EXTRACTING CLUSTERED URBAN MOBILITY AND ACTIVITIES FROM GEOREFERENCED MOBILE PHONE DATASETS

Yihong Yuan*, Martin Raubal
Department of Geography, University of California, Santa Barbara, USA, 93106
yuan@geog.ucsb.edu, raubal@geog.ucsb.edu

1. INTRODUCTION

Modeling human mobility and activities in urban systems has been a continuing research topic in several areas such as transportation planning and behavior modeling. However, most of the previous research is based on data acquired from travel diaries and questionnaires, which is a widely adopted data collection method when studying individual behavior (Yamamoto *et al.* 1999). Due to the limited number of people covered by travel diaries, these datasets fail to provide comprehensive evidence when studying the characteristics of the whole urban system, such as identifying clustering of urban mobility. Meanwhile, the development of information and communication technologies (ICTs) created a wide range of new spatio-temporal data sources (e.g., georeferenced mobile phone records). These datasets opened the way to a new paradigm in urban planning, i.e., Real-time cities (Ratti *et al.* 2007), as well as facilitating the study on behavior analysis and spatio-temporal data mining (Yuan and Raubal 2010; Miller 2009). Undoubtedly, these technologies are a major step forward in identifying and characterizing dynamic hotspots and clustering in urban systems.

In this research, we focus on extracting clustered human mobility and activities based on a mobile phone dataset from northeast China. The research is conducted from two perspectives: 1) dynamic clustering (such as the hourly mobility patterns) and 2) Points of interest (POIs) clustering (such as the home locations of residents). The second aspect is considered as an indirect result derived from the first one. There have been several studies on modelling urban dynamic patterns from mobile connection datasets (e.g., the real time Rome project at the MIT SENSEable Lab[1]), but our research focuses on extracting the implications of various clustering patterns, as well as relating these patterns to the distribution of urban infrastructures. These results would be very useful in updating environmental, urban and transportation policies. Moreover, the results can be used as informants of human activity including long-term choices such as where to live and short-term choices such as daily activity scheduling. In addition, it is highly helpful for policy makers to understand the characteristics of individual mobility with wide-spread ICT usage, as well as updating environmental and transportation policies.

2. DATASET

In this research we use a dataset from city A**, which is a major commercial and transportation center in northeast China. Figure 1 shows a basic road map of the city.

Figure 1. The basic road map of City A

The dataset covers over one million people and includes mobile phone connection records for a time span of 9 days. The data include the time, duration, and approximate location of mobile phone connections, as well as the age and gender attributes of the users. Table 1 provides several sample records. The real phone number, longitudes and latitudes are not shown in the table for reasons of privacy.

Phone #	**1350*********
Longitude	**126.*******
Latitude	**45.*******
Time	**14:36:24**
Duration	**12mins**
Receiver phone #	**1340*********

Phone #	Gender	Age
1350***** **	male	45

Table 1. Sample records from the dataset of city A

For each user, the location of the nearest mobile phone tower is recorded both when the user makes and receives a phone call, resulting in a positional data accuracy of 300m-500m. Note that the locations are recorded only when a phone call connection has been established; however, it still provides useful resources for depicting the general characteristics of individual mobility.

3. METHODOLOGIES AND PRELIMINARY RESULTS

3.1 Identifying dynamic clustering.

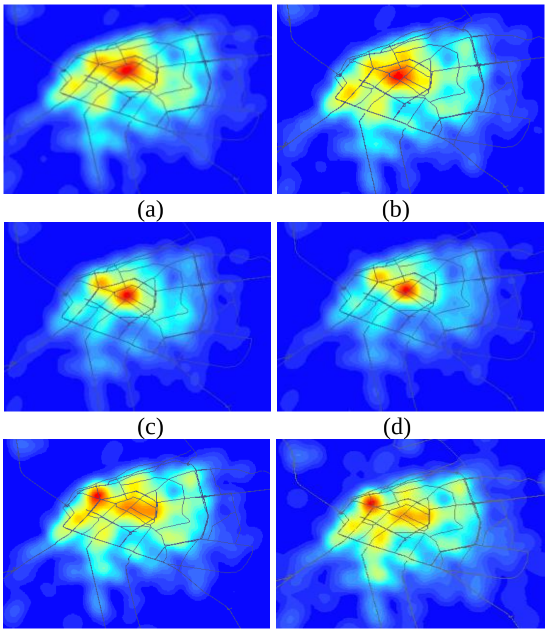
As stated in Miller (2009), identifying spatial clustering (e.g., high crime rate areas) is a major research topic in geographic knowledge discovery. In Figure 2, we locate the phone connections that occurred during three time periods (T_1: 8am-9am; T_2: 2pm-3pm; T_3: 7pm-8pm) on two separate days: one

---

* Corresponding author.
** The name of the city is not shown as required by the data provider.

[1] http://senseable.mit.edu/realtimerome/

weekday (07/23/2007) and one weekend day (07/21/2007), then we conduct a kernel density analysis to model the distribution of mobility density.

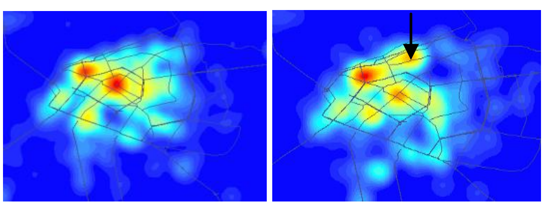


(a) (b) (c) (d) (e) (f)

Figure 2. Changing clustering of urban mobility in city A. (a) T_1, weekday; (b) T_1, weekend day; (c) T_2, weekday; (d) T_2, weekend day; (e) T_3, weekday; (f) T_3, weekend day

As indicated in Figure 2, the hotspots of urban mobility change for different time periods. In both T_1 and T_2, the densest clustering appears in the city center, whereas in T_3, the pattern is more spread over the city except for a small clustering in the northwestern part. This may indicate that the cluster in the Northwest represents resident home locations and we will further confirm this hypothesis in the later part of Section 3. Moreover, Figure 2 shows a high similarity of mobility patterns between weekday and weekend day for all three time spans.

Some clustering patterns are also different for various population groups (e.g., age, gender). Figure 3 shows the mobility hotspots for two age groups during 2-3pm on a weekday: teenagers (age 12-17) and seniors (age>60). As indicated in Figure 3, the clustering of teenagers appears both in the center and in the Northwest of the city, whereas the density pattern of seniors is more widely distributed. Due to the different demands for living resources in various population groups, such analysis can provide helpful references for updating urban infrastructures for different population groups (e.g., high schools, hospitals, etc.). For example, there appears to be a cluster of seniors in the north part of the city, the center of which is very close to a large park in City A.



(a) (b)

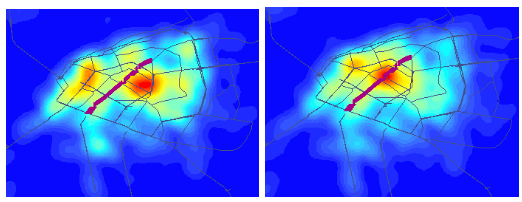
Figure 3. The clustering of (a) teenagers and (b) seniors

3.2 Identifying POIs clustering

In this research, POIs refer to regularly visited locations (e.g., home locations) of mobile phone users derived from their mobility patterns. Such analysis offers valuable input for enriching the personal profiles of users and studying urban areas according to their functions. Here we applied the methodologies described in Phithakkitnukoon *et al.* (2010) to identify the *stops* in trajectories: The trajectory of a certain individual is identified as a sequence of chronological locations:

$$R = \{(p_1, t_1) \rightarrow (p_2, t_2) \rightarrow .... \rightarrow (p_n, t_n)\}$$

Where the p_i refer to spatial locations and the t_i refer to time points. Then the trajectories are regrouped into sub-trajectories based on the restriction that any two consecutive points within a sub-trajectory are located within the cell of the same mobile phone tower. If the time duration of a sub-trajectory is longer than the temporal threshold ΔT, the sub-trajectory is identified as a *stop* for the particular user. Once the *stops* have been extracted, the home location of each user is estimated as the most frequent *stop* during the night hours and the work location is the most frequent stop during day hours on weekdays. Figure 4 demonstrates the distribution of home and work locations of users in City A. These POIs can also be combined with our previous research on user trajectory patterns to further examine the determinants of an individual's activity space (Yuan and Raubal 2010).



(a) (b)

Figure 4. Clustering of (a) home locations and (b) work locations

As shown in Figure 4, both home and work locations are clustered in the city center, however, there are slight differences between the locations of hotspots in Figure 4a and Figure 4b. The highlighted street in Figure 4 is one of the main streets and it runs across the whole city. The home locations are mostly concentrated on the southeast side of the main street, whereas the work locations are evenly distributed on both sides of the street. Additionally, the home locations show two clustering centers in the study area (one in the middle, the other on the western side), indicating that city A has multiple active sub-areas that function as residential districts.

4. CONCLUSIONS

The pervasive usage of mobile technologies highly facilitates the modeling of urban mobility from different perspectives (e.g., mobility flow between sub-areas, transportation mobility density, etc). In this research, we demonstrate that mobile information is highly effective in characterizing the dynamic clustering of urban mobility. Furthermore, we extracted the POIs based on user trajectories and discussed the distribution of work and home locations across the whole city. Although this research focuses more on the implications of clustering results rather than the methodologies of cluster detection, in a next step we will generate a dynamic cluster detection and cluster significance model to simulate the daily rhythm of human mobility in urban systems. Future research will also focus on correlating the mobility patterns with the distribution of various types of urban infrastructures. Other continuing research includes the modeling and predicting of spatio-temporal trends of urban activities based on existing mobility patterns, as well as characterizing the regularity of individual space-time paths.

REFERENCES

Miller, H., 2009. Geographic data mining and knowledge discovery: An overview. In: H. J. Miller and J. Han, Editors, *Geographic Data Mining and Knowledge Discovery (Second Edition)*, CRC Press, London, pp.3-32.

Phithakkitnukoon, S., T. Horanont, G. Di Lorenzo, R. Shibasaki and C. Ratti, 2010. Activity-aware map: Identifying human daily activity pattern using mobile phone data. In: A. A. Salah, T. Gevers, N. Sebe and A. Vinciarelli, Editors, *HBU 2010*, LNCS, Springer, Heidelberg, pp. 14-25.

Ratti, C., A. Sevtsuk, S. Huang and R. Pailer, 2005. Mobile Landscapes: Graz in Real Time. In *the 3rd Symposium on LBS & TeleCartography*. Vienna, Austria.

Yamamoto, T. and R. Kitamura, 1999. An analysis of time allocation to in-home and out-of-home discretionary activities across working days and non-working days. *Transportation*, 26, pp. 211–230.

Yuan, Y. and M. Raubal, 2010. Spatio-temporal knowledge discovery from georeferenced mobile phone data. In: B. Gottfried, P. Laube, A. Klippel, N. V. d. Weghe and R. Billen, Editors, *MPA'10 - 1st Workshop on Movement Pattern Analysis*, Zurich, Switzerland, pp. 121-126.

Yuan, Y. and M. Raubal, 2010. On Correlation between Mobile Phone Usage and Travel Behavior – A Case Study of Harbin, China (Extended Abstract). In: Geographic Information Science - 6th International Conference, GIScience 2010, Zurich, Switzerland.