2014 IEEE 17th International Conference on
Intelligent Transportation Systems (ITSC)
October 8-11, 2014. Qingdao, China

Understanding Temporal and Spatial Travel Patterns of Individual Passengers by Mining Smart Card Data

Juanjuan Zhao† Chen Tian⋆† Fan Zhang† Chengzhong Xu†§ Shengzhong Feng†
†Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China
⋆Department of Electronics and Information, Huazhong University of Science and Technology, China
§Department of Electrical and Computer Engineering, Wayne State University, MI, USA
†{jj.zhao, zhangfan, sz.feng, cz.xu}@siat.ac.cn, ⋆tianchen@hust.edu.cn

Abstract— **Metro systems have become the most preferred public transit services in many cities. It is important to understand individual passengers' spatio-temporal travel patterns inside metro. More specifically, for a specific passenger: what is the temporal access pattern? what is the spatio access pattern? is there any relationship between the temporal and spatio patterns? is this passenger's patterns normal or special? Answer all these questions can help us understanding the major reasons of why this passenger takes metro. In this paper, we analyze and understand the spatio-temporal travel patterns of individual passengers in Shenzhen, China. A systematic approach is proposed to extract temporal, spatial and anomaly features related to metro passengers. We analyze one month smart card data collected from Shenzhen. Combined with bus transaction data, we give an in-depth analysis and explanations for different groups.**

I. INTRODUCTION

Nowadays, metro systems have become the most preferred public transit services in many cities. Compared with other services, metro has the benefits of high efficiency, large volume, and fast speed. In Shenzhen (one of the four largest cities in China), the number of metro passengers reaches 2.5 million daily, which is around one-third of the total public traffic.

It is important to understand individual passengers' spatio-temporal travel patterns inside metro. More specifically, for a specific passenger: what is the temporal access pattern? what is the spatio access pattern? is there any relationship between the temporal and spatio patterns? is this passenger's patterns normal or special?

Answer all these questions can help us understanding the major reasons of why this passenger takes metro. The benefits are multi-fold as follows.

- *Policy Evaluation:* An important application is to evaluate the effects of a new policy. For example, a metro company considers the option of issuing monthly cards besides regular cards. The knowledge of individual passengers' mobility patterns can help estimating both traffic and income variation after introducing the new card type.
- *Anomaly Detection:* Abusing metro is an important problem in many cities. For example, beggars change cloths after passing the fare gantries, do their business, and change back before alighting the metro. However, their spatio-temporal patterns must be different from normal passengers, which can help metro company identifying those special passengers.
- *Social Networking:* The "familiar stranger" phenomenon is widely acknowledged in metropolitan area. One of our ongoing work is to develop a new social network by connecting passengers with similar public transportation patterns. Classify passengers, based on their different spatio-temporal patterns, is a critical step for scalable processing.

In this paper, we analyze and understand the spatio-temporal travel patterns of individual passengers in Shenzhen, China. The prevalence of smart card fare collection provides a unique opportunity for this study. By the end of 2013, the number of public transit smart card holders in Shenzhen has reached 10 million; these smart cards can be used for both bus and metro systems.

The contribution of this paper includes:

- A systematic approach is proposed to extract temporal, spatial and anomaly features related to metro passengers.
- We analyze one month smart card data collected from Shenzhen. Combined with bus transaction data, we give an in-depth analysis and explanations for different groups.

The rest of paper is organized as follows. Section II describes related work, data sources and data mining method. Section III presents an approach to extract temporal, spatial and anomaly features related to passengers. Passengers' temporal travel patterns and spatio-temporal travel patterns are analyzed in Section IV and Section V respectively. Section VI concludes the paper.

II. BACKGROUND

A. Related work

There are some studies of public transit systems by mining smart card data. Bagchi *et al.* analyze the advantages and disadvantages of smart card fare collection systems compared with traditional approaches [1]. Trepanier *et al.* propose an algorithm to infer passengers' get-off site from smart card data [2]. Sun *et al.* estimate the spatio-temporal density of passengers inside metro systems [3]. Kusakabe *et al.* try to estimate which train is boarded by passenger using transaction data [4]. Asakura *et al.* analyze how passengers change

978-1-4799-6077-4/14/$31.00 ©2014 IEEE

TABLE I: Transaction Record Format

Field	Value
CardID	unique identifier of smart card
TrmnlID	metro station id or bus id
TrnsctTime	Transaction time
TrnsctType	Transaction type(31-bus boarding, 21-metro swiped-in, 22-metro swiped-out)

travel behaviour to accommodate the change in the train timetable [5].

Liu *et al.* focus on the collective mobility patterns in time and space, also with smart card data in Shenzhen, China [6]; as a comparison, our paper focus on individual passengers' patterns. Morency *et al.* have several works closely related to our paper [7], [8], [9]. They use data mining technology to analyze the passengers' patterns in public transportation. Compared with them, our work (1) provides in-depth temporal and spatial analysis for individual travel patterns, (2) analyzes the relationship between temporal and spatial features, and (3) uses spatio-temporal analyse to perform abnormal detection.

B. Dataset

The dataset used in this study is the smart card transaction records in Shenzhen, China. The metro system has 5 metro lines by 2013. There are two types of smart card data used, one from the metro system and the other from the bus system. The additional bus data are required for in-depth analysis of the clustering results.

The whole data have more than 140 million transaction records, covering 21 consecutive weekdays from Nov 1 to Nov 30. Every time a passenger passes through metro gantry or boards a bus, a transaction is collected automatically. The data are collected from around 4 million smart cards: metro and bus account for 41.9 and 58.1 percent of the whole data respectively. The record format is shown in the Table I.

C. Data mining tools

Clustering is a data mining technique, which can find natural groups without any prior knowledge of relationships among the data. The goal of clustering is to ensure the inter-group similarity is as low as possible and the intra-group similarity is as high as possible. Clustering has been used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and so on. Among various clustering methods, we choose K-means as our classifier for both simplicity and efficiency.

III. Features Extraction

A. Data preprocessing

Every passenger can have many trips, hence the first step is to find all trips belong to a specific passenger. Each metro trip contains one boarding and one alighting records. From all transaction records, each pair of two records are joined together by matching smart card id and time.

The second step of preprocessing is to filter out passengers that rarely take metro. The distribution of passengers versus

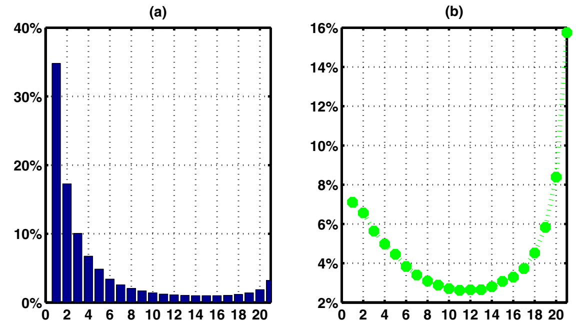

Fig. 1: The distribution of the number of passengers according to the number of active days

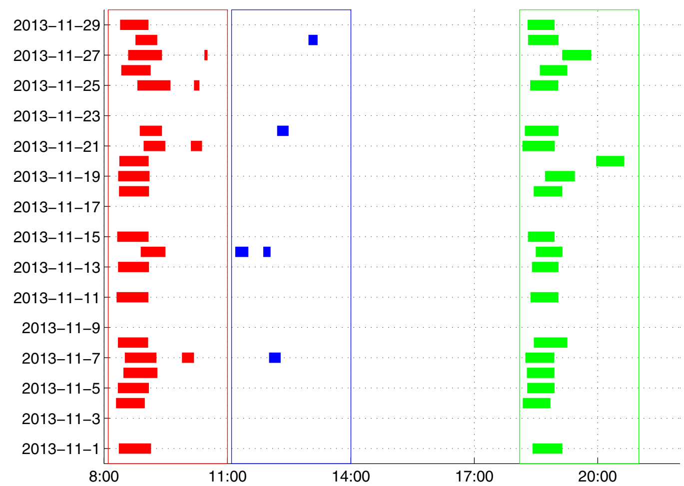

Fig. 2: The travel time of a specific passenger

the total number of active days of taking metro is shown in Fig. 1(a). It is clear that about 80 percent of passengers have active days of taking metro less than 7 within 21 weekdays. However, the top 20 percent passengers posses 68 percent of total transactions, as the distribution of transactions shown in Fig. 1(b). For passengers seldom travel, there is not enough information to reveal the temporal or spatial characteristic of them; in this paper we only analyze a passenger of his/her card's number of active days is more than 6.

B. Temporal Features extraction

Temporal analysis uses n features to express temporal characteristics of a passenger. There are two requirements. First, the chosen n features should convey temporal information as much as possible. Second, the n value should be as small as possible to improve the scalability of the analysis.

A typical passenger's travel time (*i.e.*, stay inside the metro system) is shown in Fig. 2. There are two observations. First, travels are regular in certain time periods: most days there are trips in two periods, *i.e.*, 08:00$\sim$10:59 and 18:00$\sim$20:59. Second, there are variations in each period: although there are always trips in the 08:00$\sim$10:59 period, the start time and duration still have variations.

The central idea of our temporal feature extraction is to divide time into sequential and overlapped slots. None-

	H1	H2	H3	H4	H5	H6	H7	...	H13	H14	H15	...
2013/11/1	0	0	1	1	0	0	0	...	1	1	0	...
2013/11/4	0	0	1	0	0	0	0	...	1	1	0	...
2013/11/5	0	0	1	1	0	0	0	...	1	1	0	...
2013/11/6	0	0	1	1	0	0	0	...	1	1	0	...
2013/11/7	0	0	1	1	1	0	1	...	1	0	0	...
...	...	...	...	...	...	...	...	...	...	...	...	...

S	1	2	3	4	5	6	7	...	13	14	15	...
T	20	20	20	19	7	4	4	...	20	20	13	...
F	20	39	43	24	9	6	5	...	31	32	14	...

P1			P2			P3			...		
T	F	S	T	F	S	T	F	S	...	...	...
20	43	3	20	32	14	4	6	6	...	...	...

Fig. 3: The steps of extracting temporal features

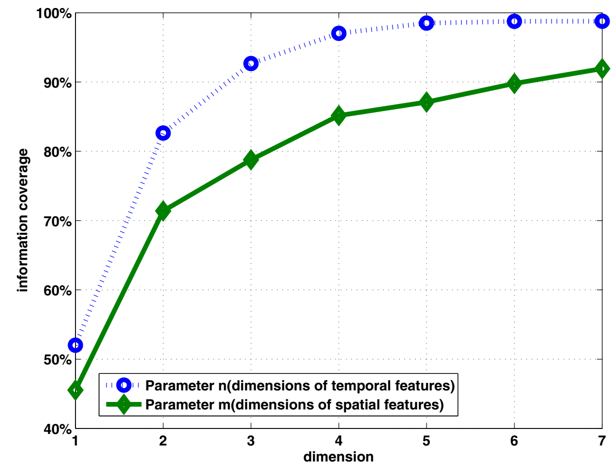

Fig. 4: The information coverage for parameter n and m

overlapped slots cannot expressively denote a duration: for example, it is hard to analyze a trip from 08:30∼09:29 by two consecutive hour slots such as 08:00∼08:59 and 09:00∼09:59 together. Instead, we use overlapped slot division such as 08:00∼10:59, 09:00∼11:59, *etc.* We choose the length of each slot to be 3 hours, since seldom a metro trip lasts more than 3 hours.

The process of temporal features extraction consists of three steps shown in Fig. 3.

step 1: For each card, the travel status of every hour every day is obtained. As the upper part of Fig. 3 shows, a matrix H is formed for every card: each row represents one weekday; each column represents one hour (H1 denotes 06:00∼6:59, H2 denotes 7:00∼9:59..., H19 denotes 00:00∼00:59); the values in this matrix tell whether the card is active in the corresponding one-hour period.

step 2: The middle part of Fig. 3 is the aggregation: S is the time slot sequence number, T is the active days in that slot, F is the active hours in that slot. Here each time slot is adjacent 3 hours. T and F are calculated using the following Equation (1) and Equation (2) respectively. D_{num} is the number of weekdays.

$$T_{1,i} = \sum_{j=1}^{D_{num}} (H_{j,i}|H_{j,i+1}|H_{j,i+2}). \tag{1}$$

$$F_{1,i} = \sum_{j=1}^{D_{num}} (H_{j,i}+H_{j,i+1}+H_{j,i+2}). \tag{2}$$

step 3: With T and F as the first and second order indexes respectively, we sort all slots as shown in the lower part of Fig. 3.

Now we need to fix the value of n, which indicates the number of temporal features. Given a certain threshold denoting the minimum coverage of information required, the selected n temporal features should provide information larger than the threshold. In this paper, we set threshold to 90%. From Fig. 4, we can set n to be 4 can meet the condition.

C. Spatial and anomaly features extraction

From the spatial perspective, we also want to extract m features, or more specifically, m OD (Origination-Destination) pairs. We first order the frequency of OD pairs in a descending order. The selection rules for parameter m is similar to temporal parameter n: the certain threshold denoting the minimum coverage of information required for spatial features are set to 85%. From Fig. 4, we conclude that m can be set to 4 in this paper.

There are two kinds of obvious abnormal trips. One anomaly is that for a specific trip, the time used from station A to station B is much more than other trips with the same OD pair. Assume the fastest passenger with zero-waiting time from A to B takes $L1$ time; we also denote the longest intervals between two consecutive trains is $L2$ (*i.e.*, the maximum waiting time) and an acceptable error denoted by $L3$. If a trip takes L time, then $L-(L1+L2+L3)>0$ is abnormal. The second anomaly is that the origin and destination stations are the same one. Such a behavior is contrary to the purpose of metro, which is travel from one place to anther.

Note that normal passengers also can have a few abnormal trips. What we need to focus is those passengers with a lot of abnormal trips. For every smart card, the probabilities for both type of anomalies are calculated, denoted by W and P respectively.

IV. Temporal Analysis

A. Clustering of metro data

We use the k-means algorithm to cluster all passengers based on their 4-dimensional temporal features. The results are 4 groups: $TGrp1$, $TGrp2$, $TGrp3$ and $TGrp4$. The centers of groups are shown in Fig.5. Clearly, passengers of $TGrp1$ have one dominant travel slot; passengers of $TGrp2$ have two dominant travel slots; passengers of $TGrp3$ have one relatively high dominant travel slot and one general travel slot; there are no significant difference among $TGrp4$ passengers' travel slots. They account for 15.83%, 32.64%, 27.11%, 24.41% of the total passengers respectively. For a

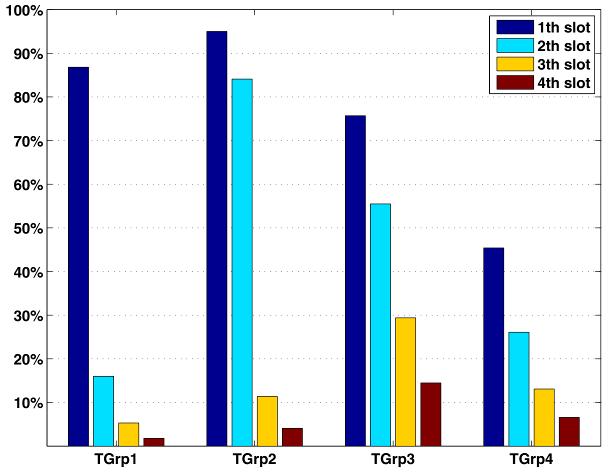

Fig. 5: Centers of $TGrp1$, $TGrp2$, $TGrp3$ and $TGrp4$

specific slot of a specific passenger, we defines p as the proportion of active days in the slot to all his/her active days; the slot is given a label among VF(very frequent), RF(relatively frequent), GF(general frequent), LF(less frequent), which satisfy certain conditions as shown in Table II; and the characteristics for $TGrps$ can be got from Table IV.

TABLE II: Standard for classification of p

Group	VF	RF	GF	LF
p	p>=80%	70%<=p<80%	50%<=p<70%	p<50%

First, let's analyze $TGrp2$. Those passengers regularly travel at two certain periods by metro: for each period, the number of active days by metro is more than 80 percent of all their own active days (the total number of days of taking bus and subway). Our investigation shows this group contains people who regularly commute back and forth to work or other activities at morning and evening peaks.

The $TGrp1$ passengers regularly travel at one certain period by metro; during that period the number of active days is more than 85 percent. The questions arise: if they go to work by metro, how do they come home? or vice versa, how do they go to work? Also, there is no distinct temporal regularity for $TGrp3$ and $TGrp4$. There are multiple possibilities: taking bus, private car, *etc.*

TABLE III: The characteristics of three groups

	1st	2nd	3rd	4th
TGrp-1	VF	LF	LF	LF
TGrp-2	VF	VF	LF	LF
TGrp-3	RF	GF	LF	LF
TGrp-4	LF	LF	LF	LF

B. Incorporate bus data

By incorporating bus data, we understand the above questions deeper. With bus data, we re-cluster all passengers using the same centers as $TGrp$ and there are still 4 groups $BTGrp1$, $BTGrp2$, $BTGrp3$ and $BTGrp4$. They account for 6.08%, 48.75%, 42.17%, 3.00% of the total passengers respectively.

TABLE IV: The joint distribution of $TGrp$ and $BTGrp$

Group	BTGrp-1	BTGrp-2	BTGrp-3	BTGrp-4	Total
TGrp-1	5.48%	7.78%	2.54%	0.04%	15.83%
TGrp-2	0.00%	31.70%	0.94%	0.00%	32.64%
TGrp-3	0.03%	3.55%	23.52%	0.002%	27.11%
TGrp-4	0.58%	5.73%	15.17%	2.94%	24.41%
Total	6.08%	48.75%	42.17%	3.00%	100%

We match the passengers to both $TGrp$ and $BTGrp$ in Table IV. For 15.83% passengers in $TGrp1$, now 7.78% move to $BTGrp2$: though metro is faster and punctual, but it is more expensive than bus; to save money and lose less time, they choose metro in one trip and bus in another trip according to their knowledge of traffic condition. There are still 5.48% passengers remain in $BTGrp1$: they could be divided into two categories, those who finished a round trip in three hours and those who chose other vehicle instead of bus or metro, such as company's commute vehicle to office, private car.

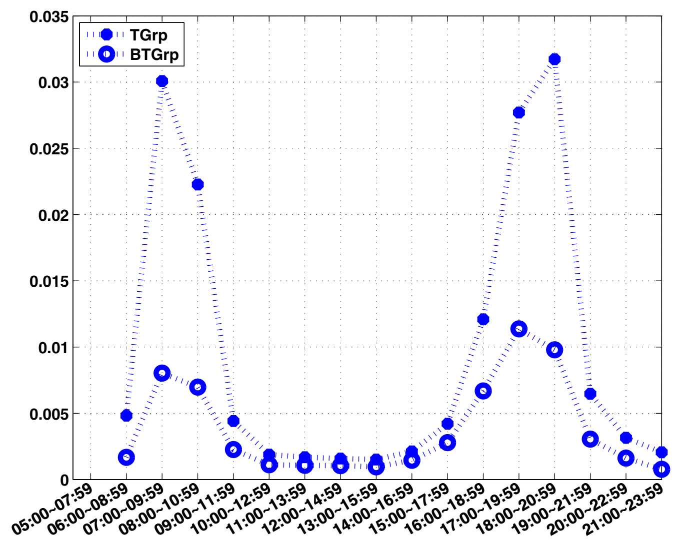

Fig. 6: The distribution of the first certain period of group 1

Fig.6 is the comparison of temporal distribution of the first certain period of group 1 between before and after combining bus data. As we can see, both curves have remarkable AM peek and PM peek hours, aside from a remarkable reduce in the number of passengers compared with the former, and the gap between peak hours' user numbers and off-peek hours' is narrow. This suggests that most of reduced passengers is traveling at peek hours, which accords with most users' habit.

Not surprisingly, nearly all $TGrp2$ passengers (*i.e.*, 31.7%) fall into $BTGrp2$; The comparison of the joint probability distribution of the first and second certain periods of passengers, between before and after reclassified, are shown in Fig.7. As we can see, the number of passengers traveling during AM and PM peek hours increase from 22% to 37% after being reclassified. The results confirm that most

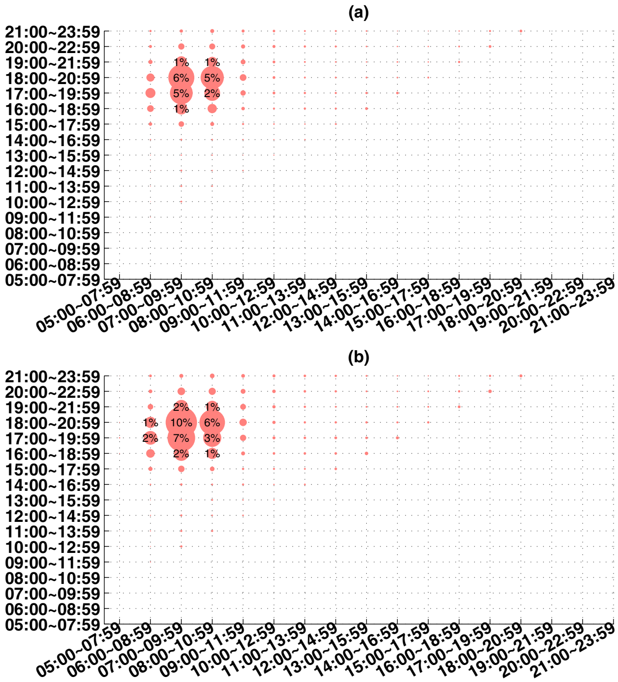

Fig. 7: Joint probability distribution of the first and second certain periods of group 2 (a)$TGrp$ (b)$BTGrp$

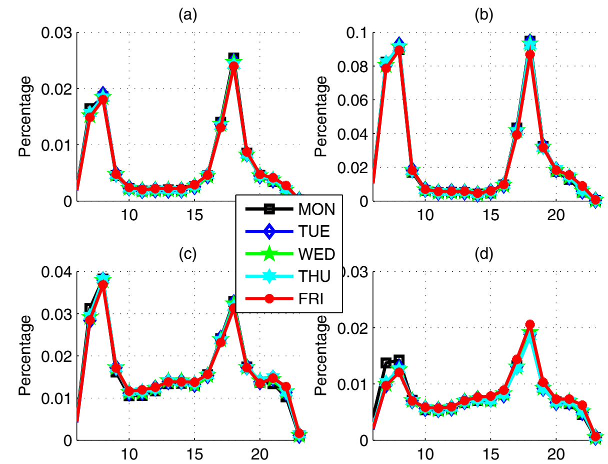

Fig. 8: Temporal distribution for (a)TGrp-1 (b)TGrp-2 (c)TGrp-3 (d)TGrp-4

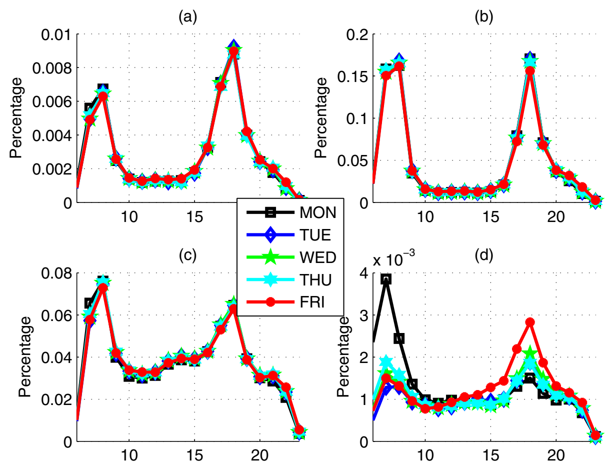

Fig. 9: Temporal distribution for (a)BTGrp-1 (b)BTGrp-2 (c)BTGrp-3 (d)BTGrp-4

passengers travel at two certain periods corresponding forth and back to work or other activities. We can also notice that there are some passengers who also regularly travel during two certain periods, but not during peek hours: the possible reason is that these passengers normally work shifts, such as as regular evening, rotating shift, or some other schedule.

In Table IV, for 27.11% $TGrp3$ passengers, most of them(*i.e.*, 23.52%) fall into $TGrp3$: those passengers mainly rely on metro for round trip, but travel time of them is somewhat irregular. Our investigation shows that some of them often works overtime or have a lot of leisure activities; and a few passengers (*i.e.*, 3.55%) fall into $TGrp2$: those passengers regularly travel during two certain periods, they sometimes travel by metro and sometimes travel by bus.

In Table IV, for 24.41% passengers in $TGrp4$, most passengers (*i.e.*, 20.9%) fall into $BTGrp2$ and $BTGrp3$: those passengers mainly rely on bus for round trip. The rest of the passengers either rely on their private car, or have no regular life patterns.

The comparison of temporal distribution for passengers in every group in weekdays before and after reclassified can be got from Fig.8 and Fig.9. A remarkable similarity exists from Monday to Friday for group 1, group 2, group 3 as shown in Fig.8(a), (b), (c) and Fig.9(a), (b), (c): the curves of five weekdays come closest to coinciding in shape; although for a specific passenger in group 3, the travel time is somewhat irregular, but for the whole is regular. morning and evening peek hours of group 1 and group 2 are more apparent than group 3 and group 4, because most passengers in group 1 and group 2 are traveling to and from work in AM and PM hours; the probability distribution function of travel time in group 2 approximately obeys gaussian; There are somewhat difference for group 4 in different day as shown in Fig.8(d) and Fig.9(d) because the passengers in group 4 have no regular travel time.

V. Spatio-temporal analysis

A. Pure spatial analysis

Metro Passengers can be spatially clustered into 4 groups using k-means algorithm on the 4-dimensional spatial features; the centers of *SGrp* 1, *SGrp* 2, *SGrp* 3 and *SGrp* 4 are shown in Fig. 10. *SGrp* 1 passengers have only one frequently accessed *OD*-pair, of which the number of access days amounts to 80% of all their own active days. *SGrp* 2 passengers have two frequently accessed od-pairs, of which the number of access days are both greater than 80% of all their own active days. *SGrp* 3 has one relatively frequently

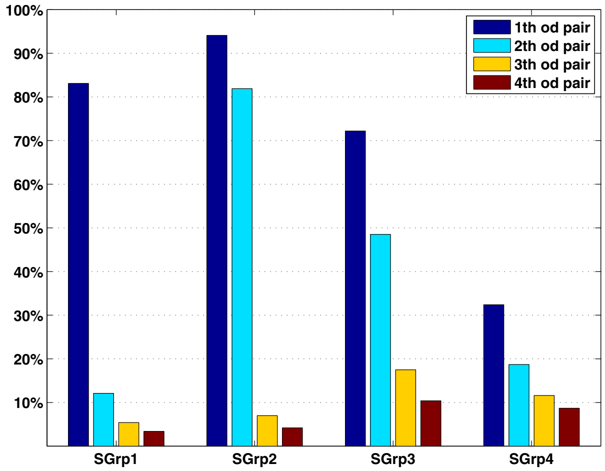

Fig. 10: Centers of *SGrp* 1, *SGrp* 2, *SGrp* 3 and *SGrp* 4

accessed od-pairs and one general accessed od-pair, There is no remarkable frequently accessed od-pair for *SGrp* 4.

B. Relationship between SGrps and TGrps

TABLE V: The conditional probability of *SGrp* given *TGrp*

Group	*SGrp*-1	*SGrp*-2	*SGrp*-3	*SGrp*-4	Total
TGrp-1	72.8%	2.2%	10.7%	14.4%	100%
TGrp-2	1.0%	70.1%	27.6%	1.3%	100%
TGrp-3	5.9%	15.3%	48.2%	30.7%	100%
TGrp-4	6.8%	0.2%	9.5%	83.6%	100%

The conditional probability of *SGrp* given *TGrp* is shown in Table V. We can observe that a *SGrp* has strong connections with the corresponding *TGrp*, as 72.8%,70.1%,83.6% of temporal group *TGrp*1, *TGrp*2 and *TGrp*4 belong to the corresponding spatial group *SGrp*1, *SGrp*2 and *SGrp*4. due to the complexity of *TGrp*3's travel slots, it mainly belongs to two groups: *SGrp*3 and *SGrp*4. The implication is that most passengers in group 1 and group 2 are spatio-temporally regular; most passengers in group 3 are relatively spatio-temporally regular; as a comparison, most passengers group 4 are irregular both temporal and spatial.

TABLE VI: Certain conditions for two classes of passengers

class	*TGrp*	*BTGrp*	*SGrp*
class-1	1	2	1
class-2	2	2	2

So far, there are three group labels for every metro user, temporal label, super temporal label (combining with bus data) and spatial label, denoted by *TGrp*, *BTGrp*, *SGrp* respectively. Although metro has many obvious advantages, why some passengers choose metro in a single trip and choose bus in another trip, instead of metro in round trips? We suspect that there some economic reasons.

In order to explain this case, firstly we select two classes of passengers satisfying certain conditions as shown in Table VI. Class-1 represents those who take metro in one trip and take bus in another trip; class-2 represents those who take metro in round trips. The comparison of average cost, travel time in five weekdays are shown in Fig.11. It is clear that first class is less than the second class in cost, travel time. It was learned that the average cost of taking metro is higher than taking bus in Shenzhen, and nearer distance from source to destination is in general more easily to find a no-transfer bus or a satisfactory transfer routine of bus, so for economic reasons, if time permitting, some passengers will choose bus.

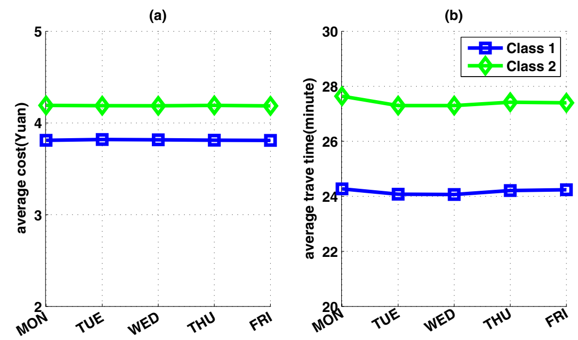

Fig. 11: Average cost and average travel time

C. Anomaly analysis

The passengers with abnormal trips or *OD* pairs are analyzed in this part. The abnormal variable W and P are defined before: W is the ratio of abnormal travel time trips of a passenger, and P is the ratio of abnormal *OD* pairs of a passenger. The distributions and the corresponding cumulative probability distributions of W and P are shown in Fig.12(a), (b) and Fig.12(c), (d) respectively. As we can see, for 98 percent passengers with travel time anomaly, the

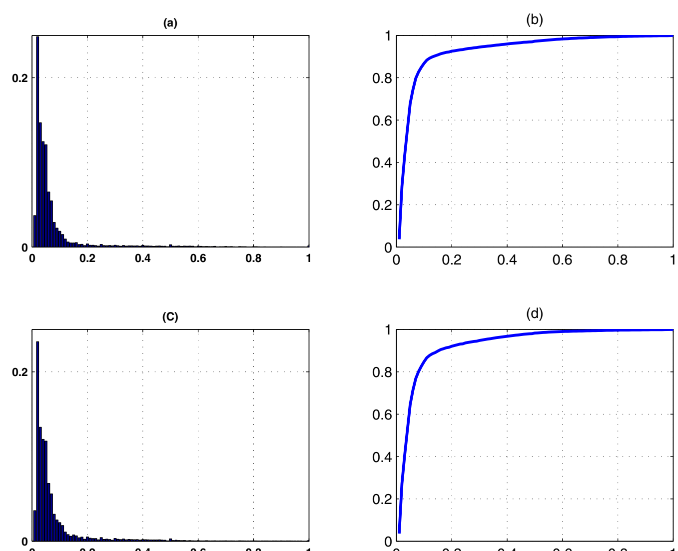

Fig. 12: (a) W distribution (b) W Cumulative frequency (c) P distribution (d) P Cumulative frequency

value of W is less than 40%, and for 98 percent passengers with OD pairs anomaly, the value of P is less than 40% too.

TABLE VII: W * P Crosstabulation

	PB=0	PB=1
$WB=0$	20485176	82597
$WB=1$	110464	70916

We also analyze the correlation between W and P. we suppose the acceptable range for anomaly RATIO is 40 percent, two values of W and P are converted into WB and PB: if $W > 0.4$, then $WB = 1$, else then $WB = 0$; similar is the relationship between P and PB. The result is shown in four fold Table VII: the test results suggest that there are strong relation between W and P.

D. Anomaly analysis

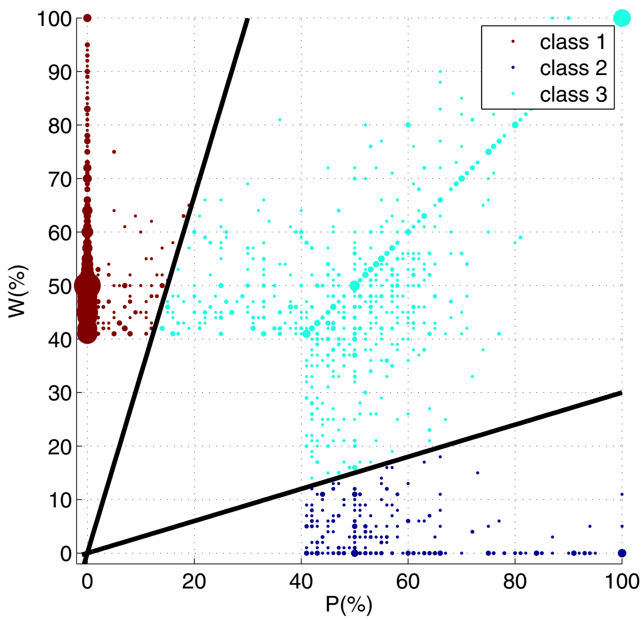

Fig. 13: Groups of passengers with anomaly travel behaviors

TABLE VIII: Certain conditions for three groups of passengers with anomaly travel behaviors

class	condition	anomaly with
class-1	a>=5	travel time
class-2	a<=1/5	OD
class-3	other	travel time and OD

The bivariate scatter plot of W and P is shown in Figure 13. The passengers can be classified into three groups, anomaly with travel time, anomaly with OD, anomaly with both travel time and OD respectively, which satisfy the condition shown in Table VIII, the variable a is defined as W/P. Why those anomaly exists? According to a survey with metro authority, we learn that: there are some passengers who involve in express logistics works; they choose metro as their vehicle; they deliver the goods to the nearest train station to destination, pass it to their colleagues who wait outside the station and will send the goods to the destination. The motivation is that they do not need to go out of the station in the entire process, as the money costed is fixed given a certain OD pair, there is no extra expense; the same station of boarding with departure cost lest, that is also the reason why some passengers board and depart at same station. Another reason is the beggar group: they actually work inside metro. All these abuse of metro should be prevented, or, at least restricted. Our mining method could help metro authorities to identify those special passengers.

VI. Conclusion

This paper analyzes and understands metro users from different angles. It is performed with smart card data collected from Shenzhen metro system and the benefits are multi-fold. We have an important understanding that: if a passenger is temporally regular, it is very possible that the passenger is also spatially regular. Also, we find some abnormal travel patterns that need to be noticed by the metro authorities.

Acknowledgment

The authors would like to thank anonymous reviewers for their valuable comments. This work is supported in part by “National Natural Science Foundation of China (No. 61100220, No. 61202107, No. 61202303)" and by the “Fundamental Research Funds for the Central Universities", and by NSF under grant CCF-1016966.

References

[1] M. Bagchi and P. White, “What role for smart-card data from bus systems?” *Municipal Engineer*, vol. 157, no. 1, pp. 39–46, 2004.

[2] M. Trépanier, N. Tranchant, and R. Chapleau, “Individual trip destination estimation in a transit smart card automated fare collection system,” *Journal of Intelligent Transportation Systems*, vol. 11, no. 1, pp. 1–14, 2007.

[3] L. Sun, D.-H. Lee, A. Erath, and X. Huang, “Using smart card data to extract passenger’s spatio-temporal density and train’s trajectory of mrt system,” in *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*. ACM, 2012, pp. 142–148.

[4] T. Kusakabe, T. Iryo, and Y. Asakura, “Estimation method for railway passengersąŕ train choice behavior with smart card transaction data,” *Transportation*, vol. 37, no. 5, pp. 731–749, 2010.

[5] Y. Asakura, T. Iryo, Y. Nakajima, and T. Kusakabe, “Estimation of behavioural change of railway passengers using smart card data,” *Public Transport*, vol. 4, no. 1, pp. 1–16, 2012.

[6] L. Liu, A. Hou, A. Biderman, C. Ratti, and J. Chen, “Understanding individual and collective mobility patterns from smart card records: A case study in shenzhen,” in *Intelligent Transportation Systems, 2009. ITSC’09. 12th International IEEE Conference on*. IEEE, 2009, pp. 1–6.

[7] C. Morency, M. Trepanier, and B. Agard, “Measuring transit use variability with smart-card data,” *Transport Policy*, vol. 14, no. 3, pp. 193–203, 2007.

[8] B. Agard, C. Morency, and M. Trépanier, “Mining public transport user behaviour from smart card data,” in *12th IFAC Symposium on Information Control Problems in Manufacturing-INCOM*, 2006, pp. 17–19.

[9] C. Morency, M. Trépanier, and B. Agard, “Analysing the variability of transit users behaviour with smart card data,” in *Intelligent Transportation Systems Conference, 2006. ITSC’06. IEEE*. IEEE, 2006, pp. 44–49.