Analysis of Urban Residents' Travel Demand Based on Multiple Data

Xiao-min QIN
Beijing Jiaotong University No. 3 Shangyuancun, China

Keywords: Didi GPS data, Bus IC card and GPS data, OD, Bus travel demand.

Abstract. In the areas where public transport system is undeveloped, IC card data can’t fully represent the residents' demand, and affect the analysis of urban residents' demand. The clustering analysis theory is used to analyze the travel original and destination after eliminate the invalid and redundant data. Reappear the original and destination on the GIS map, integrate the data and draw desire lines. This paper discusses the comprehensiveness of bus travel demand analysis based on bus and Didi data.

Introduction
Taxi traffic is a useful supplement to the bus way. It is a main transportation tool for commuter travel and non-regular travel. It's a public transportation service mode [1]. In general, the analysis of travel characteristics mainly includes the analysis of travel basic characteristics, the characteristics of travel space distribution and the analysis of travel time distribution characteristics. Domestic scholars mainly use GPS data to estimate road status, study travel behavior, guide traffic planning and establish taxi dispatching system. Yang Gege and others, studied the OD temporal and spatial distribution characteristics of external traffic hub in urban, based on the Didi GPS data in Beijing [2]. Tong Xiaojun and others use Didi GPS data to analyze residents' travel behavior in two cases, working days and rest days [3]. Ma Yunfei takes the track point data collected from 620 GPS equipped devices in Kunshan city of Jiangsu province as the research object, and analyzes the hot spots and temporal and spatial characteristics of residents' trips [4]. The optimization method of urban bus line based on GPS data of Didi operation is studied by Ba Xing Qiang and others [5]. In view of the shortcomings of the existing taxi OD survey, Zhao Lei and others proposed an algorithm based on GPS and GIS system, which can realize the calculation of the taxi OD matrix [6]. In foreign countries, the matching algorithm of GPS location data and electronic map is mainly studied by using GPS data. Taxi GPS data is also used to estimate OD, predict road traffic state, and identify travel origin and destination. GPS data mining is also an important research direction. Joe Grengs, Xiaoguang Wang and Lidia Kostyniuk and other scholars have transformed GPS data into a database, so that traffic engineers and planners can better understand drivers' driving behavior. In this paper, in order to make up for the lack of the IC card data in the area where the public transportation system is poor, we extract OD data from Didi GPS data and IC card, analyzing the urban bus residents' travel demand.

Data Processing
Preprocessing Didi Data
The collected GPS positioning data is from Cangzhou's GPS equipment installed in Didi. The vehicle reported real-time information about vehicle’s equipment status, operation status, geographic location, instantaneous speed and operation azimuth at 0.5-0.5min intervals. It is necessary to preprocess the data for the returned GPS data have some errors due to satellite positioning, atmosphere, operation error and so on. The research scope is within three expressways in Cangzhou, and the latitude and longitude of the scope is 38° 24'54.83”north-38° 14'3.92”north and 116° 44'29.35”east-117° 0'21.58” east queried on Google Earth software. The latitude and longitude outside this area are excluded from this data. The GPS data displays 0 means on load, displays 1 indicates passengers. If the GPS data displays other numbers beyond 0 and 1, the data
should be removed. Delete duplicate records or incomplete data; delete erroneous data, such as no longitude and latitude data in the longitude and latitude column, and delete obviously unreasonable data, such as boarding time later than the time of getting off.

Identification of Origin and Destination

The key point of the analysis of urban residents' travel demand in this paper is the identification of the origin and destination of the passengers. Realized identifying the origin and destination of the residents by programming. The principles are as follows:

Transform the GPS data used Oracle database into a data format that can be read by Excel.

1) Travel origin identification

The origin of the trip is the passenger status from 0 to 1, take the middle time of the passenger's state of 0 and 1 as the start time, take the middle latitude and longitude of the passenger's state of 0 and 1 as the start point of the trip. Identified in the following order:

a. The upper and lower two lines of data are numbered with the same number;

b. When the time difference between the two data with the same number is large, the data should be removed;

c. Using the lower passenger state minus the passenger status of the previous row, if the return value is 1, the intermediate time of the two row data is taken as the departure time of the origin, and the middle longitude and latitude of the two row data is taken as the location of the origin.

2) Travel destination identification

The destination of the trip is the passenger status from 1 to 0, take the middle time of the passenger's state of 1 and 0 as the destination time, take the middle latitude and longitude of the passenger's state of 1 and 0 as the destination of the trip. Identified in the following order:

a. The upper and lower two lines of data are numbered with the same number;

b. When the time difference between the two data with the same number is large, the data should be removed;

c. Using the lower passenger state minus the passenger status of the previous row, if the return value is -1, the intermediate time of the two row data is taken as the departure time of the destination, and the middle longitude and latitude of the two row data is taken as the location of the destination.

Preprocessing Bus IC Card Data and GPS Data

1) Preprocessing bus IC data and GPS data

According to the line number, bus number to analyse IC card data, the data lower than a certain threshold data is deleted; The data that deviate from the downtown area of Cangzhou is excluded; Delete the data that the speed is non-zero in GPS data; keep the data within the 90m range from the site location; match The GPS data with the IC card data and the data that is not matched is deleted.

2) Travel origin identification

Match the bus line site with GPS data: By matching the longitude and latitude of GPS data and bus site, obtained the bus site number corresponding to each GPS data, match the GPS data (with the information of the bus station) with the time of the bus IC card, thus obtain the aboard site of the bus corresponding to the bus IC card.

3) Travel destination identification

Cangzhou bus adopts a single ticket system, and the record of the card does not contain information about the destination. The travel destination identification is based on the origin data, derive by the general rule of urban bus passenger travel. In this paper, the attraction right method of bus site is used to deduce the bus destination site. On the one hand, after recognizing the aboard site of the IC card data, we can count the traffic on the bus in a certain period of time. On the other hand, the attraction strength of a bus station can be reflected by the traffic volume of the site. Therefore, the attraction method based on the attraction strength of bus stops can be used to deduce the get off stations[8][9].

The statistical study shows that the number of passenger travel stations is close to the Poisson distribution:
The attraction coefficient of the bus station $i$ on this line; $s_i$: The overall number of passengers aboard the bus station $I$; $n$: The number of sites in a single direction of this line.

The attraction of the site is expressed by the ratio of the number of passengers aboard the bus or the number of people who drop out of the bus to the total number of aboard or drop out passengers on each site in this direction. In this paper, we use the aboard station of IC card data, count the number of aboard people on every site and calculate the site attraction of each bus site:

$$w_i = \frac{s_i}{\sum_{p=1}^{n} s_p}$$  \hspace{1cm} (1)

$p_{ij}$: The number of passengers getting off from station $i$ to site $j$.

The probability of aboard on site $i$ and get off on site $j$ is affected by the number of sites and the attraction right of a site, Therefore, $p_{ij}$: the probability of aboard on site $i$ and get off on site $j$ can be expressed as:

$$p_{ij} = \begin{cases} \frac{p_{ij}w_i}{\sum_{p=1}^{n} p_{ip}w_i}, & i < j \\ 0, & i \geq j \end{cases}$$  \hspace{1cm} (2)

$X_j$: The number of passengers get off on site $j$.

The quantity of people that get off on site $j$ can be represented as the product of the total number of passengers aboard the bus at each sites before the $j$ site and the probability of getting off on site $j$:

$$X_j = \sum_{i=1}^{j-1} s_i \cdot p_{ij}$$  \hspace{1cm} (3)

4) bus passenger transfer analysis

The bus passenger transfer analysis is carried out in the following steps:

Step 1: Sort the IC card data that matched the aboard and down sites according the card number and swing card time. Operate The adjacent data records of the same card number.

Step 2: Distance analysis between transfer stations: according to the aboard site, choose the nearest site from the previous line on the same card number. Set the distance threshold to determine whether the station is passenger’s getting off station. If the distance between transfer stations is greater than 1km, it is considered a non-transfer travel. If the distance between transfer stations is less than 1km, then the turn step three;

Step 3: The time difference between two swing time: According to the service level of Cangzhou public transport system, the set time interval is 30min. If the time interval between two adjacent card swims is greater than 30min, it is considered a non-transfer trip. If the interval between two swipe cards is less than 30min, it is considered to be a transfer trip.

Step 4: Judge all records of the same card number and remove the middle transfer record and reserve the origin and destination record.

Divide the Traffic Area According the Origin and Destination Data

The K-means clustering method of SPSS software is used to cluster the origin and destination of bus and Didi. Integrating with the dividing principle, divide the traffic zone. Then import the traffic zone boundary diagram to the GIS map that has cluster data of the origin and destination. Calculated the number of the same attribute points in each traffic zone by GIS software. The OD matrix of residents' public transportation and Didi trips is obtained respectively. The output and attraction of public transportation and Didi trip in each traffic zone are calculated.
Analysis of Travel Space Characteristics

Affected by the coverage of public transport system and the information collected by bus routes, some communities will be short of bus travel data. Data volume in traffic area will further affect data enlargement. The Didi data can make up for the lack of the bus data and provide more comprehensive guidance for the bus system planning because of the high correlation between the Didi and the bus transport purpose.

Comparative Analysis of Public Transport Data and Didi Data in Traffic Zones

![Figure 1. Comparison of bus and Didi traffic volume in traffic zones.](image)

The population of Cangzhou is concentrated in the surrounding areas of Ying Bin Road, Hai He Road, Chang Lu Road and Yong Ji Road, that is, 3, 5, 6, 4, and 7 traffic zones, and their travel demand is also large. From Figure 1, we know that the bus traffic volume of residents in 6 traffic zone is far less than that of the Didi trip volume. The main reason is that the traffic zone 6 is composed of Fu Yang Avenue, Jie Fang Road, Traffic Street, Nan Chuan Road, Railway West Street and Hai He road, the canal belongs to it. Because of the barrier between the canal group and the Xinhua Group, the bus lines and bus stations are relatively less, and passengers turn to Didi trips when the bus travel cannot meet their needs.

OD Distribution Characteristics

The bus OD analysis is to get a comprehensive understanding of the passengers' daily travel conditions. Through detailed and detailed statistical analysis, the comprehensive information of the residents' travel is obtained. It provides a scientific basis for predicting the state of the future urban public transportation and evaluating the urban and rural public transportation planning. The most important purpose of OD survey is to grasp the spatial distribution of passengers' travel. The spatial distribution of trip volume is also known as the origin and destination distribution of trip volume, or the trip distribution of desire line. It can clearly show how many travel needs will be produced during the investigation time from a traffic zone to other traffic zones, thus helping to grasp the direction of the main passenger flow in the city.
Figure 2 shows the travel desire line of bus and Didi OD data in Cangzhou downtown, the OD matrix data in Transcad can be displayed in the form of desire line diagram, the desire line is a straight line connecting two traffic zones, and the line’s thickness represents the size of the travel amount between the traffic zones.

![Desire Line Diagram](image)

Figure 2. The desire line of peak hour travel in Cangzhou downtown area.

From figure 3, we can see that: first, on the interior of the road, on the Yongan Avenue the number of passengers that aboard and down bus and Didi is the most. Yongan Avenue is located inside the canal group. It is the main road of south-north of the downtown area of Cangzhou. There are more residential areas on both sides of Yongan Avenue with Cangzhou Career Academy, fourteenth middle school, etc. on both sides of the road. So there is more generation and attraction. At the same time Yongan Avenue intersects with Yongji road which is the section of S331 in Cangzhou area. So Yongan Avenue also share part of the south-north traffic that from S331 to south-north traffic in Cangzhou. The Yellow River road rank second, it link up the western group, the canal group, the Xinhua Group and the eastern group. And it meets the requirement to build as the main road. Second, in terms of the exchange between different sections, it can be found that area enclosed by Yongan Avenue, Yongji Road, the Yellow River road and Qian Tong Road have a large amount of travel and a large part of the main traffic has been undertook.

Through the above analysis, it is found that the service level of 6 traffic zone cannot meet the needs of the residents, and the service level of the bus should be improved, which is mainly to improve the public transportation of the Fuyang street and the Jiaotong street.

Summary

The paper describes the process of data handling of Didi GPS and bus IC card data, GPS data. Take Cangzhou for example to analyze the travel demand that based on Didi and bus data. By comparing the bus OD with the Didi OD data, it is easier to find the weak bus service area, and can reflect the demand of the bus residents more comprehensively. The combination of macro and micro provides a more comprehensive guidance for public transport planning.
Reference


