

ABSTRACT

Zhan, Xianyuan. M.S.E., Purdue University, December 2012. Understanding the Aggregate Level Urban Activity Pattern Using Large-Scale Geo-Location data. Major Professor: Satish Ukkusuri.

In recent years the introduction of location-based services in online social networks have provided unprecedented amount of public-generated data on human movements and activities. This novel data source is available for free and provides a new dimension of information related to human activities with greater details.

Investigating urban activity pattern is crucial for understanding the driving force behind human movements in urban areas. Most of previous studies focus on understanding and modeling human mobility using data collected from mobile phones, bank notes movements, and subway smart-card transactions etc. However, these studies fail to link human movements with their purposes (activities). Although other studies use the combination of mobile phone data and Point of Interest (POI) data collected from internet, the information from POI for each place does not necessarily correspond to the actual activity that individual is participating in. In this research, the location-based information (“check-ins”) shared in an online social media, i.e., Twitter, is used to explore and interpret the aggregate level urban activity pattern. With detailed activity categories in each data record, we are able to discover more realistic and detailed descriptions of human dynamics and urban activity characteristics.

The first goal of this research is to investigate the urban aggregate level activity pattern. In this work, the individual check-in behavior is analyzed, which shows that the individual number of check-ins follows a truncated power law and the individual radius of gyration follows a Weibull distribution. The temporal characteristics are also investigated, which reveals distinct patterns in different activity categories. In order to study the data in a more detailed level, a virtual grid is constructed by dividing the city map into square cells of 200 by 200 meters. Using the grid system, we can

visualize the spatial distribution of popular places, and compute the site ranking distribution for the cells. These spatial distributions show important features about the unique activity patterns for each city. The kernel density estimation technique is also applied to reveal the evolution of activity centers in both time and space. Same analysis is conducted in New York, Chicago, Los Angeles and San Francisco, where city level activity patterns are compared. The findings show the existence of underlying laws and patterns that govern human urban activities, which enhance our understanding about the urban dynamics and human movement in the urban area.

The second goal of this study is to explore the possibility of using this new type of data to infer the urban land use. A new approach of using unsupervised clustering algorithm and social media check-in data is proposed. Data is aggregated and normalized by time periods and activity categories for each cell. K-means algorithm is then used to cluster all the cells into 4 classes. Optimal number of clusters and empirical verification confirm that this approach can effectively predict the residential, commercial, open space and recreation sites with reasonable accuracy level. This shows the potential of using large scale geo-location data in urban land use inference.