entrance of Yaowang Mountain Scenic Spot is about 180hm2. It is concluded that Yaowangshan may have crowded feelings in the scenic spot. The number of people in the actual scenic spot is 128,000. The number of people in the scenic spot is 17.2% of the actual scenic spot. Based on this ratio, the number of mobile phone signaling users of Yaowangshan will be identified as 58116. Multiplied by a percentage of 17.2%, it can be inferred that the mobile phone signaling sample reaches 99996. Visitors in Yaowangshan may have a feeling of crowding. The daily crowding time is about 10:00-16:00, which is kept for 6h; It is about 9:30-17:00. It stays 7.5h, 1.5h more than usual.

3.3 Area division and extraction of network base stations

During the planning process, we are unable to study the problems that arise within the region as a whole

The research area must be divided into different parts, the travel demand relationship between these parts must be analyzed, the nature of the problem should be found, and the solution should be found. The different parts of the division are the districts.

The division of the scenic area causes a large number of microscopic sources to be transformed into a small number of macroscopic sources, and a large number of disordered individuals are gathered in a limited number of independent spaces. The basic unit that forms the analytical characteristics is realized

The phenomenon of travel and the relationship between holidays and weather. Reduce the difficulty of travel analysis and forecasting. Area division is based on the internal differences of the target area. Past division

Methods The demographic characteristics, land characteristics, social characteristics, economic characteristics, etc. of the main areas are based on these "indirect features" to infer the consistency of the characteristics of the scenic spots, thereby dividing the areas.

This chapter combines the division of the area with the mobile communication network base station, maps the base station area to the land, analyzes the base station location and the base station to ask the crowd to move, and obtains the direct reflection characteristics and travel occurrence.

Flow trend data with attractive crowds. According to these "direct features" divisions, the accuracy will be greatly improved compared to traditional areas based on experience and "indirect features."

The area division feature based on the base station should reflect the location of the base station, and at the same time reflect the occurrence and attraction characteristics of the line. This chapter selects the base station latitude and longitude to reflect the location of the base station. In reflecting the characteristics of travel

In terms of the traffic characteristics of the base station, in order to not excessively distinguish the specific value of the base station traffic and preserve the characteristics of the base station, this chapter selects the difference between the base station attribute and the peak value of the holiday of the base station traffic and the peak of the non-holiday. The feature selection of the scenic area based on the base station of the mobile communication network base station is set in order to reflect the source and sink of the travel during the planning process, and therefore, the base station sample used for the area division is as shown in the table 2.

Table 2. base station sample used for the area division

feature		description	effect
Base station location characteristics	longitude latitude	Geographical latitude and longitude	Characterizing the geographical location of a base station
Base station characteristics	Traffic semantic category number	Semantic attribute category	Characterizing the attributes of the base station, reflecting the trend of changes in the basic semantic features of the base station
	Peak-Valley	The difference between the peak of the holiday and the peak of the non-holiday	Distinguish the size of the traffic volatility of the holiday

Due to the difference in signaling collection mechanism and actual signaling data quality, not all mobile phone movements can be detected within a certain period of time. Assume that the probability

of a call, SMS, and shutdown event is equal for each mobile phone. In the absence of spatial data loss, the mobile phone signaling collection can be regarded as a simple random repeated sampling process, and the detected mobile phone is pumped. Medium sample. The statistic of the sample is used to estimate the characteristics of the parent (all mobile phone users), and the error is only related to the sample size and the overall standard deviation. Take different approaches in different application scenarios:

When you only need to obtain a proportional relationship, you do not need to expand the results. For example, the external travel ratio based on the sample is the actual external travel ratio of the area. The absolute amount of the line is estimated under the condition that the maternal capacity is clear, and the maternal volume and the proportional relationship can be used for estimation. For example, if the total number of mobile phone users in the area and the proportion of external commuting are known, the external commute number is the product of the total number of mobile phone users and the sample external commute. The absolute amount of the line is estimated under the condition that the parent capacity is not clear, and the sample expansion is required according to the sampling rate. For example, the external travel volume of a certain area is obtained by analyzing the sample data of a single operator, and the total external travel volume of the area is known. The factors that need to be considered for the expansion at this time include the effective sample size, the market share of the operator, and the mobile phone retention of the resident. Rate (permeability), mobile phone detection rate.

The market share and mobile phone ownership rate can be obtained through questionnaire survey. The mobile phone detection rate is closely related to the coverage of the base station, the analysis period and the scope of analysis, and it is difficult to directly detect it. Since it is difficult to directly determine the sampling rate of each region, the average sampling rate (effective sample to total population ratio) can be approximated during the expansion.

The results of mobile phone signaling analysis are directly used to determine the planning method. The traditional planning analysis method mainly relies on experience to draw qualitative conclusions. The mobile phone signaling analysis can realize the quantitative description of the contact interval. In the regional passenger flow forecast, the development of each node play area is relatively mature, and the influencing factors change slowly. The prediction based on the current OD is more accurate than the “four-stage” model prediction. It can overcome the gravity model and can only consider the distance factor and cannot consider the travel mode and travel route of tourists. Defects in the selection factor. Time granularity optimization of basic features of base station traffic semantic calibration. This paper extracts four basic characteristics (fixed amount of scenic spot, inflow of scenic spot, outflow of scenic spot, future new quantity) for base station traffic semantic calibration, but only one hour of this time granularity analysis semantic feature, in order to adapt to the second Level traffic semantic calibration requires the extraction of more elaborate, multi-granular semantic calibration features.

A method for dividing a scenic area of a base station of a mobile communication network considering geographical conditions and boundaries. In this paper, the location characteristics and characteristics of the base station are used to divide the scenic area by clustering method. In order to truly reflect the role of mobile communication network data in the scenic area division, the data value is reflected, and the urban geographical conditions and administrative boundary conditions such as rivers and highways are not included. To a traffic cell division method based on a mobile communication network base station. The next step is to reflect these physical conditions into the partitioning method. Excavate the role of scenic road semantic base stations in the acquisition of traffic road network information. In this paper, the scenic road base station is divided into a class of traffic semantics, and a base station-based road network information detection system is constructed. The next step is to calibrate the secondary road semantics of the main road such as expressway and trunk road to obtain the movement speed of the road network user and identify the traffic. Way, determine the traffic status of the scenic spot, etc.

4. People move time and space situation

The time and space situation of the crowd movement is to analyze the concentration of residents in different regions at different times within the city area. The crowd movement macroscopically reflects the traffic occurrence and attraction of the city, and uses the data of mobile communication network to analyze the temporal and spatial variation characteristics and evolution rules of urban population movement, and grasp the pulse of urban traffic travel macroscopically. This paper uses SupenIl's printing software and Oracle database to build a GIS database platform, analyzes the real-time mobile aggregation of residents under the coverage of mobile communication network on the network, and draws the time and space change map of the crowd movement.

Specific steps are as follows:

(1) The original mobile phone data is a record table per minute, which records the interaction information between the user's mobile phone and the communication network. Use the 0IIacle database platform to extract the time and space change information of the crowd movement, and convert the record data into

The amount of tourists covered by each base station at each moment. (2) Convert the amount of base station coverage users into the number of scenic users. (3) Import the base station coverage information table into the Supem printing GIS software, draw the base station point data, and create the thematic map of the base station coverage user volume.

(4) Make a special map of the base station scenic spot visitors into a crowd distribution cloud map.

After 7 o'clock, tourists began to arrive at the scenic spot. After 18 o'clock, during the off-hours, the scenic spot will be closed, and the color of the scenic spot will gradually fade. The road network is the basic skeleton of the scenic spot. The traffic situation analysis of the road network is to accurately calculate the traffic system behavior of the scenic spot, scientifically formulate the regional management decision of the scenic spot, and give full play to the potential of the scenic spot facilities. The basics. This paper proposes a base station that uses the calibrated scenic road semantics to collect real-time crowd traffic of the road network and identify the theoretical thinking of the road network traffic situation. Based on the measured base station's hourly inflow (hlnow) and hourly outflow (outnow) data, the road network load is evaluated, and the spatial and temporal distribution and spatial-temporal evolution characteristics of the road network crowd situation are analyzed qualitatively and quantitatively.

5. Conclusion

The scientific requirements of urban and rural planning in the new era are getting more and more important. It is more and more important to “speak with data and let data speak”. This is also the requirement of the “big data” era, that is, to replace or make up for traditional technical means through massive data analysis. Master the rules of people's travel activities. Massive mobile phone signaling data analysis can effectively reflect the movement rules of people, provide data support for urban planning and traffic planning, and the application prospect is very broad. The mobile phone signaling data collection and processing method was explored, and combined with urban and rural planning practice, the application of mobile phone signaling data was demonstrated from the perspective of OD relationship between the city area and the district administrative district, and a wide range of regional passenger flow distribution features were realized. In the future, it is necessary to dig deeper into the application of mobile phone signaling data analysis in other aspects of planning management, such as regional planning, traffic special planning, and important facility layout planning. In addition, it is also very important to innovate mobile phone signaling acquisition methods to improve data quality and sampling frequency, increase the collection of 3G and 4G user data, and increase the collection of mobile phone network signaling data.

Acknowledgments

This work was partially supported by Science and Technology Innovation Program for College Students (No. 201811488014).

References

[1]CALABRESE F, SEVTSUKA, RATI'T C. Cellular census: explorations in urban data collection[J]. IEEE Pervasive Computing, 2007, 6(3): 30-38

[2]CALABRESE F. Real-time urban monitoring using cell phones: a case study in Rome[J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(1):141-151.

[3]Yang Xu, Shih-Lung Shaw, Ziliang Zhao. Understanding aggregate human mobility patterns using passive mobile phone location data - A home-based approach[J]. Transportation, 2015(42): 625-646.

[4]Shan Jiang, Joseph Ferreira, Jr., and Marta C. Gonzalez. Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore. IEEE TRANSACTIONS ON BIG DATA, 2017(3):1-7