

# Research on Methods of Processing Transit IC Card Information and Constructing Transit OD Matrix

Xiuhua Han<sup>1</sup>, Jin Li<sup>1</sup>, and Han Peng<sup>2</sup>

<sup>1</sup> Transportation School, Jilin University, 5988 Renmin Street, Changchun, China, 130022, hxh1952@163.com; li\_jin@jlu.edu.cn

<sup>2</sup> North China Institute of Water Conservancy and Hydroelectric Power, 36 Beihuan Road, Zhengzhou, China, 450011, penghan20@yahoo.com.cn

**Abstract.** Transit OD matrix is of vital importance when planning urban transit system. Traditional transit OD matrix constructing method needs a large range of spot check survey. It is expensive and needs long cycle time to process information. Recently transit IC card charging systems have been widely applied in big cities. Being processed reasonably, transit passenger information stored in IC card database can turn into information resource. It will reduce survey cost a lot. The concept of transit trip chain is put forward in this paper. According to the characteristics of closed transit trip chain, it discusses how to process IC card information and construct transit OD matrix. It also points out that urban transit information platform and data warehouse should be constructed, and how to integrate IC card information.

## 1 Introduction

Developing urban transit is of vital importance in releasing urban congestion. Transit preference rule is universally acknowledged by various urban managing sectors. Recently, information technology is widely applied in urban transit field. Transit IC card charging systems are adopted in many cities. As aiming at charging and reckoning statistics information, transit IC card information management system is typically designed for information storing and processing, so lots of transit passenger flow and its space-time distributing information in database is deposited and cannot be applied reasonably.

Transit passenger flow OD (Origin-Destination) matrix is an important basis of planning urban transit system[1]. Traditional method of constructing OD matrix needs a large range of passenger trip spot check survey or transit passenger flow survey to achieve relative basic data. It requires a lot of manpower and expenditure.

The data processing needs heavy workload and long time. It is costliness to achieve data. So this kind of survey is merely done once after several years.

Nowadays, transit IC card is applied in various big cities. The cardholder percentage of office worker and student group is more than 90%. It is of practical sense to plan dynamic transit system, to improve quality of transit operating scheme, to reduce correlative survey cost by proper processing urban transit IC card information and to accumulate available transit passenger trip information.

## 2 Analyzing Transit IC Card Data

### 2.1 Time Distributing of Resident Transit trip

Transit passenger distributing is of obvious asymmetry. Hereinto passenger flow rhythm taking a day or a week as a cycle is most obvious. Thus passenger flow in holidays differentiates from that in workdays. Properly processing transit IC card data, we can find card-holders trip time distributing rule during transit operating period[2]. Transit trip time distributing status of all card-holders in a workday is illustrated in Figure 1. Card-holder trip periods distributing status of Changchun transit line 306 is illustrated in figure 2.

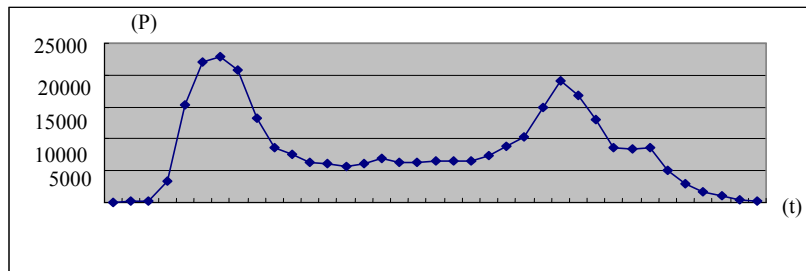


Fig. 1. Changchun Card-holder Trip Periods Distributing Status in a Whole Day

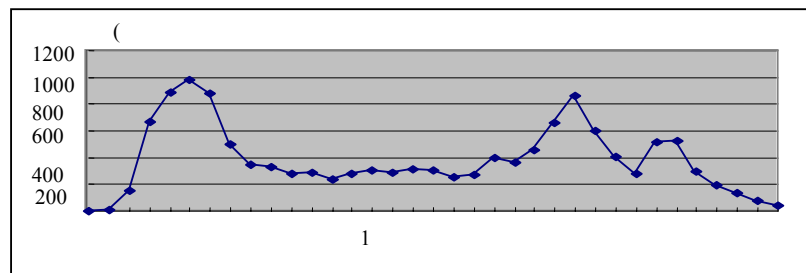


Fig. 2. Card-holder Trip Periods Distributing Status of Changchun Bus Line 306

The figures illustrate that transit passenger flow has obvious peak periods in the morning and evening. Passenger distributing is more concentrated in morning peak period. Evening peak period occurs during 16: 00 to 18: 00. It is as well off duty passenger rush hour. There is a sub peak period as to 19: 00, during which students leave school and people in service industry ring out to take bus. Transit passenger flow in rush hour accounts for a higher percentage in a whole day. The card-holder percentage during 6: 30 to 8: 30 accounts for 26% in whole day; the percentage of evening rush hour from 16: 00 to 18: 00 accounts for about 21%.

Since 2007, Changchun everyday card-scanning amount is as to 460 thousand person-time, accounting for about 38% of 1300 thousand person-time of all-day transit passenger flow. Morning and evening rush hours are card-holder centralized periods. The percentage of the periods ought to be higher than this value. According to the analysis of transit passenger spot check, the characteristics of card-holders and all passengers during transit trip periods are by and large consistent. Taking IC card data as swatch, the analysis can reflect the general characteristics of urban transit passenger flow time distributing.

## 2.2 Analyzing Transit IC Card Database Structure

Database structure of transit IC card charging system can be divided into two categories. One is when transit line is long and it is charged regarding various sections, a passenger scans his card severally when he gets into and gets off the bus to ensure passenger actual trip section and deduct the corresponding fare from IC card. The attributes of tuple in database table include passenger IC card number, line number, bus number, consuming date, stops and the time of getting on and off as well. Another category is when it uses a single fare in the whole journey, passenger scans IC card only when he gets into bus. Thus the tuple in database table doesn't include the attribute of getting off time. Beijing transit IC card system involves both circumstances. The database structure of transit IC card in many cities such as Changchun falls into the second category, the main attributes of tuple in database table are illustrated in Table 1.

**Table 1.** Status quo IC Card Database Structure

IC card NO.	Card-scanning time	line No.	Bus No.
249987	15:27:09	226	40445
394562	15:31:13	226	40445
179879	15:34:26	226	40445
394562	15:31:13	226	40445
.....	.....	.....	.....

Among the information needed to construct transit OD matrix, passenger ride OD information ought to include “getting into bus time”, “getting off time”, “stop of getting on”, “stop of getting off”, etc. But information such as stop of getting on and off is lack in today's most databases. In many cities where single fare is used in

whole journey, card-scanning time of getting on is input in transit IC card database, whereas getting off time is lack, so it is very difficult to deduce the corresponding stops.

Moreover, information of transit line, relevant stops and transit scheduling of various lines is as well demanded to construct transit OD matrix.

### 2.3 Analyzing Transit Trip Chain of IC Card-holder

The transit trip chain here means the transit trip route originated from a passenger's first stop of getting on of his first trip, to the stop of getting off of his last trip in a day. Three factors of stop, time and route are encompassed in transit trip chain.

Transit passengers of morning and evening rush hours are mainly office workers and students. It can be generally called work trip. Card-holders account for a high percentage of these passengers. In workdays, the time and route of getting on is of obvious rule and symmetry. Thus the transit trip chain is simple and close. The getting on stop of first transit trip is the getting off stop of ultimate transit trip.

Close transit trip chains can be divided into following categories.

(1)invariable round trip chain

There is a unique through transit line between trip origin and destination. The transit stops are invariable. Nothing but two stops of same line are involved in a trip chain. Most passenger round trips fall into the simplest mode of transit trip chain.

(2)round trip chain of a choice of lines

There are at least two transit lines of same direction between trip origin and destination. Out and home trip are of different lines. But the stops of getting on and off are invariable. Nothing but two stops is as well involved in a trip chain.

(3)trip chain including midway transfer

There is not through lines between trip origin and destination. One or more midway transfer are demanded. At least 5 stops are involved in a round trip chain. The sections of the transit trip chain can be divided into two categories, i.e., invariable line and lines of choice. The midway transfer stops can be different. Transit trips in non-rush hours are typically occasional trips, such as trips for shopping, medical care as well.

Table 2 shows part of tuple in database table of card-scanning ranking by card number and card-scanning time in Changchun Transit IC Card Database.

**Table 2.** IC Card Record Ranking by Card Number and Time

Card number	Time	Route No.	Bus No.
1459	07:47:55	10	60017
1459	17:01:22	10	60007
21586	05:57:42	62	60017
21586	16:45:54	362	60071
312656	06:05:30	3	50008
312656	06:47:30	364	250009
312656	16:57:06	364	251020
312656	17:43:27	3	50022

### 3 Constructing Methods of Transits O-D Matrix

#### 3.1 Determining Transit trip OD

Based on the above analysis, when transit O-D matrix is constructed, the key information is O-D stops of each trip, which is absent in IC card database. It should be estimated by characteristics of transit trip chain.

(1) In fixed line round trip chain and optional-route round trip chain table, only card-scanning time information of getting on is involved. Combined with transit schedules, direction and getting on stop can be inferred. As the transit chain is closed, the return stop of getting on is the foregoing stop of getting off. So judging stop of getting on of each record is all right.

(2) Midway-transfer transit trip chain is a little complex. But it is generally considered that the stop of getting off before transfer and the stop of getting on after transfer are of the same. Similarly, determining stop of getting on when card-scanning is enough. Constructing transit OD matrix need nothing but O-D stop information of each trip. Midway transfer can be neglected. Thus transfer information is meaningful to planning of transit transfer hub.

From the above discussion, because of the characteristics of closed transit trip chain, in order to construct OD matrix structure, we can determine the stop of getting off, depending on information of getting on stop of each trip.

#### 3.2 Coordinating and Sorting Record

The original card-scanning record must be coordinated and sorted, which are usually collected by IC card management center, following charge statistics and reckoning. Main steps are as follows:

(1) Transmit all the intraday data into a central database, rank up and index them according to IC card number;

(2) Add a attribute "trip classification," to Database structure setting conditions to screen IC card records being ordered. According to the categories of transit trip chain, the field is given four different category values. Type 4 is the "other" category, which beyond the foregoing categories;

(3) Establish transit trip chain category database table. There is null attribute in the table, such as stops of getting on and get off, moving direction and so on. According to category number, records will be respectively to transmit into database tables of different types;

(4) Link category database tables to various vehicle Schedules in the same day, judge the trip direction and stop site of same vehicle number when cards are scanned. Fill its code into air field as illustrated by Table 3;

(5) After stops are determined, transmit the major fields into IC card historical database to preserve. The structure is shown by table 4.

Computerized management of Transit enterprise's vehicle scheduling information is very important. Now many urban transit enterprises have not done this, which causes difficulties to IC card data comprehensive utilization of.

**Table 3.** Transit Chain Categories Database Table

Card No.	Date	Route	Vehicle No.	Time	Direction	Get on Stop	Get off Stop	Trip Type
390712	20070423	361	020287	07:05:39	1	36103	36108	1
390712	20070423	361	020305	18:16:05	2	36108	36103	1
390718	20070423	306	040083	07:13:25	2	30601	30607	1
390718	20070423	306	040076	17:21:42	1	30607	30601	1

**Table 4.** IC Card Historical Information Database Table

Card No.	Date	Route	Time	Get on Stop	Get off Stop	Trip Type
390718	20070423	306	07:13:25	30601	30607	1
390718	20070423	306	17:21:42	30607	30601	1

### 3.3 Calibrating Vehicle Schedules

In bus management and operation, travel Schedules is the important basis for vehicle operation .It provides a vehicle for arrival time and leaving time, when it arrived at the flow along the site in each one-way trip and turnover trip. Vehicle Schedules normally are demarcated by different season, time, sections and the traffic situation around, and are tested by the vehicles actual operation. Nowadays urban traffic conditions are complicated, which disturb buses' running greatly. When traffic Schedules convicted passengers disembark stop, it is difficult to get totally accurate matching time, which is a fuzzy relationship. Certain time warp is allowed.

Fuzzy set is described by the membership function. Generally speaking, bus sends up at origin station. The time arrived at the site and stopping time are random variables. In order to simplify data processing, in terms of different directions and sections, on the basis of a detailed traffic Schedules calibration, membership function of each route can be established for operation time in peak and non-peak hour.

Trapezoidal curve can be used to express "stop" and "trip time" fuzzy relationship. Trapezoidal membership function allows a certain time range, which is vehicles arrival time and leaving time, to ensure the corresponding relation of card-scanning time and stop of getting on, shown in Figure 3.  $t_n$  means the time when a bus arrive at stop  $n$  , AB means stopping time, Some passengers might scan cards after drive, function is the non-isosceles trapezoid. At card-scanning moment fall on CD section, stop  $n-1$  can be determined as the stop of getting on.

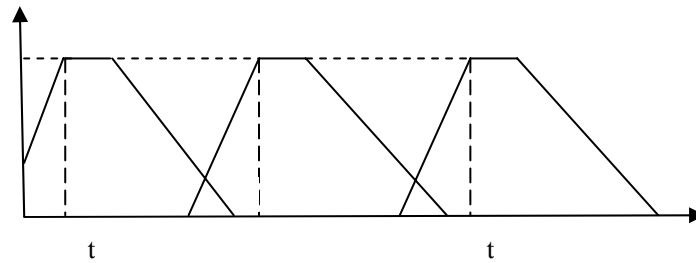


Fig. 3. Trapeziform curve of subjection degree function

### 3.4 Constructing Transit OD Matrix

#### (1) Line OD matrix

According to stop information of getting on and off of IC card record, passenger flow matrix of various transit stops OD matrix can be constructed of any transit line of whole day or different time periods. Take stops code of the transit line as "origin line" and "destination row". Take the line number as key character. Read the record in turn which takes in the same day's database table. Accumulates 1 in the corresponding matrix form, then the line OD matrix is generated. Line OD matrix may substitute for traditional passenger flow investigation. It can provide detailed passenger flow information at various stops along the route every day, to help schedule department reasonably disposition transportation resources.

#### (2) Net OD matrix

Net OD matrix involves all urban transit lines, the matrix dimension and all urban transit stops are of the same. The generating process is: The trip chain table of category 1 and 2 generates each line and row of OD matrix separately; for records of category 3, neglect the midway transfer process. Generate OD matrix between stops of various transit line. Net OD matrix is sum of the three categories, which can be generated according to the record of the entire day or various time periods.

The dimension of Net OD matrix is too large. The same stop of various lines uses same identical code. It can effectively reduce the dimension. Based on net OD matrix, the urban area transit OD matrix can be generated

#### (3) Urban area transit OD matrix

The urban area transit OD matrix takes inhabitant plot as trip OD points. From urban territory view, it can provide inhabitant transit trip information for transit planning and urban comprehensive transportation planning. For IC card information takes transit stops as trip OD, it is recommended that trip plot division method should be combined with kinds-gathering method on proximate stops by distance. Figure 4 illustrates the trip plot gravity center, merging 1,700 transit stops in Changchun. Based on trip plot, urban area transit OD matrix can be generated by merging transit net stops OD matrix.

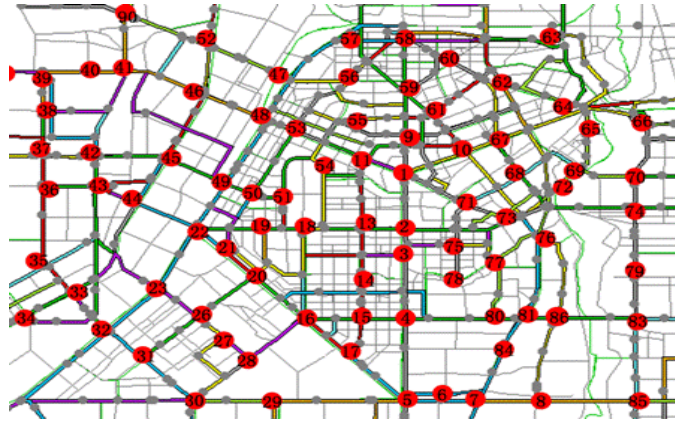


Fig. 4. gravity center sketch map after merging trip plots

## 4 Conclusions

It systematically researches on processing method and comprehensive application of transit IC card information. The concept of transit trip chain is proposed. Methods of information processing and transit OD matrix constructing are discussed, depending on existing IC card database structure. Main conclusions as follows:

(1) Transit trip time distribution characteristics of card holders and all passengers by spot checking are basically consistent. Taking IC card data as sample, we can analyze the overall time distribution characteristics, which can reflect to urban transit passenger flow.

(2) It introduces the concept of transit trip chain. The enclosed trip chain can be constructed according to the recording characteristics of card holder. We could only judge passengers' getting on stops to simplify IC card information processing.

(3) Use IC card information to structure public transportation OD matrix may save the cost of massive passenger flows investigation, and the data is reliable, timeliness. Transit line OD matrix can provide the real-time information for daily transit operation scheduling. The urban district public transportation OD matrix may provides the dynamic information for the usual public transportation line adjustment and for the regular public transportation system plan.

The current various urban transit IC card database is generally of different systems, which is viewed as information isolated island. Massive data leave unused. It cannot apply effectively in transit enterprise operation scheduling and transit planning, which means enormous waste of information resource. We should unify regular and rail transit to build urban transit information platform, and integrate transit IC card information with transit enterprise operation management information, the transit inquiry information, urban GIS information and the transit vehicles GPS information and so on, to construct transit data warehouse[3], to realize correlation information fusion, effectively utilize comprehensive IC card information.



## References

1. W. Wang, X.M. Yang and X.W. Chen, *Planning Methods and Management Technologies of Urban Transit System*, Science Publication, 2002, p. 62.
2. F.M. Shi, "Research on Constructing Methods of Transit OD Matrix Based on IC Card Data", *Master's degree dissertation of Jilin University*, 2004, p. 15.
3. X. Dai, X.W. Chen and W.Y. Li, "Study of Data Mining Technique for Bus Intelligent Card Data Processing", *Computer and Communications*. 24(1), 40-42 (2006).