Research and Simple Calculations on Public Transport Card Data

Yefan Yang*
School of Transportation, Wuhan University of Technology, Wuhan, China

*Corresponding author: yefanyang2021@whut.edu.cn

Abstract. Big data is one of the hottest topics nowadays. Under the background of smart city and intelligent public transport, a large number of scholars begin to use big data of public transport in urban transport planning. This paper, combined with dozens of the latest journal papers, combined with their own understanding, focuses on the development of bus IC card big data, application status and research prospects, forming a reading report.

Keywords: Bus IC card, Big data mining, Passenger flow analysis.

1. Introduction
Public transport is an important part of urban transport, and well-developed public transport has a positive effect on relieving the pressure of urban traffic and reducing air pollution and carbon emissions. However, long waiting times, overcrowding and inconvenience in changing to public transport during the morning and evening rush hours are some of the main factors that prevent people from choosing public transport. Access to information on bus passenger flow is fundamental to the operation of public transport, providing a reliable basis not only for the daily regulation of public transport, but also for the optimisation of the bus route network. Therefore, a reasonable data analysis and processing method for bus IC cards is an important way to establish an efficient urban transport system. In the face of this massive amount of public transport data, we need to process and analyse it effectively.

2. An overview of bus pass data research
In recent years, with the improvement of city bus information system, the popularisation of city bus IC cards, the promotion of mobile payment, the introduction of IC card payment preferences and preferential policies for interchange, etc., city bus IC card payment and mobile payment have become more convenient and cheaper, and passengers using IC cards account for an increasingly high proportion of all people travelling on buses. According to the China Urban Passenger Transport Development Report (2017), the number of bus passengers using the bus one-card accounts for about 50% of the total passenger traffic nationwide, and the swipe rate of the bus one-card in some domestic cities has exceeded 80%.

Every time a bus passenger uses an IC card to swipe a bus, the charging system records a piece of data, and by swiping the card throughout the day it is possible to obtain - a bus IC card passenger's bus system travel information throughout the day. The bus IC card system data generally includes the card
number, card date and time, the bus route and vehicle number (or card device number), the amount of consumption and other information. In recent years, the Ministry of Transport has issued the "Opinions on Comprehensively Deepening Transport Reform", which states that the Ministry of Transport will study and formulate a framework for the development of intelligent transport and the realization of a national network of ETC and public transport cards. The use of public transport cards is becoming more and more widespread.

In recent years institutions such as the Massachusetts Institute of Technology in the US, the University of Westminster in the UK, the Montreal Institute of Technology in Canada and Southeast University and Tongji University in China have started to conduct research on the potential uses of public transport IC card data. This section focuses on the content of data mining applications, the basic ideas, technical features, problems and current domestic research progress of the analysis methods based on the big data of public transport IC cards, taking into account the current situation of domestic and international research.

3. Public transport card data research application

3.1. Analysis of bus passenger flow projection based on bus IC card information

Based on the bus IC card data, the method of determining the passenger flow at bus hubs can be proposed. By combining the actual data, the temporal, dynamic and short-term characteristics of passenger flow at bus hubs are analysed to reveal the temporal and dynamic change patterns of passenger flow at bus hubs.

![Flow chart for determining passenger boarding points](image1)

**Figure. 1** Flow chart for determining passenger boarding points

The minimum running time $T_{min}$ for a vehicle between stops can be derived from the average speed of the bus and the stop spacing. It should be noted, however, that in real life not all passengers will choose the IC card payment method, so there is still a discrepancy between the boarding passenger flow obtained by IC card alone and the actual boarding passenger flow. Each bus route should also be analysed in relation to the actual IC card swipe brigade.

The analysis is carried out to calculate the downstream passenger flow using the formula.
Where, in equation: Bj is the number of drop-off passengers at city bus hub j; Si is the number of boarding passengers at city bus hub i; Pij is the probability of boarding at city bus hub i and getting off at city bus hub.

The main factors affecting the probability of passengers getting off at a bus hub in a city Pij; are the distance passengers travel by public transport, the transfer capacity of the bus hub, the attraction intensity of the bus hub and the nature of land use around the bus hub.

Therefore, the analysis of bus passenger flow projections needs to be combined with the use of a large amount of data obtained from preliminary surveys, based on which a comprehensive analysis is carried out.

3.2. Calculation of bus full load index based on bus IC card information

Through MYSQL database software, the collected bus IC card information (including bus IC card number, card swipe date and time, consumption amount, fare amount, line number, vehicle number, pick-up point, vehicle up and down, etc.), bus GPS information (including vehicle number, recording time, line number, bus number, latitude and longitude), bus station information (including station number, station name, line number, direction, station type, latitude and longitude, etc.), bus vehicle information (including vehicle number, route, fuel type, rated passenger capacity, time of purchase, vehicle value), respectively, to establish a database.

In addition to the passenger flow projection analysis based mainly on IC card data as described in 3.1, it is also possible to combine GPS data for more convenient passenger flow analysis. Based on the vehicle number and station number information in the bus IC card information table, the entire journey of each bus passing the starting station of the bus route is recorded as one trip, and the bus vehicle departure schedule can be integrated by the corresponding card swipe time records. Then, combined with the rated passenger capacity of each vehicle in the bus vehicle information, the total rated passenger capacity of the bus line during peak hours is obtained.

Firstly, the maximum number of passengers in the bus IC card section is expanded. Compare the total number of IC card swipers on the bus line with the bus company's statistics on the bus line passenger flow to find the ratio of the bus line passenger flow to the IC card passenger flow, and
expand the sample by this ratio. First of all, we will expand the sample for the largest passenger flow section in the bus IC card. Based on the actual passenger flow and the rated capacity of the expanded sample, the full capacity of a single bus line is calculated.

This method reduces manpower input, improves the accuracy of calculations, is more efficient in measurement and has better accuracy of indicator results than the traditional manual station-based full rate survey method. It has some value.

3.3. Estimation of passenger waiting time per capita based on bus IC card information and GPS.

First, the bus IC card data and GPS data were cleaned separately: to remove trajectory drift, data points with vehicle speed over 30 m/s calculated from the trajectory points were removed as anomalies to achieve cleaning of bus GPS data. If the time difference between the IC swipe time and the matched vehicle arrival time exceeds 15 min, it is marked as unmatched data. For unmatched swipe data caused by lost vehicle GPS data or damaged equipment, the Bayesian decision tree algorithm is used to infer the passenger boarding time and obtain the passenger's boarding time and stop. Data reliability was ensured and the coefficient of variation and the correlation coefficient between the sampled distribution and the true distribution were used to quantitatively evaluate the similarity between the passenger arrival intensity calculated using the data and the true arrival intensity to ensure data completeness.

The cleaned data was then used to calculate the intensity of passenger arrivals at each station in the city and to model the passenger arrival pattern dynamically in conjunction with the non-temporal Poisson process.

Finally, to ensure the simplicity of the algorithm and to avoid the process of solving curve integrals, the variation of passenger waiting time per capita with information such as electronic station signs and mobile phone arrival alerts.

This model solves the disadvantage that questionnaire and video capture methods are time consuming and do not reflect dynamically the changes in passenger arrival times.

4. Future research directions for public transport card data

As the process of intelligent and informative public transport construction accelerates, bus IC card swipe information will become more comprehensive and richer, providing a large amount of usable basic data for the analysis of public transport travel characteristics of urban residents. In the future, the data can be mined and the scope and field of application of data technology can be expanded to provide more accurate and comprehensive data analysis.

At the same time, there is still a portion of the population that is not part of the full public transport travel chain, which makes it difficult to calculate the travel characteristics of public transport passengers and the full load factor using only the majority of the population's travel data, and the scope of the calculation does not cover all public transport travelers. The smart bus card does not currently have a co-branding system like bank cards, and fares in many cities are only related to the length of the route, plus there is a great deal of uncertainty about public transport in cities. It is also necessary to strengthen public transport construction to obtain more comprehensive data with a full set of facilities for more accurate data analysis, so that the most appropriate decision can be made.

Bus card data is not just used alone to integrate, in conducting analysis, we use a large amount of GPS data, land data, data from bus stops bus vehicles and so on. In the future, we need to further study big data, build mathematical models, design intelligent platforms for automatic dispatching, optimisation and decision-making, synthesise and analyse various data to select the right model for calculation, and combine with other technologies in order to realise a smart city and intelligent transport in the era of big data.
References

[1] Ministry of Transport. Report on the development of urban passenger transport in China (2017) [M]. Beijing: People's Transport Publishing House Co., 2017: 37-39.

[2] Fei Ye. Research on the OD projection method of public transport travel based on IC card data [J]. Computer Application and Software, 2018,35(8):190-194.

[3] LU H,BURGE P,HEYWOOD C,et al. The impact of real-time information on passengers' value of bus waiting time [J]. Transportation Research Procedia, 2018(31):18-34.

[4] WANG Y,ZHANG D,HU L,et al. A data-driven and optimal bus scheduling model with time-dependent traffic and demand [J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(9): 2443-2452.

[5] Li Jun, Deng Hongping. Research on passenger travel classification based on bus IC card data [J]. Journal of Chongqing Jiaotong University (Natural Science Edition), 2016, 35 (6): 109-114.

[6] TU Y, YANG J. Analysis and forecast of passenger flow based on public transportation IC card and GPS data [C]// Computer Science and Network Technology (ICCSNT), 2016 5th International Conference on. USA: IEEE, 2016: 281-285.

[7] Li Ying, Weng Xiao Xiong. A transfer recognition method based on bus IC card and GPS data [J]. Journal of Guangxi University (Natural Science Edition), 2017, 42 (2): 579-586.

[8] Qin Zheng. Research on passenger boarding and alighting stations based on bus IC card and GPS data [J]. Western Transportation Science and Technology, 2017 (8): 115-119.

[9] Zuo Happy. Research on the prediction method and application of bus passenger flow at different time levels based on IC card data [D]. Nanjing: Southeast University, 2016.