

HAINAN HUANG, Ph.D.^{1,2}
E-mail: hhnan021@foxmail.com

JIAN RONG, Ph.D.²
E-mail: jrong@bjut.edu.cn

PENGFEI LIN, Ph.D. Candidate²
E-mail: pengfei_lin@qq.com

JIANCHENG WENG, Ph.D.²
(Corresponding author)
E-mail: youthweng@bjut.edu.cn

¹ College of Transportation and Civil Engineering
Fujian Agriculture and Forestry University
63 Xiyuangong Road, Minhou County, Fuzhou,
350108, China

² Key Lab of Traffic Engineering
Beijing University of Technology
100 Pingleyuan, Chaoyang District, Beijing, 100124, China

Traffic and Space
Original Scientific Paper
Submitted: 11 Aug. 2018
Accepted: 3 Oct. 2019

UNDERSTANDING DAILY TRAVEL PATTERNS OF SUBWAY USERS – AN EXAMPLE FROM THE BEIJING SUBWAY

ABSTRACT

The daily travel patterns (DTPs) present short-term and timely characteristics of the users' travel behaviour, and they are helpful for subway planners to better understand the travel choices and regularity of subway users (SUs) in details. While several well-known subway travel patterns have been detected, such as commuting modes and shopping modes, specific features of many patterns are still confused or omitted. Now, based on the automatic fare collection (AFC) system, a data-mining procedure to recognize DTPs of all SUs has become possible and effective. In this study, DTPs are identified by the station sequences (SSs), which are modelled from smart card transaction data of the AFC system. The data-mining procedure is applied to a large weekly sample from the Beijing Subway to understand DTPs. The results show that more than 93% SUs of the Beijing Subway travel in 7 DTPs, which are remarkably stable in share and distribution. Different DTPs have their own unique characteristics in terms of time distribution, activity duration and repeatability, which provide a wealth of information to calibrate different types of users and characterize their travel patterns.

KEY WORDS

*daily travel pattern; smart card data;
station sequence; subway user; data mining;*

1. INTRODUCTION

Urban rail transit has become an indispensable option for daily travel in China, especially for commuters in the metropolises such as Beijing and

Shanghai [1, 2]. In 2015 the passenger traffic of the Beijing Subway was 3.32 billion times, and the average daily passenger volume reached 9.11 million times [3]. In Shanghai, the share of rail transit passengers reached 53% in public transportation while the proportion of bus transit decreased to 33% [4].

With millions of people choosing rail transit as their primary travel mode, the congestion occurs in the peak hour. Many measures for passenger inflow control have been taken in the Beijing and Shanghai Subway to limit the number of passengers entering the subway station and waiting at the platform [5, 6]. Other measures such as pre-trip information are also used to provide travellers with help in route selection and avoiding crowded stations. For example, a mobile phone App function has been launched for passengers inquiring about the real-time congestion of the Beijing Subway. However, this does not essentially solve the problem of subway congestion. Only with in-depth understanding of the user's travel behaviour characteristics and travel patterns, the nature of subway traffic jams can be discovered and a reasonable solution can be proposed [7, 8]. Understanding the daily travel patterns (DTPs) of subway users (SUs) at the individual level can provide richer and useful information to reveal the regulation and characteristics of travellers, and enhance system performance or better assess network investments [9-11].

Travel patterns of transit passengers have been discussed with the automatic fare collection (AFC) data in many research papers. Every transit rider's travel pattern is detected based on the temporal and spatial characteristics by trip chains [8, 11, 12], multi-week activity sequences [13] or other factors [14, 15]. The authors in [8] have classified the travel patterns into five clusters (very low, low, medium, high, very high) with regularity levels of transit customers by the DBSCAN algorithm and found that approximately 41% of transit riders fall into the categories of high and very high (namely commuting patterns), which are further studied in [12]. The authors in [11] focused on detecting the detailed analysis of the characteristics of four user groups (exclusive commuters, non-exclusive commuters, leisure travellers and non-commuter residents), and the results indicate that visitors and registered users have different characteristics in temporal and spatial variability, activity patterns, sociodemographic characteristics, and mode choices. The authors in [13] have investigated passenger heterogeneity and defined eleven clusters (including conventional 9-to-5 commuters, a variety of non-working routines, and non-conventional work routines) as the activity patterns of London public transport users. According to the user's travel map, the authors in [14] used the BP model to classify public transport passengers into the commuting and non-commuting ones. Public transport passengers from the Beijing subway are divided into five categories in [15], including standard commuter passengers, flexible commuter passengers, high-frequency frequent passengers, life class passengers and short-term low-frequency passengers, and travel intensity, time dimension and spatial dimension have been detected to illuminate the travel characteristics at the aggregate level.

It can be concluded that the current research on passenger travel patterns focuses on passenger classification methods, while the analysis of different types of passengers is mainly concentrated on the level of aggregation. These analyses of trip patterns put emphasis on the identification of different modes rather than the fine-grained analysis of trip characteristics [16]. Although the time and space regularity of different travel types are detected, the daily travel characteristics of subway users are still not clear enough for subway operation optimization and line network planning. Meanwhile, while these works highlight the potential of clustering algorithm

to classify the travel patterns, the approaches are still limited with the clustering variables which might ignore or be affected tremendously by some abnormal data, as well as the number of clusters [17]. Moreover, it takes a significant amount of work to reduce the noise and extract indicators before the travel pattern processing [18, 19].

Therefore, a simple data mining method was launched to analyse DTPs of SUs from a more detailed perspective, and to mine travel information that can be used for daily operation monitoring and future network planning of the subway system. The existing studies neither fully utilized all AFC records for mining user daily travel patterns nor explored the detailed characteristics for understanding the typical DTPs. These features distinguish this paper from the others. The main contribution of this study is twofold. First, this paper attempts to propose an easy way of identifying and interpreting DTPs of SUs. DTPs will be represented by station sequence (SS), which is obtained by comparison of the stations in SU daily subway trips. Second, the data mining approach is used to understand the daily popular travel behaviours of the Beijing Subway with a week AFC data. DTPs proposed in this paper can more clearly understand the user's travel characteristics, such as the choice preferences of travel time and stations, one-way trip or round trip, random trip or long-lasting trip. Through the analysis of the characteristics of DTPs, the travel habits and needs of most users can be further understood.

The remainder of this paper is organized as follows. The real datasets used in the study are introduced and the proposed data mining process is explained in the following Section. The experimental result is presented and discussed in Section 3. Finally, the paper is concluded by summarizing the research findings and suggesting directions for future research.

2. DATA MINING PROCESS

2.1 Data foundation

At present the Beijing Subway adopted an all-purpose system for billing, wherein the users need to swipe the smart card both inbound and outbound. With the exception of the airport line, the transfers between lines do not require re-swipe. When a passenger swipes into the station, the AFC system records the information of the entry time, line and station of the inbound. When the card is

Table 1 – An example of the Beijing Subway AFC analysis data

Card code	Entry line number	Entry station number	Entry time	Exit line number	Exit station number	Exit time
86270***	13	37	10/17 8:35	13	27	10/17 9:00
86270***	13	27	10/17 18:50	8	13	10/17 19:26
86270***	13	37	10/18 8:27	13	27	10/18 8:52
86270***	13	27	10/18 18:46	8	13	10/18 19:21
86270***	13	37	10/19 8:20	13	27	10/19 8:45

swiped out, the AFC system will fill in the previous record with the information of the exit time, line and station of the outbound [20]. Other information such as deal type, card type, ticket type, device ID, etc. are also recorded in the transaction record.

In order to present the daily travel patterns, seven main fields of the Beijing Subway AFC data are extracted in the data pre-processing. Table 1 shows a random user’s extracted transaction records in a week, which are comprised of five subway trips (each trip a line) with the user’s card number, the entry and exit time, inbound and outbound line numbers and station numbers.

2.2 Station sequence extraction

A trip chain comprising temporal and spatial information is considered a useful way to present and analyse travellers’ behaviours [21, 22]. In order to depict the individual activity sequence, the stops (e.g. home and work, origin and destination) are used to define the travel patterns [23-25]. In this

paper, the station sequence (SS), a string of numbers consisting of the station code (SC), is selected to represent the users’ trip chains. SC refers to the user’s inbound and outbound station by comparing whether the station is the same with the previous ones. The main process is shown in Figure 1, and the specific steps are as follows:

Step 1: Process the site number. The site number (SN) mainly consists of two swipe records, the line number and the station number. These two indicators need to be joined to name SN (shown in Equation 1). The line number and the station number will be changed to character strings, and then they can be joined together instead of simple numeric additions. When the line or station number is less than 10, a 0 is automatically added to the front to form two digits. SN has four characters, of which the first two digits indicate the line number and the last two digits indicate the station number.

$$SN = (Entry/Exit)Line\ Number + (Entry/Exit)Station\ Number \tag{1}$$

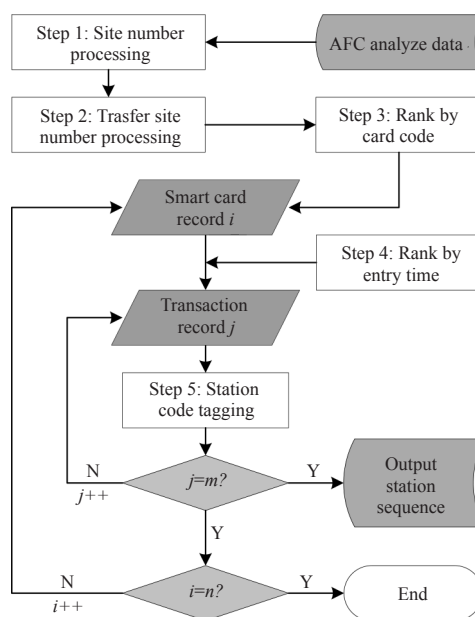


Figure 1 – Station sequence extraction method of SUs

Step 2: Process the transfer site number. Since there are two or more lines passing through the transfer site, SN is different at the same location of different lines. Therefore, SN has to be unified into the smallest one when there are two or more SNs of the transfer site.

Step 3: The transaction records are sorted by card code and renumbered as i ($1, 2, 3, \dots, n$), and the i -th user's transaction records are extracted. n refers to the total number of SUs in a day.

Step 4: The i -th user's transaction records are ranked by entry time and numbered as j ($1, 2, 3, \dots, m$), where m refers to the total number of the i -th user's transaction records.

Step 5: SCs are tagged by comparison of SNs, shown in *Figure 2*. And the core steps are described from Step 5.1 to 5.3.

Step 5.1: $SC(1)$ and $SC(2)$ are assigned 1 or 2, shown in *Equation 2*.

$$SC(1) = 1$$

$$SC(2) = \begin{cases} 2, & \text{if } SN(2) \neq SN(1) \\ 1, & \text{if } SN(2) = SN(1) \end{cases} \quad (2)$$

where $SC(1)$ and $SN(1)$ refer to SC and SN of the entry station in the $j=1$ transaction record, $SC(2)$ and $SN(2)$ refer to SC and SN of the exit station in the $j=1$ transaction record.

Step 5.2: SCs from 2nd to m -th transaction records are tagged. Check the rest of the trip records whether the entry or exit station are the same as the previous SNs.

For the j -th transaction record, if $SN(2j-1)=SN(x)$ (x from 1 to $2j-2$), $SC(2j-1)=SC(x)$; if not, then $SC(2j-1)=\max\{SC(x)\}+1$. If $SN(2j)=SN(x)$ (x from 1 to $2j-1$), $SC(2j)=SC(x)$; if not, then $SC(2j)=\max\{SC(x)\}+1$.

The abbreviations used in *Figure 2* are described as follows: $SC(2j-1)$ and $SN(2j-1)$ refer to SC and SN of the entry station in the j -th transaction record; $SC(2j)$ and $SN(2j)$ refer to SC and SN of the exit station in the j -th transaction record.

Step 5.3: all transaction records of the i -th user are tagged and $SS(i)$ is output.

$$SS(i) = SC(1) - SC(2) + \dots + SC(2j-1) - SC(2j) + \dots + SC(2m-1) - SC(2m) \quad (3)$$

where $SS(i)$ refers to the i -th user's station sequence.

Step 6: Complete the station sequence extraction for all SUs.

Taking the user's transaction record in *Table 1* as an example, the user's SS is identified and shown in *Table 2*. Some trip characteristics of this user can be easily determined with SS, e.g., the rider travelled in only two different stations during the week, as well as the rider travelled in a round trip on the 17th and 18th but a single journey on the 19th. Other information such as the frequency of stations, travel days, travel time, and duration of trips and between trips can also be calculated by SS at the individual level or at the aggregate level. The empirical data mining analysis indicates that the identified SS is a succinct and visual way to understand SUs' travel choices.

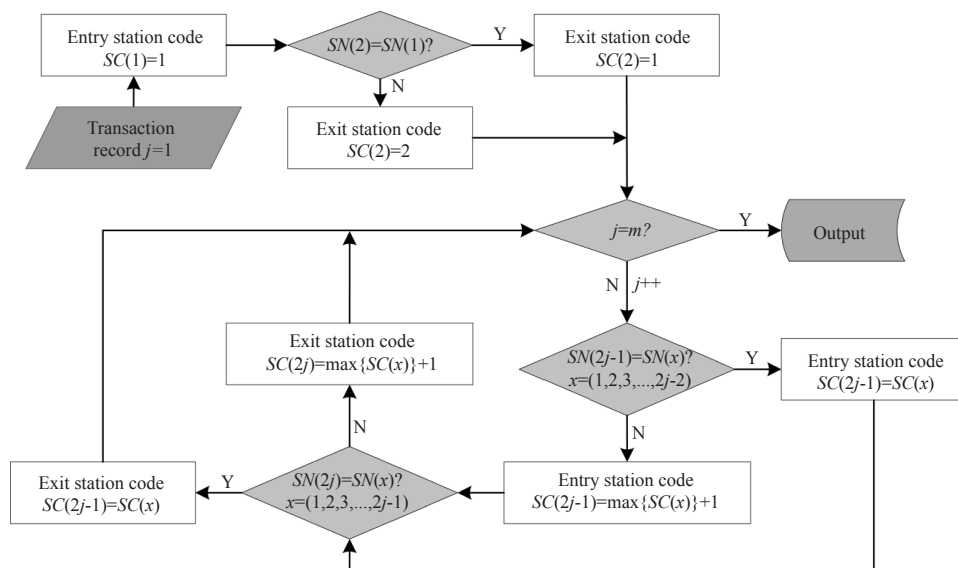


Figure 2 – Station code tagging processing

Table 2 – An example of SS extraction

Card code	Entry time	Entry site number	Exit site number	Entry station code	Exit station code	SS
86270***	10/17 8:35	0813 (1337) *	1327	1	2	1-2+2-1
86270***	10/17 18:50	1327	0813	2	1	
86270***	10/18 8:27	0813 (1337) *	1327	1	2	1-2+2-1
86270***	10/18 18:46	1327	0813	2	1	
86270***	10/19 8:20	0813 (1337) *	1327	1	2	1-2

* Lines 13 and 8 intersect at this transfer station. Therefore, SN is 0813 or 1337, and 1337 is replaced by 0813 during the identification process.

3. CASE ANALYSIS AND DISCUSSION

Using the station sequence identification method proposed in Section 2, the Beijing Subway AFC data are identified and SSs are calculated by a Python program. To demonstrate DTPs, the data were collected in a typical week from Sunday, October 16 to Saturday, October 22, 2016. There were no special holidays within 7 days before and after the selected week, so the data could avoid any holiday effect. There are 7,181,371 smart cards with 32,743,509 transaction records in the selected week. It only takes a few minutes to finish the data mining work on about 2,729,934 smart cards with 4,677,644 transaction records per day. About 3,800 to 6,300 types of SS per day are obtained through data mining. There are about 4,000 types of DTPs in average on weekdays, and the number rises to around 6,000 on the weekends.

More detailed information is detected and shown in the following section. DTPs features of Beijing SUs are introduced first. The most popular DTPs are shown and discussed in Section 3.1, where DTPs are presented by SS.

For the high similarity of different DTPs in a day, the detailed distribution of entry time and activity duration with one-day data (Tuesday) are investigated in the subsequent analysis, then the repeated travel days with a week of data are extracted to mine

information of regular riders in Section 3.2. The inbound time distribution mainly reflects the peak hours of different trips in different DTPs, which is helpful in optimization of the vehicle scheduling. The activity duration characterizes the length of activity of different DTPs, which can be used to determine the travel destination or activity nature of the traveller. With regard to the repeated travel days, it can reflect the repeatability and regularity of the user's travel in a week, and provide basic data support for managers to identify and optimize the commuter travel services.

The typical trips of Top 7 DTPs are summarized and discussed in Section 3.3 to understand the behaviour of SUs and to provide more useful finer features for daily management optimization.

3.1 Daily travel patterns based on SS

Table 3 shows the results of the main DTPs proportions of the Beijing Subway users. T_i is used to present the i -th DTP sorted based on the proportion of all users each day. Only seven DTPs account for more than 1%, while the remaining types are less than 1% (no comment in this study). The proportion of Top 7 DTPs has not changed in this week's ranking, and their total proportion keeps around 93%. About 40% SUs take a one-way trip (T1) in a day, and the proportions of T1 on workdays are smaller

Table 3 – The main DTPs proportion of the Beijing Subway users [%]

DTPs	SS	Sun.	Mon.	Tue.	Wed.	Thu.	Fri.	Sat.	Average	Standard deviation
T1	1-2	48.20	38.61	37.02	37.15	37.74	41.59	44.12	40.63	3.93
T2	1-2+2-1	25.54	35.16	36.13	36.06	35.80	30.55	26.84	32.30	4.27
T3	1-2+2-3	9.11	10.02	9.97	10.04	9.87	10.86	9.98	9.98	0.47
T4	1-2+3-1	5.15	5.03	5.24	5.13	5.16	4.69	5.50	5.13	0.22
T5	1-2+3-4	2.65	2.42	2.46	2.42	2.35	2.65	2.95	2.56	0.19
T6	1-2+2-3+3-1	2.02	1.89	2.04	2.04	2.02	2.09	2.27	2.05	0.11
T7	1-2+2-3+3-4	0.90	0.79	0.84	0.85	0.85	1.03	1.08	0.91	0.10
Total		93.59	93.93	93.71	93.70	93.78	93.46	92.75	93.56	0.36

than on weekends. About 30% SUs take a round trip (T2) in a day, and the proportions of T2 on workdays are higher than on weekends. The standard deviations of the Top 7 DTPs are extremely small, which indicates that the proportions of the most popular travel patterns are highly stable, although there are some differences between workdays and weekends.

It can also be calculated from Table 3: (1) 73% of SUs travel in two different stations in a day, 90% less than three and about 94% less than four, which implies that most subway riders often choose and travel in fixed stations; (2) most SUs choose the alighting station of the last trip as the boarding station of the next trip.

Meanwhile, the distributions of Top 7 DTPs on workdays and weekends are also highly similar and stable. Taking T1 as an example, the distribution of the entry time of T1 in a week is presented in Figure 3. The horizontal axis in the figure indicates the entry time of trips in hours (1 means 0:00-1:00, 2 means 1:00-2:00 and so on), and the vertical axis indicates the number of trips. The distribution curves of T1 trips on workdays are almost the same with pinnales from 7:00 to 9:00 and from 17:00 to 19:00. On the other hand, the distribution curve of T1 trips on Sunday is similar to the curve on Saturday, all of which are without significant peaks.

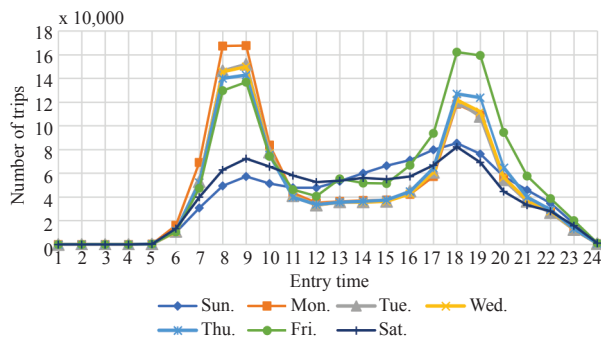


Figure 3 – Distribution of entry time of T1 in a week

3.2 Detailed characteristics of DTPs

Distributions of entry time analysis

Distributions of entry time analysis present the percentages of Top 7 DTPs based on the entry time, which clearly show the passenger's preference for travel time. In Figures 3 and 4, the majority of Top 7 users depart around morning peak hours (7:00 am~9:00 am) and return during evening peak hours (5:00 pm~7:00 pm). In particular, T6 and T7 have three peak hours (Figure 4b), wherein the first peak hours are the same as T2~T5, the second peak hours

are from 12:00 to 14:00, and their third peak hours start at 18:00 and last until 22:00. The distributions of T6 and T7 suggest two types of riders: (1) one's first and second trips aim to commute, and the third trip to return; (2) one's first trip aims to commute, the second and third trips to return by passing another stop for a short visit.

Activity duration analysis

Durations of stay at the same station have been employed as the simplest attributes correlating with the trip purposes [13, 21]. Activity duration analysis in this section focuses on activity intervals between trips of SUs, which is comprised of the transfer duration (TD) and the whole duration (WD). TD refers to the activity duration between trips, which is counted from the exit time of the trip to the entry time of the next trip. WD refers to the activity duration between the first trip and the last trip, which is counted from the entry time of the first trip to the exit time of the last trip in a day. Figure 5 presents the conceptual definition of TD and WD in the sketch of a SU one-day trips, whose DTP is T5, TD is 8 h and WD is 10 h.

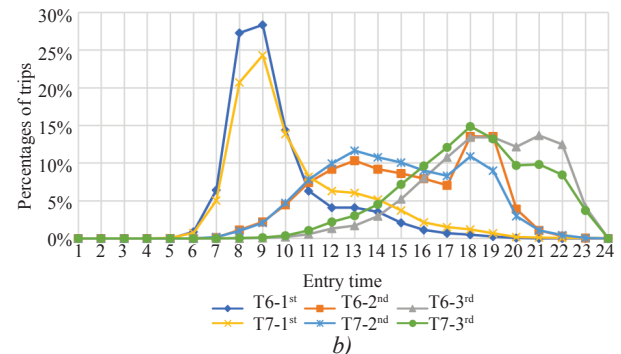
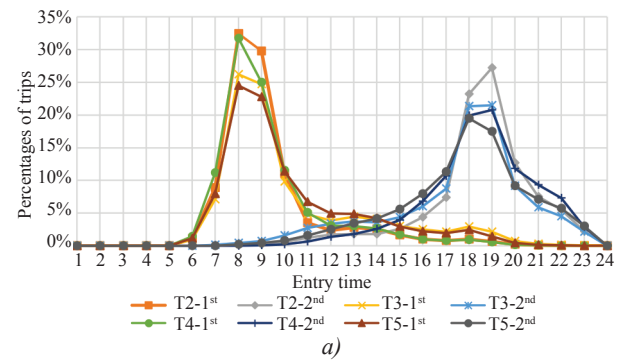


Figure 4 – Distribution of entry time of T2~T7 in a working day

(T_i-1^{st} means the first trip of the i -th travel pattern; T_i-2^{nd} means the second trip of the i -th travel pattern; T_i-3^{rd} means the third trip of the i -th travel pattern)

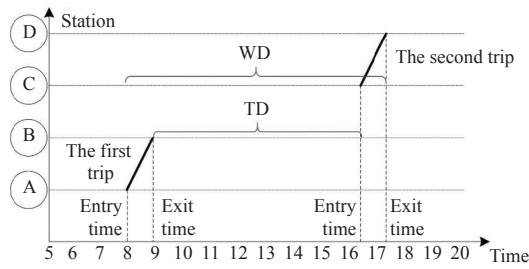


Figure 5 – A sketch of the spatial and temporal distribution of one-day trips

Figure 6 shows the distributions of TD and WD of T2~T7 on a working day. Some superficial conclusions can be observed: (1) TD and WD have the same distribution trend and WD is generally one hour later than TD; (2) TD and WD of T2 and T4 are unimodal distributions, while high peaks of TD are clustered at 10 h and WD at 11 h; (3) TD and WD of T3 and T5 are bimodal distributions, with a smaller peak at 4 h to 5 h comparing with T2 and T4; (4) TD and WD of T6 and T7 are equilibrium distributions, though TD-1st is bimodal (peaks at 2 h and 10 h) and TD-2nd is a unimodal (a peak at 2h). These duration characteristics indicate that the

average length of travel for these DTPs is about 1 hour, and the duration of travel between trips is about 9 hours. The interval between two trips in the morning period is relatively small, suggesting that only simple activities have been done here, such as dropping off children to school.

Repeated travel days analysis

SUs' travel patterns may differ daily, because they would change their routes up to different purposes, and even if they go to the same place, they might choose other modes of transportation [25, 26]. The repeated travel days analysis tries to assess the loyalty of SUs with the variable R , which refers to the number of days in a week that the user travels with the same DTP. Table 4 shows the proportions of the repeated travel days of Top 7 DTPs, which declines significantly as R increases. Most users do not maintain the same DTP every day; however, there is still a certain percentage of users who insist on the same DTP. About 16% of T1 users and 32% of T2 users travel in the same DTP more than 3 days. The authors further extract the data ($R \geq 5$) and the distribution of the repeated

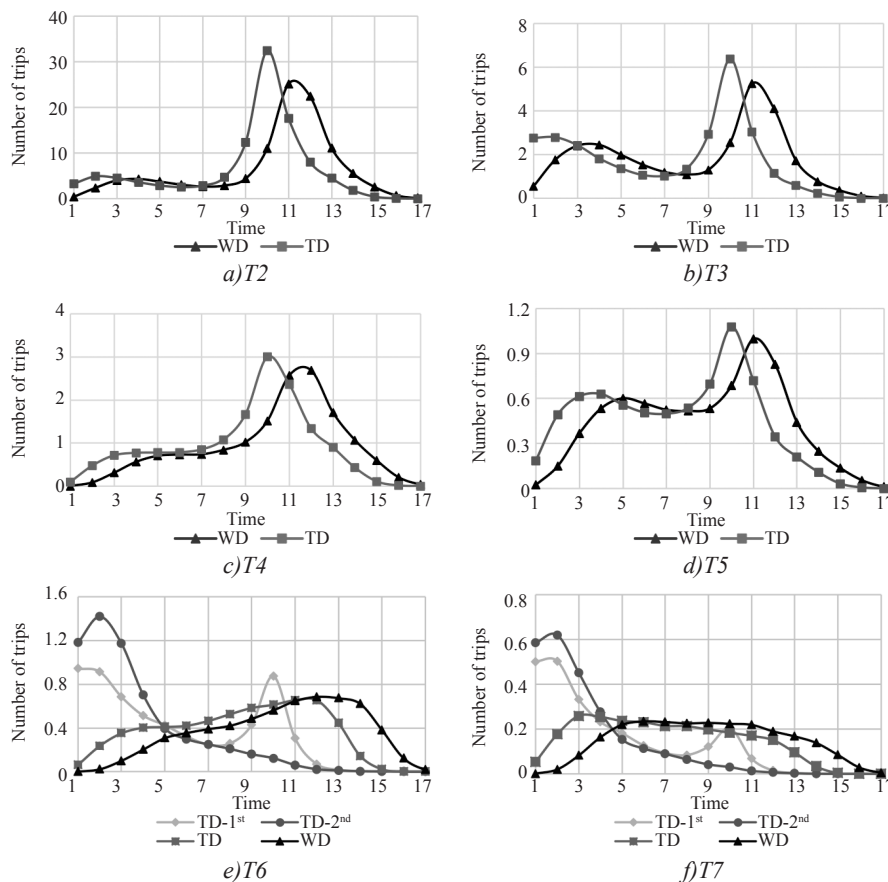


Figure 6 – Distributions of TD and WD of T2~T7 on a working day (The ordinate unit is 104; TD in e) and f) is the sum of TD-1st and TD-2nd, where TD-1st and TD-2nd refer to the user's first and second TD respectively.)

Table 4 – The proportion of the repeated travel days of Top 7 DTPs in a week [%]

Type number	SS	R						
		1	2	3	4	5	6	7
T1	1-2	61.31	22.57	8.25	4.37	2.65	0.71	0.15
T2	1-2+2-1	51.59	16.28	10.77	9.64	8.54	2.61	0.57
T3	1-2+2-3	76.63	13.96	4.79	2.67	1.63	0.27	0.05
T4	1-2+3-1	79.26	12.61	4.41	2.26	1.26	0.17	0.03
T5	1-2+3-4	88.17	8.65	1.97	0.82	0.35	0.03	0.00
T6	1-2+2-3+3-1	89.13	8.97	1.45	0.33	0.10	0.01	0.00
T7	1-2+2-3+3-4	94.76	4.62	0.51	0.09	0.02	0.00	0.00

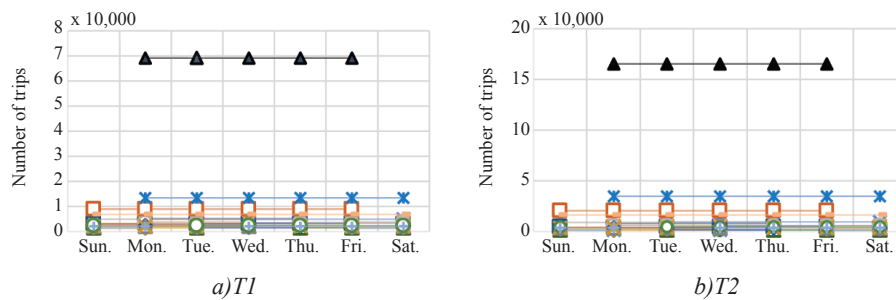


Figure 7 – The repeated travel days ($R \geq 5$) distribution of T1 and T2

travel days of T1 and T2 is shown in Figure 7. The horizontal axis indicates the number of travel days, and the vertical axis shows the number of users who travel continuously. The lines in the figure connect the days when the user travels in the same DTP. It can be easily found that most users have the same T1 and T2 from Monday to Friday ($R=5$, continuous 5-day-travel with the same DTP), then from Monday to Saturday and from Sunday to Friday ($R=6$). This distribution is consistent with the characteristics of office workers and student groups in Beijing, whose travel is very regular, with a fixed number of travel days and a fixed DTP.

3.3 Typical trips of DTPs

With features summarized from detailed analysis of DTPs, the typical trips of Top 7 DTPs are presented in Figure 8 to better understand the travel preferences of Top 7 DTPs. The boarding and alighting time shown in the figure are not the exact times but the approximate times. The similarities and differences between Top 7 DTPs are visually demonstrated in the diagrams. The first trips of Top 7 DTPs in the morning peak hour are similar and the return trips are of uniqueness, which indicates that most SUs' first trips in the morning are more fixed

in travel time than the return trips in the afternoon or in the evening, having been proved by [11, 18, 19, 22, 25].

The typical trips of DTPs can be used to show finer and common characteristics of different types of SUs, and provide a visual representation of the specific characterization to better understand DTPs [27]. For example, T1 users only travel a single trip by subway one day, implying that they might choose trips by other modes of travel, such as taking a bus, a customized taxi or a neighbourhood car-pool which has become popular in China, or they might find it not easy to reach the subway stations by walking [28]. In this case, the station distribution of T1 may provide a wealth of information about inconvenient or insufficient subway stations, which can be used to optimize the subway network.

4. CONCLUSION

A data mining methodology is proposed for analysing the daily travel pattern features observed by smart card data. Daily travel patterns (DTPs) are reflected by station sequence (SS), which is calculated from station code (SC). By the use of the Beijing Subway's weekly smart card data, the authors have found that the Beijing Subway DTPs are concentrated in seven types. The characteristics of the Top 7 DTPs are discussed in three aspects;

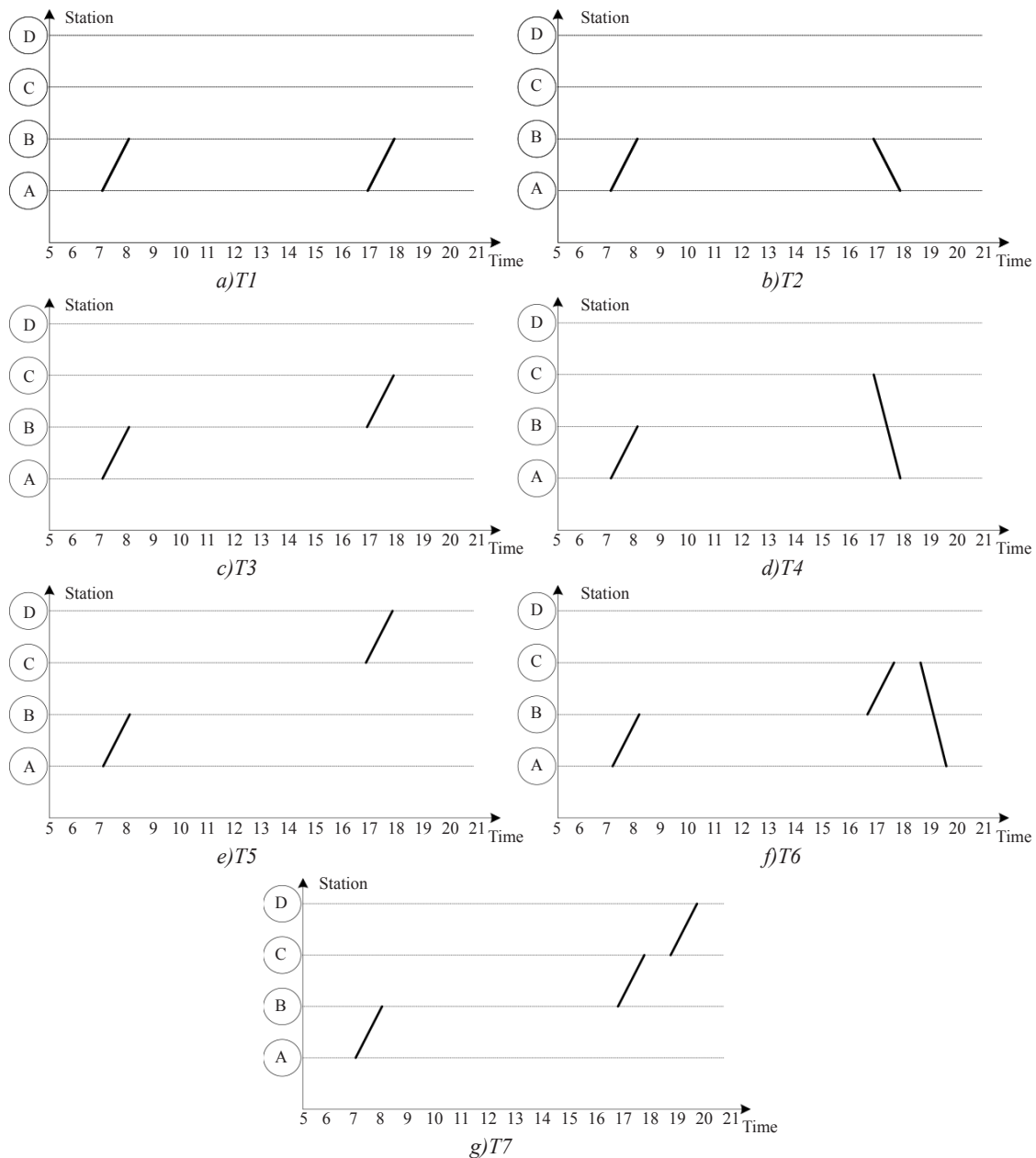


Figure 8 – The spatial and temporal distribution of Top 7 DTPs

the distribution of entry time, the activity duration, and the repeated travel days. In summary, the typical trips of the Top 7 DTPs are extracted and demonstrated in the sketch of the spatial-temporal distribution. The results show that most types of SUs have remarkable stability in share and distribution of all trips, even though most individual users do not travel at the same DTP every day. Meanwhile, different DTPs have their own unique characteristics in terms of time distribution, activity duration and repeatability, which provide a wealth of information to calibrate different types of users and characterize their travel patterns.

By applying the proposed method, the subway managers are able to understand the features of the most popular DTPs. For example, this study represents the users' segments who travel to fixed stations and who contribute the same DTPs in a week. The proposed method can also be applied to the estimation of a weekly travel pattern (WTP) of SUs or other kinds of transportation with the same trip information, such as actual origins and destinations. In terms of data applications, DTP can be extracted in all AFC systems with transaction records regardless of the entry mode (via smart card, QR code, or face) selected by the passenger. However,

DTPs might have different manifestations while the subway networks structure and urban land development and utilization differ [29, 30]. In addition, the dataset used in this study does not include the users' personal attributes nor the changes of the rail transit schedules and transport operators' policies [31]. In the future, the proposed method will be applied to the assessment of users' travel patterns on the rail transit control centre with more details identified.

ACKNOWLEDGEMENT

This research was funded by the National Natural Science Foundation of China (NFSC), Grant number 51578028 and the Education Department of Fujian Province, China, Grant number JAT160167.

黄海南^{1,2} Email: hhnan021@foxmail.com

荣建² Email: jrong@bjut.edu.cn

林鹏飞² Email: pengfei_lin@qq.com

翁剑成(通讯作者)² Email: youthweng@bjut.edu.cn

¹ 福建农林大学交通与土木工程学院
福州市闽侯县溪源宫路63号

² 北京工业大学交通工程重点实验室
北京市朝阳区平乐园100号

地铁用户日出行模式分析：以北京地铁为例 摘要

日出行模式 (DTPs) 反映了用户的短期出行行为和实时特征, 有助于地铁规划者更好地了解地铁用户 (SU) 的出行选择和规律性。虽然目前有些常规出行模式 (如通勤和购物模式) 已经被检测出, 但大部分出行模式的精细化特征仍然还较模糊。基于自动票价收集 (AFC) 系统, 利用数据挖掘程序识别所有用户的日出行模式成为可能和有效手段。本研究中利用 AFC 系统构建用户的站点序列, 并通过站点序列对用户日出行模式进行标识。以北京地铁一周的全样本刷卡数据为例, 分析北京地铁用户的日出行模式特征。结果表明, 超过 93% 的北京地铁用户日出行模式集中在 7 种模式, 这 7 种出行模式在占比及分布等方面具有显著的稳定性。同时, 不同的日出行模式在时间分布、活动持续时间和重复性方面拥有各自独特的特征。这些丰富的信息可用于刻画地铁用户及其出行模式的特征。

关键词

日出行模式; 智能卡数据; 站点序列; 地铁用户; 数据挖掘

REFERENCES

[1] Long Y, Thill J. Combining smart card data and household travel survey to analyse jobs–housing relationships in Beijing. *Computers, Environment and Urban Systems*. 2015;53: 19-35.

[2] Sun Y, Shi J, Schonfeld PM. Identifying passenger flow characteristics and evaluating travel time reliability by visualizing AFC data: a case study of Shanghai Metro. *Public Transportation*. 2016;8: 341-363.

[3] *Beijing Transportation Development Annual Report*. Beijing Transportation Research Center; 2016. Chinese.

[4] *Shanghai Comprehensive Traffic Operation Annual Report*. Shanghai Urban and Rural Construction and Transportation Development Research Institute; 2017. Chinese.

[5] Jiang ZB, Liu W, Zhu BQ. Reinforcement learning approach for coordinated passenger inflow control of urban rail transit in peak hours. *Transportation Research Part C*. 2018;88: 1-16.

[6] Xu XY, Liu J, Li HY, Jiang M. Capacity-oriented passenger flow control under uncertain demand: Algorithm development and real-world case study. *Transportation Research Part E*. 2016;87: 130-148.

[7] Boyle DK, Foote PJ, Karash KH. Public transportation marketing and fare policy. *Transport in the New Millennium*. 2000;A1E06.

[8] Ma XL, Wu YJ, Wang YH, Chen F, Liu JF. Mining smart card data for transit riders' travel patterns. *Transportation Research Part C*. 2013;36: 1-12.

[9] Yang Y, Herrera C, Eagle N, González MC. Limits of predictability in commuting flows in the absence of data for calibration. *Scientific Reports*. 2014;4: 5662.

[10] Ren Y, Ercsey-Ravasz M, Wang P, González MC, Toroczkai Z. Predicting commuter flows in spatial networks using a radiation model based on temporal ranges. *Nature Communications*. 2014;5: 5347.

[11] Ortega-Tong MA. *Classification of London's public transport users using smart card data*. S.M. Thesis. Cambridge, MA: Massachusetts Institute of Technology; 2013.

[12] Ma XL, Liu CC, Wen HM, Wang YP, Wu YJ. Understanding commuting patterns using transit smart card data. *Journal of Transport Geography*. 2017;58: 135-145.

[13] Goulet-Langlois G, Koutsopoulos HN, Zhao J. Inferring patterns in the multi-week activity sequences of public transport users. *Transportation Research Part C*. 2016;64: 1-16.

[14] Liang Q, Weng JC, Lin PF, Zhou W, Rong J. Public transport commuter identification based on individual travel graph. *Journal of Transportation Systems Engineering and Information Technology*. 2018;18(2): 100-107. Chinese.

[15] Zhou QR, Zhao P, Yao XM. Passenger classification for urban rail transit by mining smart card data. *Journal of Transportation Systems Engineering and Information Technology*. 2018;18(1): 223-230. Chinese.

[16] Briand A, Côme E, Trépanier M, Oukhellou L. Analyzing year-to-year changes in public transport passenger behaviour using smart card data. *Transportation Research Part C*. 2017;79: 274-289.

[17] Morency C, Trépanier M, Agard B. Measuring transit use variability with smart-card data. *Transport Policy*. 2007;14: 193-203.

- [18] Kusakabe T, Asakura Y. Behavioural data mining of transit smart card data: A data fusion approach. *Transportation Research Part C*. 2014;46: 179-191.
- [19] Zhao JJ, Qu Q, Zhang F, Xu CZ, Liu SY. Spatio-temporal analysis of passenger travel patterns in massive smart card data. *IEEE Transactions on Intelligent Transportation Systems*. 2017;18(11): 3135-3146.
- [20] Li M, Wang YH, Jia LM. The modeling of attraction characteristics regarding passenger flow in urban rail transit network based on field theory. *PLOS ONE*. 2017;12(9): e0184131.
- [21] McGuckin N, Nakamoto Y. Trips, chains, and tours: using an operational definition. *National Household Travel Survey: Understanding Our Nation's Travel (NHTS), 1-2 November 2004, Washington DC, USA*; 2004.
- [22] Weng JC, Wang C, Wang YY, Chen ZH, Peng S. Extraction method of public transit trip chains based on the individual riders' data. *Journal of Transportation Systems Engineering and Information Technology*. 2017;17(3): 67-73. Chinese.
- [23] Barry J, Newhouser R, Rahbee A, Sayeda S. Origin and destination estimation in New York City with automated fare system data. *Transportation Research Record*. 2002;1817: 183-187.
- [24] Liu MJ, Mao BH, Gao F, Guo JY, Gao LP. Analysis on commuter's activity chain choice behaviour. *Sixth International Conference of Traffic and Transportation Studies Congress (ICTTS), 5-7 August 2008, Nanning, China*. American Society of Civil Engineers; 2008. p. 222-230.
- [25] Yao XP, Zhao P, Han BM, Zhou QR. Home district identification for urban rail transit travellers by mining automatic fare collection data. *Journal of Transportation Systems Engineering and Information Technology*. 2016;16(5): 233-240. Chinese.
- [26] Kitamura R. Sequential, history dependent approach to trip chaining behaviour. *Transportation Research Record*. 1983;944: 13-22.
- [27] Stead D, Marshall S. The Relationships between urban form and travel patterns. An international review and evaluation. *European Journal of Transport and Infrastructure Research*. 2001;1(2): 113-141.
- [28] Mao ZD, Ettema D, Dijst M. Analysis of travel time and mode choice shift for non-work stops in commuting: Case study of Beijing, China. *Transportation*. 2018;45: 751-766.
- [29] Hsu CI, Guo SP. CBD oriented commuters' mode and residential location choices in an urban area with surface streets and rail transit lines. *Journal of Urban Planning and Development*. 2006;132(4): 235-246.
- [30] Sun YS, Schonfeld PM. Schedule-based rail transit path-choice estimation using automatic fare collection data. *Journal of Transportation Engineering*. 2016;142(1): 04015037-1-8.
- [31] Thorhauge M, Cherch E, Rich J. How flexible is flexible? Accounting for the effect of rescheduling possibilities in choice of departure time for work trips. *Transportation Research Part A: Policy and Practice*. 2016;86: 177-193.