Understanding individual and collective mobility patterns from smart card records: A case study in Shenzhen

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1109/ITSC.2009.5309662">http://dx.doi.org/10.1109/ITSC.2009.5309662</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Institute of Electrical and Electronics Engineers</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Tue Apr 02 08:20:37 EDT 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/58794">http://hdl.handle.net/1721.1/58794</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Understanding individual and collective mobility patterns from smart card records: 
a case study in Shenzhen

Liang Liu, Anyang Hou, Assaf Biderman, Carlo Ratti
SENSEable City Lab
Massachusetts Institute of Technology
Cambridge, MA, USA
{liuliang, enhanya, abider, ratti}@mit.edu

Abstract—Understanding the dynamics of the inhabitants’ daily mobility patterns is essential for the planning and management of urban facilities and services. In this paper, novel aspects of human mobility patterns are investigated by means of smart card data. Using extensive smart card records resolved in both time and space, we study the mean collective spatial and temporal mobility patterns at large scales and reveal the regularity of these patterns. We also investigate patterns of travel behavior at the individual level and show that the concentricity and regularity of mobility patterns. The analytical methodologies to spatially and temporally quantify, visualize, and examine urban mobility patterns developed in this paper could provide decision support for transport planning and management.

Keywords—mobility pattern; smart card; intelligent transportation system (ITS); visualization; regularity

I. INTRODUCTION

The city never sleeps. The human movement constitutes the pulse of the city. Observing and modeling human movement in urban environments is central to traffic forecasting, understanding the spread of biological viruses, designing location-based services, and improving urban infrastructure. However, little has changed since Whyte [1] observed in his "Street Life Project" that the actual usage of New York's streets and squares clashed with the original ideas of architects and city planners. A key difficulty faced by urban planners, virologists, and social scientists is that obtaining large, real-world observational data of human movement is challenging and costly [2].

In recent years, the large deployment of pervasive technologies in cities has led to a massive increase in the volume of records of where people have been and when they were there. These records are the digital footprint of individual mobility pattern. As websites have evolved to offer geo-located services, new sources of real-world behavioral data have begun to emerge. For example, Rattenbury et al. [3] and Girardin et al. [4] used geo-tagging patterns of photographs in Flickr to automatically detect interesting real-world events and draw conclusions about the flow of tourists in a city. In addition, as city-wide urban infrastructures such as buses, taxis, subways, public utilities, and roads become digitized, other sources of real-world datasets that can be implicitly sensed are becoming available. Ratti et al. [5], Reades et al. [6] and Gonzalez et al. [7] used cellular network data to study city dynamics and human mobility. McNamara et al. [8] used data collected from an RFID-enabled subway system to predict co-location patterns amongst mass transit users. Such sources of data are ever-expanding and offer large, underexplored datasets of physically-based interactions with the real world.

In this paper, we introduce a novel method for understanding human mobility in dense urban area based on millions of smart card records. We show how these data regarding the position and intensity of digital footprint can be used to infer cultural and geographic aspects of the city and reveal urban mobility pattern, which corresponds to human movement in the city.

In particular, the main contributions of this paper are: (1) demonstrating the potential of using smart card records as data sources to gain insights into city dynamics and aggregated human behavior; (2) exploring the relationship between spatiotemporal patterns of smart card usage and underlying city behavior and geography; and (3) studying patterns in smart card usage, including an analysis of how factors such as the time of the day affect this prediction.

We believe this work not only has direct implications for the design and operation of future urban public transport systems (e.g., more precise bus/subway scheduling, improved service to public transport users), but also for urban planning (e.g., for transit oriented urban development), traffic forecasting, the social sciences [9]—in particular, studying how people move about a city—and

Jun Chen
School of Transportation
Tongji University
Shanghai, China
chenjuntom@126.com
the development of novel context-based mobile services. In addition, we expect that similar types of analyses can be applied to other sources of urban digital traces such as those provided by parking management (e.g., San Francisco’s SFpark) and cellular networks [7]. Our work thus emphasizes the increasing role that data mining and visualization techniques will play to assist the aforementioned fields in analyzing traces of human behavior. Our work seems to open the way to a new approach to the understanding of urban systems, which we have termed “Urban Mobility Landscapes.” Urban Mobility Landscape could give new answers to long-standing questions in urban planning: how to map vehicle origins and destinations? How to understand the patterns of inhabitant movement? How to highlight critical points in the urban infrastructure? What is the relationship between urban forms and flows? And so on.

This paper is organized by the following sequences: in section II, we describe the data sets and the feature extraction process; in section III, we investigate the aggregated spatiotemporal patterns of urban mobility; in section IV, the individual mobility pattern is explored; in section V, we draw the conclusion and discuss the future work.

II. DATASETS DESCRIPTION

The datasets used to describe urban mobility patterns cover two major public transport modes, i.e. bus, subway. For smart card data, according to the survey conducted by the smart card company, there are 55% passengers using smart card in bus trip and 61% passengers using smart card in subway trip [10]. The smart card data is from 5 million smart card users’ transit records for one month, through December 1st, 2008 to December 31st, 2008. Every day there are 1.5 million transit records from the users. The smart card data description and sample is showed in table 1.

<table>
<thead>
<tr>
<th>Field</th>
<th>Memo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date</td>
<td>December 5th, 2008</td>
</tr>
<tr>
<td>Time</td>
<td>15:02:36</td>
</tr>
<tr>
<td>CardId</td>
<td>Anonymous unique user card id</td>
</tr>
<tr>
<td>Tradetype</td>
<td>Different transit behavior: 31- bus</td>
</tr>
<tr>
<td></td>
<td>boarding, 21- subway check-in, 22-</td>
</tr>
<tr>
<td></td>
<td>subway check-out</td>
</tr>
<tr>
<td>terminalid</td>
<td>For bus, it represents bus id; for subway,</td>
</tr>
<tr>
<td></td>
<td>it represents the stop id</td>
</tr>
<tr>
<td>trademoney</td>
<td>Final fare payment after discount (cent)</td>
</tr>
<tr>
<td>Tradevalue</td>
<td>The full payment/the original fare (cent)</td>
</tr>
<tr>
<td>transfer</td>
<td>Transfer discount (cent)</td>
</tr>
</tbody>
</table>

From the smart card data, we could infer two major public transit modes: bus and subway. For bus journey, we could get the boarding time and travel fare; for the subway journey, we could get the time and location of check-in and check-out, from which we could infer the OD feature of trip. Moreover, the transfer information could be derived from the smart card data.

Right now Shenzhen Metro Phase 1 is composed by east part of line 1 and south part of line 4. The east part of line 1 is from Luo Huo Railway station to Shijiezhichuang; the south of line 4 is from Futian Port to Shaoniangong. The total length of Shenzhen Metro Phase 1 is 21.866 kilometers, and there are 19 subway stops in total (the detailed information is showed in figure 1. The land use information is in table 2).

III. AGGREGATED SPATIOTEMPORAL MOBILITY PATTERN

Before exploring the implications of urban mobility landscape to urban planning, we discuss temporal and spatiotemporal patterns and highlight how these patterns reflect underlying cultural and spatial characteristics of Shenzhen.
**A. Temporal patterns of public transit passengers**

Through statistics of smart card records of the bus and subway in different day, their temporal mobility patterns could be inferred. The temporal pattern of bus trip and subway trip is showed in figure 1.

Figure 2(a) and 2(b) shows that during weekday there are obvious peak hours for bus trips. AM peak begins from 7am and reaches the peak at 8am; PM peak begins from 17pm and reaches peak at 18pm. This rhythm reflects the daily life patterns of citizens in Shenzhen, most of government institution and enterprises begin their work around 9am and finish their work around 18pm. The morning peak proportion is 26%, the evening peak proportion is 20%, the sum of two peaks achieves 46%, and the trips in peak hours almost occupy the half of daily travel demand. On Sunday, the peak hour is 10am, 14pm and 17pm, but in total is smooth, because the inhabitants can choose the free travel time on Sunday. The night activity is more frequent during night on weekend than on weekday, which means more citizens choose to go out for recreation. The morning peak of passengers on Saturday is still big, it is 10.2%, it means on Saturday a large proportion of citizens work during Saturday (From the proportion, it contribute half of daily working population), this reflects the corporation composition of Shenzhen, because most of HongKong, Taiwan and Japan enterprises ask employees to work half day on Saturday, their daily travel cause the peak hour on Saturday, and the small noon peak. In some sense, the travel feature of Saturday is the combination of weekday and Sunday, in the morning it likes the weekday and after the noon it is the same like Sunday. However, the Saturday peak hour is 18PM.

Figure 2(c) and 2(d) shows the temporal mobility pattern for subway. During weekday, the subway trip mobility pattern is almost the same with bus trip mobility pattern. The only difference is that subway AM peak hour is one hour after bus AM peak hour, which means subway is more reliable than bus. The inhabitants spend less travel time in subway than in bus.

**B. Spatiotemporal patterns of subway passengers**

Through detailed analysis of daily check-in and check-out in different subway stops, we could get the spatiotemporal patterns of subway travel. The proportion of check-in and check-out in different day is showed in Figure 3.

Figure 3 (a) and 3 (b) is basically the same, which means that the check-in and check-outs is almost similar. The proportion remains the same during different date and different bus stops, which means they have great similarity. The largest proportion subway stop is Dajuyuan, Huaqiang Road and Shijiezhichuang. The proportion of Saturday and Sunday are similar, the three largest proportion data is Laojie, Huaqiang Road and Shijiezhichuang.

From the proposition change of from weekday to weekend, Laojie Subway station(Dongmen Pedestrian Street) has largest increase, than is Huaqiang Road, Luohu Station and Futian Port station. Meanwhile, the proportion decreases in Guomao, Dajuyuan, Gouwu Gongyuan and Chegongmiao. This describes the life rhythms of citizens’ lives. The working place activity decrease during the weekend, while the two biggest recreation center- Laojie and Huaqiang Road increase the activity.

**C. Daily check-in and check-out of different subway stops**

From figure 3 we could know the unique mobility pattern for individual subway stop, however, what we are more concerned is the connection between different subway stops. While the most efficient way to measure the connections between different subway stops is deriving OD matrix. According to check-in and check-out timestamps and locations for each individual passenger, we can inter the OD matrix for arbitrary time period, which is a 19*19 matrix.

From the analysis results, we find the OD pair between different two points is in the same magnitude and there is little difference between them. So we sum the OD pair, and display them through GIS platform to compare directly. The summarized OD pair is showed in Figure 4.
From figure 4(a), it is easy to understand that the largest connections in weekday are Guangxia and Huaqiang Road, Shijiezhichuang and Huaqiang Road, Shijiezhichuang and Dajuyuan, Huangqiang Road and Laojie. The first three shows the working travel, the last one is the recreation. The working travel is more than the recreation travel.

Figure 4(c) shows the largest connections in Sunday are Laojie and Huaqiang Road, Shijiezhichuang and Huaqiang Road, Huaqiang Road and Gangxia. The recreation trips are more than working trips.

Figure 4(b) indicates the mobility pattern of Saturday is the mix of weekday and weekend, work travel and recreation travel. Though Saturday travel is less than weekday, but the connections are more concentrated, i.e. several stops have more connections than weekday.

From figure 4(a), it is easy to understand that the largest connections in weekday are Guangxia and Huaqiang Road, Shijiezhichuang and Huaqiang Road, Shijiezhichuang and Dajuyuan, Huangqiang Road and Laojie. The first three shows the working travel, the last one is the recreation. The working travel is more than the recreation travel.

Figure 4(c) shows the largest connections in Sunday are Laojie and Huaqiang Road, Shijiezhichuang and Huaqiang Road, Huaqiang Road and Gangxia. The recreation trips are more than working trips.

Figure 4(b) indicates the mobility pattern of Saturday is the mix of weekday and weekend, work travel and recreation travel. Though Saturday travel is less than weekday, but the connections are more concentrated, i.e. several stops have more connections than weekday.

D. Spatiotemporal patterns during the AM and PM peak hours

Through temporal analysis we can know the AM and PM peak mobility pattern. But how do they distribute in spatial scale and do they have obvious directions? Through spatial mapping of AM and PM peak flow, we can answer these research questions.

Generally, the check-in during AM peak hour is more close to the residential center, and the check-out is more close to the working zones. On the contrary, the check-in during PM peak is more close to the working zones and the check-out is more close to the residential area. Based on these assumptions, we could detect working zones and residential zones.

Through analyzing different proportion of check-in and check-out in AM peak 8am and PM peak 18pm, we could use simple figure to detect residential area and working area. Check-in(check-out) proportion in different stops in peak hour is showed in Figure 5.

From Figure 5, obviously Shijiezhichuang and Gangxia represents the residential centers, the two biggest communities, Shijiezhichuang represents western communities, including Nanshan district and Xin’an and Xi’xiang community in Bao’an district; Gangxia represents northern communities, including Meilin, longhua communities. While Guomao, Dajuyuan, Huaqiang Road, Gouwu Gongyuan, ChegongMiao are the center of working area.

Figure 5. Check-in(check-out) proportion in different stops in peak hour

Through analysis of different proportion during the peak hours, we can distinguish the residential area and working area, but we can not tell the directions of the travel. Thus we must calculate the related OD matrix and show it on GIS platform, from which we could obtain the direction of morning and evening direction. AM and PM peak OD in different day is showed in Figure 6.

From Figure 6 we could understand that the AM peak in weekday represents very obvious rules: from residential area to working area is unidirectional flow, and the main flow direction is from west to east. The center of trip generation is Shijiezhichuang and Gangxia, which represents the western and northern residential center.
The spatiotemporal pattern is also very obvious during PM peak in weekday, which is from working area to residential area, and the main flow direction is from east to west. The center of the trip generation is Huaqiang Road and Dajuyuan.

Figure 6. AM and PM peak OD in different day

In sum, if we summarize the travel OD in peak hours, we could find the daily mobility patterns is simple and clear, every morning the inhabitants move from residential area to working area, while in the evening they travel from working zone back to residential area, some of them choose to go to recreation area.

In all of the subway stations, Shijiezhichuang and Gangxia are the center of residential area, representing western and northern inhabitants; Guomao, Dajuyuan, Kexueguan, Huaqiang Road, Gouwu Park and Chegongmiao are the center of working zone; Laojie and Huaqiang Road are the center of shopping and recreation. Inside of these, Huaqiang Road has special statuses, which is the center of both working area and recreation area. The mobility patterns of peak hours show the uni-CBD model in Shenzhen, which causes the clock pendulum movement of urban citizens.

IV. THE INDIVIDUAL MOBILITY PATTERN OF SUBWAY TRANSIT PASSENGER

According to the subway trip record, we could calculate the each user’s subway stops in one week sampling period and the frequency that each user shows up. These patterns will help us understand the travel pattern of different individual inhabitants.

A. passengers reaching different number of stops

According to the smart card record, we summarize the stops that different subway users reach, and the result is show in figure 7.

From figure 7 we could know that, most of the users only commute between two stops, the percentage is 61.5%, while the stops is inside the 4 stops is above 94%, which describe most of people are only active in several points. If we use logarithmic coordinate, we can get figure 7(b). It shows that users and the number of stops have linear relationship (equation 5.3).

$$y(x) = 6.1246 \exp(-1.0471x)$$  \hspace{1cm} (1)

From the result we get previously, we can know that the transit system is in steady condition. The mean value of number of reaching stops in different people is a constant value. According to the maximum entropy principle, we can judge that it belongs to exponential distribution. It also means that most people just access to a small part of areas under the physical constrain.

Figure 7. the different user percentage in different stops

B. percentage in different stops which passengers visit

For each of the smart card user, in his different travel stops, the percentage of different stops is different. Thus, we calculate the percentage of different stops above 5 stops. The result is showed in figure 8. In this graph, we only list the stops which are less than 10 stops.

From figure 8 we could know that, most of the people are concentrated on the first two stops, the sum of top 2 reaches 52%, which means most of the travels are concentrated in a small part of areas, home and office. If we use the logarithmic coordinate again to redraw the data in figure 8(a), we can get Figure 8(b). It is almost a linear line. The regression result shows that the power law is nearly -1, the same as Zipf law. The equation is shown in equation (2):

$$y(x) = 0.4144x^{-1.0954}$$  \hspace{1cm} (2)

We know that if people have a constant proportion among different stations in a macro flow system, its distribution should be exponential distribution. Due to the physical constrains, most people will have a high proportion activity in a small area.
C. First Passage Time

The concept FPT (First Passage Time) means the time interval between two consecutive appearances in one location. In this research we calculate the time interval between the first appearance and the following appearance. Firstly we do not distinguish the arrivals and departures. The calculation result is shown in figure 9. To further illustrate the mobility pattern, we then calculate the FPT of departure behavior, which is shown in figure 10.

From the figure 9 we can see that FPT is very regular. The proportion is decreasing from 1 to 8 hours. The large proportion of 1-hour FPT means that the duration for an activity in one location is less than one hour. Beyond the 8-hour, we can see that a peak shows in 10-hour. This is the time interval between morning peak and evening peak of working days (18:00-8:00 = 10 hours. It means people go to office at 8:00 and leave office at 18:00). Other high FPT peaks show on the multiples of 24. Obviously, people have a regular mobility pattern in the location and the time.

From the analysis, we observe that the individual mobility pattern is extreme concentric and regular, in both spatial and temporal scale. The strong spatial-temporal regularity, such as a lot of trips are carried out between home and office, and the 24-hour cycle of activity. Planning and operation should consider these characteristics in transportation facility and service design.

V. Conclusion

Novel aspects of human mobility patterns were addressed by means of smart card data with time and space resolution. This allowed us to study the mean collective behavior at large scales and focus on the regularity of human mobility pattern. Through the analysis, we show the regularity of the mobility patterns. The strong relationship between mobility pattern and land use properties is investigated in this study, which could help us better plan the related facilities and services. We also investigated mobility patterns at the individual level. We observed that the individual mobility pattern is extreme concentric and regular, in both spatial and temporal scale.

The case study in Shenzhen using smart card records to understand urban mobility pattern is carry out and analysis results show that it is of great importance to understand the functioning of metropolitan mobility systems in order to enhance life quality, to protect the environment and to achieve sustainable development.

In the future, we would like to incorporate contextual features into our urban mobility landscape system, such as weather, season, special events (e.g. concerts or soccer matches), public transportation schedules and locations, and data from additional urban infrastructure (e.g., cellular networks). We also plan to fuse the different data sources to derive high level understanding of urban dynamics. The most important aspect of the research is to understand how these urban monitoring systems can become a tool for urban planning and policy making.

REFERENCES