Social Computing for solving Urban Challenges: Pushing the methodological frontier

Satish V Ukkusuri, Purdue University

The rapid development of Internet and Web 2.0 technologies, the global adoption of smartphones and social media applications, and the worldwide socio-economic and cultural trends have given rise to a novel rapidly-progressing computing paradigm, *social computing*, which takes a computational approach to modeling of social behavior and develops applications to support social interactions. Social computing applications can provide some novel datasets to understand urban systems such as transportation, water, building usage etc and user behavior and dynamics in a scale that was unimaginable before. In general, thus transportation modeling and analysis can significantly benefit from social computing methods, platforms and data.

Social computing approaches can be motivated by three major shifts in our thinking about traditional methods (e.g., econometric models) of transportation analysis and modeling. *First*, datasets from most social computing platforms cover a large number of individuals over many days providing the ability to analyze vast amounts of data instead of settling for small-scale datasets - accepting ``the unreasonable effectiveness of data'' (Halevy et al., 2009). *Second*, social computing approaches embrace the real-world messiness in the data (e.g., missing observations) rather than depending on comprehensive data and develops methods that can account for the noisiness in these data sets. *Third*, these approaches follow a data-driven modeling path focusing on predicting behavior rather than explaining behavior or drawing correlations between individual behavior and socio-demographic attributes.

The next few paragraphs describe few applications of social computing data and methods to solve transportation and disaster management issues in urban areas.

1.1.1. Understanding urban mobility and activity patterns

The introduction of location-based services in social computing applications has enabled people to share their activity related choices (check-in) in their virtual social networks. Users share their activity locations when they visit restaurants, shopping malls, movie theatres and so on. Users' locations are also available when they post status updates or media on their social networks. These data sources, containing instantaneous observations of the activity-locations, provide an unprecedented amount of user-generated data on human movement and activity participation. They have the potentials of revealing urban dynamics in real time. Researchers can see how different places of a city are used in the course of a day and thus visualize the urban pulse. With this data, we can find the spatio-temporal patterns of aggregate and individual mobility in the city (Hasan et al 2013); characterize the patterns activity participation behavior by overlaying activity information on geographic references, which are important problems related to mobility and activity behavior analysis.

1.1.2. Modeling and predicting behavior

Although the presentation of social media data on geographic spaces can provide interesting information, this data can provide further knowledge about human behavior. Most importantly this gives an opportunity for a large-scale analysis of human mobility behavior; such analysis opens the door for comparing activity behaviors of different urban areas potentially at the global level. With the availability of new activity related data sources from mobile phones and smartphones, there is a renewed interest in modeling individual behavior including activity and location choices over time. Using the longitudinal geo-location data available from social media it is now possible to develop complex models inferring individual behaviors. As location-based data sets from social media are becoming increasingly available, researchers have been using them to gather interesting insights on various behavioral aspects. Studies from social science, computer science, and transportation science have used innovative ways to extract meaningful patterns and model behaviors with diverse applications.

1.1.3 Land use prediction

Emerging location-based services in social media tools such as Foursquare and Twitter are providing an unprecedented amount of public-generated data on human movements and activities. This novel data source contains valuable information (e.g., geo-location, time and date, type of places) on human activities. While the data is tremendously beneficial in modeling human activity patterns, it is also greatly useful in inferring planning related variables such as a city's land use characteristics. Previous research (Zhan et al, 2014) a comprehensive investigation on the possibility and validity of utilizing large-scale social media check-in data to infer land use types by applying the state-of-art data mining techniques. Two inference approaches are proposed and tested in this paper: the unsupervised clustering method and supervised learning method. The land use inference is conducted in a uniform grid level of 200 by 200 meters. The methods are applied to a case study of New York City. The validation result confirms that the two approaches effectively infer different land use types given sufficient check-in data.

1.1.4 Disaster Analytics

Social Media data has been used for understanding various issues in disaster management including preparedness, response and recovery. Particularly, user location information has been used to understand the displacement of people immediately after the disaster and over time. For instance, data from 2013 Oklahama tornado was used to understand various issues during the disaster (Ukkusuri et al 2014). The potential of social media data in extracting relevant and useful information during natural disasters, and could potentially be an additional data source to better understand the individual behavior during crisis. The data records were categorized by user group and the most

frequently used words are ranked to track the variation of common interests for each user group. In addition, the data is classified into different content categories and their temporal variation patterns are analyzed. A sentiment analysis is conducted to quantify the sentiment in the data, which reveals variation of public mood and perception over time. The techniques used in this work can be readily used in analyzing similar major social crisis and natural disasters (e.g. hurricanes and earthquakes), which can at present provide valuable complementary information to users, first responders and emergency preparedness agencies in crisis awareness and response. Different stakeholders can determine the needs and activities of people during disasters by using the methodology proposed in this study with the help of social media data.

The presentation will discuss various methodological advances related to the use of social media data (both geo-located data and text data) for solving various research questions in the urban infrastructure space.

References

Halevy, A., Norvig, P., Pereira, F., 2009. The unreasonable effectiveness of data. *Intelligent Systems, IEEE* 24(2), 8-12.

Hasan, S., Schneider, C.M., Ukkusuri, S.V., Gonzalez, M.C., 2013a. Spatiotemporal Patterns of Urban Human Mobility. *Journal of Statistical Physics* 151(1-2), 304-318.

Hasan, S., Zhan, X., Ukkusuri, S.V., 2013b. Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. ACM, p. 6.

Ukkusuri, S.V., Zhan, X., Sadri, A.M. and Ye, Q. (2014). Use of social media data to explore crisis informatics: Study of 2013 Oklahoma tornado. Transportation Research Record: Journal of the Transportation Research Board, Vol. 2459, pp. 110-118 .

Zhan, X., Ukkusuri, S.V. and Zhu, F. (2014) Inferring urban land use using large scale social media check-in data Networks and Spatial Economics, Vol. 14(3), pp. 647-667.