Research on Bus Passenger Flow Analysis Method Based on Intelligent Transportation

Zhouhu Xie
School of transportation, Beijing Jiaotong University, Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Beijing, China
531017636@qq.com

Abstract. In recent years, with the rapid development of intelligent transportation, IC card data of passengers can be obtained through on-board instruments, and passenger flow information of bus stations can be obtained by using computers without manual investigation. However, smart bus technology has not been fully popularized in many small and medium-sized cities, so it needs to be combined with other data for calculation and analysis. How to use multi-source data to comprehensively grasp the current situation of public transport passenger flow information, so as to meet the needs of public transport scheduling management and network optimization, has certain research value. Based on the new method of bus passenger flow data collection, this paper calculates the bus IC card holders' upstream and downstream stops.

1. Introduction
In recent years, many cities in China have been supporting the popularization and application of intelligent public transportation in Daxing. On April 28, 2014, the Ministry of Communications issued a document deciding to speed up the construction of the second batch of demonstration projects for intelligent application of urban public transportation nationwide. Among them, it explicitly mentioned supporting the construction of demonstration projects for intelligent application of urban public transportation in 26 cities nationwide, including Taiyuan, Shijiazhuang, Hangzhou, Ningbo and Shanghai. Due to the rapid progress and application of intelligent public transportation system in the field of transportation, making good use of GPS system and IC card information has become a research hotspot. More and more cities across the country will have the feasibility of collecting data on passengers' swipe cards and vehicle arrival and departure times, providing a wider application space for multi-source information for public transport transfer identification.

The bus IC card has massive data, high reliability and good availability. However, in many cities such as A city, there is no bus stop for passengers in the card swiping information stored in IC cards, so the actual bus boarding and alighting positions of passengers cannot be correctly calculated. Therefore, this aspect has very important research value and is helpful to fully mine and utilize multi-source data, thus further meeting the needs of public transportation dispatching management and network optimization.

Due to the rapid progress and application of modern intelligent systems in the field of transportation. More and more scholars are doing in-depth research on how to make good use of these
modern technologies. It will also attract the attention of experts in the field of traffic flow in the next few years, which will also bring opportunities for traffic development in the era of big data. At present, the research on how to use multi-source data to analyze the law of bus passenger flow will be a hot topic in the future, which needs further exploration and research.

Xiao Dai (2005) made classified statistics of bus CARDS according to passengers' swiping time, so as to judge the average time points in the middle of the bus traveling between stops, and then paired them to judge the terminal position of the bus [1]. Weixiong Cha (2010) studied the method of using Matlab to conduct statistical analysis of the information recorded by boarding the bus and swiping the card at each time period [2]. Zhang yu (2010) established the application platform of bus IC card data mining with SPSS [3]. Xiangguo Wu (2011) put forward the bus OD matrix derivation based on the existing IC card data and GPS data analysis method combined with other bus information [4]. XiangjieCong (2016) analyzed the derivation of bus boarding station of shenzhen B706 bus passengers based on the algorithm of location matching and time matching [5]. Bo Wang (2018) Using Public Transport System to Identify Passenger Flow in and out of Stations of Chengdu Metro Lines 1, 2 and 4 and Transfer Volume between Subway and Conventional Roads Based on Bus IC Card Data [6].

To sum up, due to the rapid progress and application of modern intelligent system in the field of transportation. More and more scholars have made in-depth studies on how to make good use of these modern technologies. It will also attract the attention of traffic and passenger flow experts in the next few years, which also brings opportunities for the development of transportation in the era of big data.

At present, how to use multi-source data to analyze the law of bus passenger flow will be a hot spot in the future, which needs further exploration and research.

2. Data structure
J Public Transport Group has equipped GPS surveillance equipment for all operating vehicles and installed an IC card swiping device to store vehicle running track data and IC card swiping data in real time. These automatically collected basic data on bus operations provide an opportunity for innovative bus surveys.

Through the investigation of the passenger load of motor vehicles, non-motor vehicles and buses in the cross-river section of the selected 10 urban areas, the passenger volume of public transport in the urban area is grasped; passengers (500 samples) and urban and rural areas are passed through the urban area. The survey of bus passengers (700 samples) and external hub passengers (400 copies) was conducted to understand the travel characteristics of public transport passengers in A city City and their satisfaction with public transportation. These surveys provide scientific and rational plans for the urban public transport in A city City. Planning provides basic information.

2.1. IC card data
Bus IC card credit card data records the detailed information of each passenger's card every time. The data sheet consists of card chip number, card type, credit card time, vehicle company, line number, vehicle number, etc. More than 40,000 records (every passenger's information for each card is a record). Due to confidentiality requirements, part of the data has undergone confidentiality conversion processing.

The calculation results are shown in Table 1.
Table 1. IC card data.

| Card number | Card type | Using time       | Company | Line |
|-------------|-----------|------------------|---------|------|
| 5           | 0         | 12/6/2016 7:02:36| 2       | 9    |
| 5           | 0         | 12/6/2016 10:54:41| 2       | 9    |
| 5           | 0         | 12/6/2016 14:31:25| 2       | 9    |
| 6           | 0         | 12/6/2016 5:29:30 | 2       | 9    |
| 7           | 0         | 12/6/2016 10:21:28| 3       | 9    |

2.2. GIS data
The bus data obtained after processing is shown in Table 2.

Table 2. GIS data.

| Car number | Number plate | Longitude   | Latitude   | Time          | Line |
|------------|--------------|-------------|------------|---------------|------|
| 46         | 13           | 119.645290  | 29.093935  | 2016/6/12    | 9    |
| 46         | 13           | 119.644298  | 29.093347  | 2016/6/12    | 9    |
| 46         | 13           | 119.643268  | 29.092867  | 2016/6/12    | 9    |

2.3. GPS data
For further analysis of subsequent algorithms, the project team further processed the bus GPS data sheet (1,100,000 records in total) and obtained a vehicle-to-station information table (about 100,000 records). As shown in Table 3.

Table 3. GPS data.

| Number plate | Line | Station | Arrival time | Departure time |
|--------------|------|---------|--------------|----------------|
| 13           | 9    | 1       | 8:08:42      | 8:15:52        |
| 13           | 9    | 3       | 8:16:30      | 8:16:22        |
| 13           | 9    | 3       | 8:17:10      | 8:18:12        |
| 13           | 9    | 3       | 8:18:30      | 8:19:42        |
| 13           | 9    | 3       | 8:20:28      | 8:21:32        |

This time, a total of 200 buses were covered in all-day departure data covering 67 routes. The line sampling rate was 80.6% (92.0% in urban areas and 68.4% in urban and rural areas). Passenger flow accounts for only 1.9% of the total passenger flow throughout the day. It can be said that the city’s bus lines are fully covered.

3. Bus passenger flow data analysis algorithm
The data analysis process can be said to be the most important part of the entire data processing. It takes a lot of complicated algorithms to set up the process [7].

The first step: Match the credit card information of the IC card with the GPS information table through the time key segment to obtain the relationship table between the credit card information and latitude and longitude.
The second step: Match the first-step result with the line vehicle information table through the vehicle number field to obtain the relationship table between the card information, latitude, longitude, and line name.

Step 3: Match the results of the second step with the line and site information tables by using the latitude and longitude key fields to obtain a relationship table of credit card information, latitude, longitude, site name, and line name.

Part 4: Utilize the train of thought based on the travel chain, that is, assuming that the card is swiped n times, judging the order of getting on and off the passengers in order of priority, and changing the flow of passengers.

![Passenger flow data analysis flow](image)

**Figure 1.** Passenger flow data analysis flow.

The results of the analysis can be viewed by the user, and can also be stored and recalled, and can be analyzed or compared again in the future, and can also be used as a basis for making decisions.

3.1. *Get on the station calculation method*

Through the establishment of IC card data sheets with line number bus to stand away from the relation table data, find the charge time inside certain limits (the project for the positioning way station 50 meters radius, at the end of the first site is 150m range) in GPS coordinate information, combined with theory-oriented and GIS site coordinates, matching to a line of site location, ultimately determine the credit card on site.

3.2. *Get off the station calculation method*

Taking advantage of residents' travel is often a feature of a closed travel chain. If a resident regularly chooses a bus route, his demand for the route will usually be two stops, and the previous bus stop and the next bus stop must be similar or the same. According to the judge of the starting station of the next trip, the nearest station from the previous trip to the starting station can be found through GPS, which is considered as the end point of the previous trip. In general, the travel path of passengers is basically unchanged in a certain period of time, and more travel data can be used to help judge the travel destination of passengers.

However, the means of transportation selected by passengers can be more than just public transport. For example, they can take taxis or subways. The modern technologies we use cannot capture this trip
information. Looking at the trajectory, there will be a line that is not completely closed. However, according to the law of passengers travel, passengers travel routes, time, and methods are often fixed. Therefore, only hoofs and more days of credit card records are needed to find out records that have similar travel schedules on other dates and continue to travel on a bus-based basis.

Based on the two situations mentioned above, we have summarized the following judgment methods: for passengers who meet the discontinuous travel bus chain and transfer at different stations, there will be other travel modes between bus transfers, which can be judged according to the nearest distance between two adjacent bus lines. Since the setting of urban bus stops is mostly between 300 meters and 500 meters, when the difference between the bus lines recorded by two adjacent swipe cards is more than 500 meters, passengers can only use a non-bus mode to reach the next transfer point.

The calculation results are shown in Table 4.

### Table 4. Site calculation result.

| Card number | Line | Vehicle number | Boarding site | Take a ride | Boarding time | Drop off site |
|-------------|------|----------------|---------------|-------------|---------------|---------------|
| 1           | 9    | 46             | Yara          | 6:49:20     | 7:08:38      | Baifuyuan     |
| 1           | 9    | 46             | West alley    | 7:08:38     | 6:49:20      | Baifuyuan     |
| 1           | 9    | 46             | Baifuyuan     | 13:00:14    | 13:46:51     | Hengde Auto Park |
| 1           | K1   | 45             | Fuhua food    | 13:00:14    | 13:46:51     | Baifuyuan     |
| 1           | K1   | 45             | Three mile shop | 9:07:04 | 6:49:20 | Flashstar Stainless Steel |
| 1           | K1   | 45             | Shuangfeng avenue | 12:44:23 | 6:49:20 | Jin’an Industrial Park |

#### 3.3. Site traffic

Site passenger flow refers to the number of passengers arriving and departing at all the bus lines of a certain site within a unit time, which reflects the bus passenger flow demand status of the site and is an important reference for the bus station reconstruction. Select several sites with large passenger flow and representative sites. Calculate and judge on the basis of statistical processing of bus information, and get the number of passengers getting on and off the bus. Then use the name of the site as a key to separate counts, and finally count the number of passengers per station based on the card swipe rate of the passengers.

Through the data, we can calculate the total passenger flow of the node, and the settlement result is shown as follows.

### Table 5. Site calculation result.

| Line | Departure interval | Length (km) | Total passenger flow | Full day capacity | The average by distance |
|------|--------------------|-------------|----------------------|-------------------|------------------------|
| 1    | 10—15              | 16.9        | 3659                 | 51.9%             | 6.3                    |
| 2    | 10—15              | 20.3        | 234                  | 25.7%             | 9                      |
| 3    | 5—12               | 15.3        | 3678                 | 64.3%             | 4.67                   |
| 4    | 10—15              | 10          | 1400                 | 46.6%             | 6.78                   |
| 5    | 10—15              | 14.2        | 1500                 | 53.6%             | 5.34                   |
| 6    | 10—15              | 8.1         | 457                  | 35.5%             | 4.34                   |
| 7    | 8—20               | 10.6        | 3478                 | 38.3%             | 5.67                   |
| 8    | 8—20               | 10.6        | 467                  | 46.9%             | 5.67                   |
| 9    | 8—20               | 10.6        | 532                  | 49.5%             | 5.67                   |
| K1   | 10—20              | 14.2        | 389                  | 54.2%             | 8.97                   |
| K2   | 10—30              | 14.3        | 570                  | 73.8%             | 10.4                   |
| K3   | 10—15              | 16.9        | 425                  | 51.9%             | 6.4                    |
Thus, the cross-sectional passenger flow of A city at different points in time and space can be obtained, and then the peak flow of nodes can be obtained. Planners can master the main passenger flow distribution points and bus flow corridors through big data, so as to provide reference for the following bus line network and station planning.

4. Conclusion
In many small and medium-sized cities, the construction of intelligent transportation system has just started. Due to the technical problems of equipment level, the bus data cannot accurately count the number of people on and off the bus at the station, so it needs to be processed and calculated in a unified way with other kinds of data. The examples in this paper can provide technical means and methods for the application of big data and public transportation planning and design, and provide theoretical support for the method innovation of transportation planning in line with the development of the times.

- In the collection of multi-source data for public transport, the different urban public transport IC card penetration rates and the degree of intelligence in the public transportation system are different, which will affect the reliability and comprehensiveness of the data, and errors need to be made in response to factors affecting data reliability. the study.

- In the judgment of the bus stop, the broken bus transit chain will often appear. Only relying on experience to judge the stop site based on the distance and time interval information, the error will be large, and further improvement and research are needed.

- China's information technology is still in a period of rapid development. Research on the use of modern information technology to collect and analyze bus passenger flow data is still in the early stages of exploration, and the way in which bus passenger flow data is mastered will continue to evolve. The research in this paper is still in the preliminary stage. The ideas and processes in data collection and analysis have yet to be simplified and further refined.

References
[1] Dai Xiao, Chen Xuewu. “Analysis and Processing Method of IC Card Data for Single Bus Line,” Urban Transportation, vol. 4, 2005, pp. 77-80, doi:10.13813/j.cn11-5141/u.2005.04.016.
[2] Cha Weixiong, Liu Jinxing. “IC Card-swiping Counting Algorithm for Bus Stations and Its Matlab Implementation.” Traffic Information and Safety, vol. 2, 2010, pp. 93-97.
[3] Zhang Yu. Research on OD Acquisition Method of Dynamic Bus Passenger Flow Based on Data Mining. Beijing Jiaotong University, 2010.
[4] Wu Xiangguo. Derivation and Application of OD Matrix of Residents' Public Transportation Based on IC Card and GPS Data. Shandong University, 2011.
[5] Guan Nana, Wang Bo. “Analysis on Passenger Flow Characteristics of Chengdu Metro and Conventional Road Public Transport Transfer Based on Bus IC Card Data,” Urban Rail Transit Research, vol. 9, 2018, pp: 148-150. doi:10.16037/j.1007-869x.2018.09.038.
[6] Cong Xiaojie. “Bus Passenger Flow Analysis Based on IC Card and GPS Big Data,” urban planning society of china Urban Traffic Planning Academic Committee. Collaborative Development and Traffic Practice-Proceedings of 2015 China Urban Traffic Planning Annual Meeting and 28th Academic Symposium. Urban planning society of china Urban Traffic Planning Academic Committee: Urban Traffic Professional Research Institute of China Urban Planning and Design Institute, May. 2015.
[7] Zhou Rui. Bus Station Passenger Flow Calculation Method Based on IC Card Data. Beijing Jiaotong University, 2012.