

Transit OD Generation Based on Information Technology

LIBING CHI^{1,a}, JIAN WANG^{2,b}

¹School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin, China

² School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin, China

^a16725684@qq.com, ^bWang_jian@hit.edu.cn

Keywords: GPS data; IC card data; Transit OD matrix

Abstract: As developing in computer and information technology, data access and collection have become more and more convenient. In many cities' transit system of the world, transit vehicle GPS data and passenger IC card data could be provided with database. This paper focuses on how to use the passenger IC card data (only record once per trip) and transit vehicle GPS data to generate the transit OD matrix. With the characteristic of transit trips, The continuity of transit trips is defined. By this definition, this paper presents a search method to generate the transit OD matrix. The validity of this method has been tested in the modeling Zhengzhou city's comprehensive transportation system.

Introduction

How to use public transportation IC card data has done in many literatures [1-3]. Previous studies focused on the use of public transportation IC card data to get the demand on-site car; in addition, considering certain characteristics of the bus passenger flow to estimate approximate number of passengers alighting on stops. The gravity model is adopted to generate OD between bus stops with constraints on the number of passengers boarding and alighting. More complex approach was to develop a bi-level mathematical programming model, the upper-level issue is responding to least squares model, the lower-level is the problem of public transportation network equilibrium assignment [4]. Due to lack of bus GPS data, these methods gain the information from the bus IC card and application of a number of vehicles on the site, meanwhile applied to the bus OD estimation process.

With the development of information technology in recent years, more and more cities begin to collect the bus GPS data. Recently some studies pay attention to explore the basis of the bus GPS and bus IC card data fusion to generate a bus OD [5]. However, the assumptions lead to that the method can be applied only in a single bus line. How to effectively utilize large amounts of data in computers and IT development is still a very interesting research direction.

This paper focused on how to use public transportation IC card data (only the case of credit card on the bus) and bus GPS data to generate a bus OD. Different from previous studies, In this study, not only presenting a model to estimate the bus OD, but searching alighting stops is conducted by the bus travel characteristics of the credit card records with meeting certain conditions to reflect the public transport of bus travel the inherent law of the OD. Bus OD generated from the existing methods are mostly corresponding to a certain extent. The new method is to use the concept of continuity of bus travel "to reproduce" the number of on-site car and bus OD generated by the new method in the overall evaluation on the bus travel through the survey data. As the hypothesis conditions, most applications on a virtual network are to verify its effectiveness. This method has been successfully used in the traffic model of zhengzhou, and generated the article dynamic bus OD with more than ninety thousand of the GPS data and millions of bus card record.

Bus GPS and IC card data

This study is based on two basic database of bus GPS and bus IC, data record provided by the Zhengzhou Public Transport Company (May 18, 2010 all-day bus operators).

Bus GPS data can describe the detail time of each bus arriving in the stations etc. The data tables can provide the information comprising of company names, bus lines, the number of vehicles, operation directions, the station number, station name, arrival time. There are a total of 90 million records in a all-day data (per vehicle to reach each site for a record).

Tab.1 Example of GPS data for transit vehicles

Company	Line	Vehicle number	Uplink and downlink	Site number	Name of the station	Arrival time
The second company	62	3847	Uplink	1	Garden Road Liuzhuang	7:28:18
The second company	62	3847	Uplink	2	Lu gang xiao zhen	7:30:37
The second company	62	3847	Uplink	3	Tian Rong international building materials Port	7:32:12
The second company	62	3847	Uplink	4	New LiuLu garden road station	7:33:59

Public transportation IC card data can record detailed information of each riding on transit vehicle, and the information tables are consisted of six fields: the card number, card type, using card date, using card time, using bus line, the vehicle number. There are a total of 129 million records in a all-day (each passenger per ride card information for a record).

Tab.2 Example of transit IC data

Card Type	IC card number	Card time	Take the vehicle number	Line
Classic Card	268970105	7:29:12	3847	62
Classic Card	268970105	8:25:53	1955	61
Classic Card	268970105	19:37:52	1958	61
Classic Card	268970105	19:50:46	4483	9
Adult concession card	292659705	7:24:34	2614	41
Adult concession card	292659705	18:33:09	2615	41

The analysis of bus OD needs to match the bus GPS and IC card data on line, additionally using vehicles and taking time. After the inspection for matching, a total of 88 million bus GPS data and 118 million bus IC card data can be in exact match. The rest of the data is not directly used due to various reasons caused by no match between bus GPS or IC card.

Characters of resident's bus travel

The analysis of survey data on the bus characteristics of residents are from Zhengzhou City, June 2010 household travel characteristics survey. The principles of extraction of data sources are: (1) taking a person as a unit, the survey object (personal) has one trip at least in a day; (2) the survey object is completed in survey space, namely the first trip is from his/her home and the last trip destination is also his/her home. According to the above principles, there are a total of 8547 subjects extracted from survey data. Since the families surveyed are subjected to uniform distribution in space, we can approximately consider that the survey objects are similarly subjected to uniform distribution in space.

(1) The distribution of frequency of bus usage

Only 11.32% travelers have a bus trip within public transport survey in one day, and 68.71% travelers take two bus trips in one day. Analysis of residents of bus travel, the bus mode chosen by traveler is dominant role with 70.98% trips in all.

Tab.3 Travel mode structure for persons with transit trips

Travel modes	Passenger trips	Proportion
Bus	18945	70.98%
Other	7744	29.02%
Total	26689	100.00%

(2) The continuity of bus travel

The continuous characteristics with two bus trips in one day are: the next travel origin is generally the destination of this trip, namely the other transit modes are rarely used by passengers between two trips when to choose travel modes. Therefore, we can define the continuity of a bus trip: when travelers have many trips in a day, if the destinations of their next trip were the origins of this trip, this bus trip can be called the continuity trip. For the last trip in one day, if the destination of trip is the origin of this trip, the last trip is also the continuity trip.

According to the definition above, if there is only a bus trip in one day, this bus trip is non-continuous. The continuity trip is a proportion 89.79% with the statistics source. If we take out the trip with one travel process, the continuity trips can be 94.7%. In other words, the most bus trips are the continuity trips.

Tab.4 Rate of continuous transit trips

Bus travel classification	All bus travelers		Use the bus twice and more than twice the travel	
	Number of times	Proportion	Number of times	Proportion
Continuous bus travel	17010	89.79%	17010	94.7%
Non-continuity of bus travel	1935	10.21%	959	5.3%
Total	18945	100.00%	17969	100.0%

Analyzing the non-continuity of bus travel for travel using public transportation more than once, about 60% are with non-continuity trips. This indicates that the destination is closer to the travel origin of the non-continuity trip. In addition, taxi and car (take) modes chosen are often accompanied by bus non-continuity trips. This shows that there are no convenient bus lines until the next time.

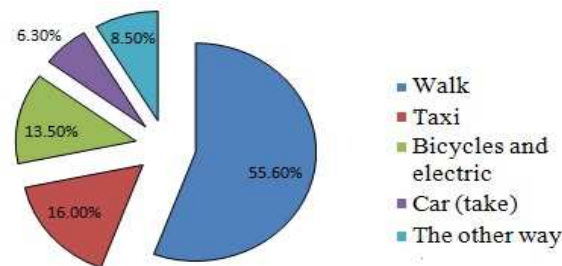


Fig.1 Non-transit travel mode composition for trips following the non-continuous transit trips

Search algorithm between O and D

(1) Typical bus trips

Using the concept of bus continuity trip, the records with many using in a card within IC card data on Zheng Zhou city network. The typical transit trip behavior can be summarized as the following three forms: (A) Travelers (morning); the first bus trip from the stop i , getting off at stop j ; (at night) the second bus trip from the j stop and get off at the stop i ; (B) Travelers (morning); the first bus travel in the stop i , transfer at stop j , and get off at n -stop, (at night) the second bus travel in the n -stop, transfer at stop j , and get off at the stop i ; (C) Travelers (morning) first bus travel in the stop i , get off at stop j , at the second bus travel in the j stop ... and get off near the n -stop, (at night) the most the first bus travel in the n -stop on the car, get off at the stop i . The typical transit trip behavior is provided in Figure 2.

(2) Algorithm description

Firstly, the records of same number cards should be extracted(including each card on the train line, direction, and site information) and sorted by charge time. Search every get-off records, and the search method: the next records site location as the target, the line get-on this time as the search direction, on-line follow-up site for the search site collection. When the shortest distance of a station in the follow-up site of this on the is found, the site is recorded as get-off site. When the card records the the last one, the target location was the location of the first sites charged.

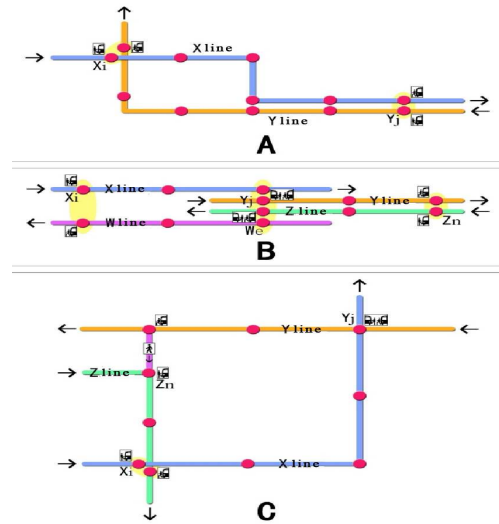


Fig2. Three typical situations for transit trips

In other words, when passengers who have twice and more records in the IC card c , the $n(n \geq 2)$ credit card records could be expressed as according to the time: $(l_1^x, l_2^y, l_3^z, \dots, l_n^w)$, where l_n^w is n times bus line of the w station, $w < \bar{l}_n$, \bar{l}_n is the largest serial number station (terminus) of the n times bus line. According to the above description, next credit card by the card location of the site is generally oriented goals when passengers continuously charged the card. Therefore, get off site $l_n^{i^0}$ can be calculated by next formulas:

$$l_1^{i^0} = \min_i \|l_1^{x+i} - l_2^y\|, (i = 0, 1, 2, \dots, \bar{l}_1 - x)$$

$$l_2^{i^0} = \min_i \|l_2^{y+i} - l_3^z\|, (i = 0, 1, 2, \dots, \bar{l}_2 - y)$$

$$l_n^{i^0} = \min_i \|l_n^{w+i} - l_1^x\|, (i = 0, 1, 2, \dots, \bar{l}_n - w)$$

Obviously, the search algorithm on every card records (including the last credit card records) can be found in the corresponding get off site. But not all the bus travel are continuous, so there will be searches that get-off site and get-on site is the same ($i = 0$). In this case, the OD between the card records stations could be regarded as invalid.

(3) Algorithm evaluation

The search algorithm showed good adaptability to be able to handle a variety of complex situations. Firstly, for the continuity of bus travel, the search algorithm can get accurate get-off site, typical bus travels cases (A) and (B) as shown in Figure 1. According to the proportion of the continuity of bus travel in the travel survey of households, the accuracy of the search algorithm can achieve about 95%. Secondly, for some of the non-continuity of bus travel, the search algorithm can be accurate to get off site. Figure 1, a typical bus travel (C) of the case, for example, when the passengers between the two bus trip mixed up with other ways, the search algorithm is still applicable. According to the characteristics of non-continuity of bus travel, the search algorithm could find a relatively accurate or relatively close method to get off site most of the non-continuity of bus travel. Finally, for some non-continuity of bus travel, get-off sites found through the search algorithm is invalid (as the situation that the get-on site and the get-off site are the same one). When this records are regarded as invalid, the bus station between the OD can get higher accuracy by the search algorithm.

The generation of the bus OD

The search algorithm can get most of the card records of the OD between bus stations, but it can't deal with only one credit card record and some non-continuous bus travel card records. As the large number data of the OD between bus stations get a high accuracy, OD between bus stations would be got through by loft on-site car (credit card) number for the whole sample.

Card records corresponding to get off site can also match its corresponding off time by bus GPS data. For many of the same card number records, the OD between bus stations could be transformed to the bus OD by the judgment of the transfer site. Because of bus GPS data, the transfer site definitions can be very precise, such as the combination of the two stations (site on the car and get off the site) the spatial distance, vehicular arrival and the time interval for the comprehensive judgment. As normal circumstances, the level peaks on the get off at the time of a passenger in a site before the next time in less than 25 minutes can be considered the site (ie, get off the site and the site on the car) for the transfer points. Peak within 30 minutes can be considered as a transfer site (due to the peak is very crowded, and passengers may not be squeezed into the vehicles to take the line first to reach). When a site has been identified as a transfer site, the bus trip and the next one will be combined (to retain the previous site on the train and get off of after times site). All card records does not exist to meet the conditions of the transfer site, the OD between bus stations was converted into a bus OD. Different from bus station OD, bus OD expression is public transportation amount of spatial relations, rather than public transit passenger volume of spatial relations.

Tab.5 Results of transit OD generation

ID	Travel sequence	On the bus line	Uplink and downlink	On the train site number	On the train site time	Travel sequence	Get off line	Uplink and downlink	Get off the site number	Get off site time
1	1	62	0	1	7:28:18	2	61	0	31	8:29:05
2	3	61	1	5	19:38:17	4	9	1	24	20:17:46
3	5	22	0	12	7:27:20	5	22	0	12	8:16:13
4	6	22	1	1	19:52:49	6	22	1	1	20:30:00
5	7	41	0	5	7:22:29	7	41	0	5	8:06:04
6	8	41	1	15	18:32:30	8	41	1	15	19:03:26
7	9	130	0	11	9:27:26	10	88	1	9	10:31:02
8	11	210	0	12	16:06:06	11	210	0	12	16:36:21
9	12	95	1	2	18:08:49	12	95	1	2	18:43:07
10	13	77	1	28	19:31:39	13	77	1	28	19:36:04

Zhengzhou bus OD generated results in Table 5. In Table 5 with a record on the train line and get-off the line is inconsistent with the description of bus travel after the transfer. When getting on line and getting off line consistent (right now travel sequence time is also consistent), there is no change to bus travel. It can be got from Table 5, the bus OD obtained by the new method is dynamic.

Figure 3 is a bus company statistics IC card credit card traffic generated bus OD allocation in the model line statistical results of the comparison. It can be seen, each line card statistics of traffic and the bus company forecast results have a very high goodness of fit (0.97), which also illustrates the bus OD generated by the above method has a very high accuracy.

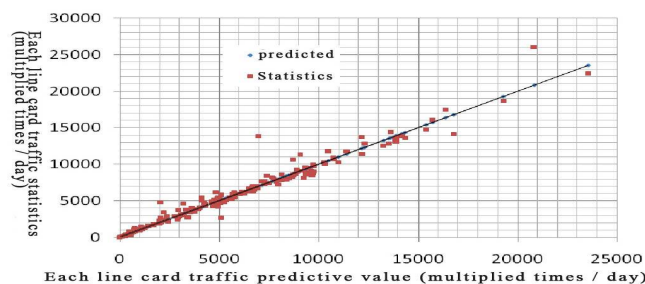


Fig3. Comparison of forecasted and reported line ridership

It should be noted that the above method is applicable to passengers in the IC card bus travel OD generated. A considerable portion of bus passenger is coin-operated ride the city bus. The features of bus travel card passengers (usually urban resident population as the main body) and the coin passengers (floating population) are significant difference. Therefore, the generated based on the IC card data bus OD should not loft to generate the overall bus OD.

In the case of the lack of coin-operated passenger, bus travel data can be assumed to constitute a card passengers urban resident population, and using the above-generated credit card passenger bus trip OD, bus service level measurements and land use data (population, employment, etc.) to calibrate

passenger transport model [6]. The flow of population distribution can be used to predict on the basis of the calibration of the passenger traffic model and the coin passenger bus travel OD (or floating population). Credit card passengers and coin-operated passenger bus trip OD are merged to generate the overall bus OD. Figure 4 shows the results for the overall bus OD in Zhengzhou city, assigned bus network.

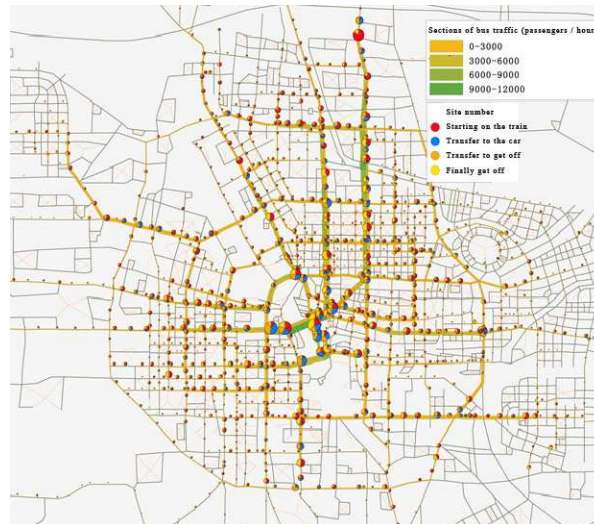


Fig4. Transit assignment results in AM peak period for Zhengzhou city

Obviously, the exact extent of the coin passenger bus travel OD have a certain impact on the quality of the overall transit trip OD, but the bus trip data is difficult to get through the survey. The future can be combined with sections of cross-section survey of passenger transport, back stepping.

5 Conclusions

This paper is based on the analysis of bus travel characteristics of the concept of continuity of bus travel. Based on the concept of continuity and bus system (bus GPS and IC card data), this paper constructs a search algorithm to generate the OD between the bus station, which to generate other types of bus OD. Different from previous studies is that the proposed method can be easily applied to the reality of the bus system; bus OD generated can reproduce the inherent law of bus travel. The bus OD generated should be noted as to the bus GPS and IC card data. It can accurately reflect the general urban population (or credit card passengers) bus travel distribution, and the floating population (or coin passengers) bus trip distribution in future is still needs further research.

References

- [1] Xue-Wu Chen, Dai Xiao, Chen Qian. Bus IC card information collection, analysis and application [J]. Civil Engineering, 2004, 37 (2): 105-110.
- [2] Shi Fumin. OD matrix construction method based on the IC card data bus. Journal of Jilin University [D], 2004, 11-39.
- [3] Zhou Tao, Zhai Changxu, Gao Zhigang. Based on the OD of the bus IC card data projection technology [J], the urban traffic, 2007, 5 (3): 48-52.
- [4] Zhou Jing, Zhang Lunke. using the IC card data to estimate the bus OD matrix model and algorithm . Theory and practice of systems engineering, 2006 (4): 130-135.
- [5] Xu Jianmin, Xiong Wenhua, You Feng. Based on GPS and IC card single-wire bus OD generation method [J]. Microcomputer Information, 2008, 22: 221-222.
- [6] Wu Zixiao, Ren Xifeng, Hu Jingyu. Based on the bus GPS and IC card data traffic modeling and new ideas [J], the urban traffic, 2011, 9 (1): 47-51.
- [7] INRO Consultants Inc. (2010). EMME 3 User's Manual.