ABSTRACT

Hasan, Samiul , Purdue University, July 2013. Modeling Individual Activity Behavior and Dynamic Traffic Patterns using Large-scale Geo-location Data. Major Professor: Satish V. Ukkusuri.

The widespread use of social media provides extraordinary amounts of user-generated data every day. The recent introduction of the location-based services in smartphone-based social media applications allows people to share their activity related choices at the level of specific geo-referenced location and time. This geo-location data offers us, in new ways, people's attitudes, interests and activity patterns over a wide network and over multiple months/years that was unimaginable before. However, while this new data is available, there are currently limited methodologies that allow us to analyze these large data sets to obtain *useful information* by characterization and analysis of individual level activity patterns and network level dynamics that are important for understanding urban dynamics. This dissertation presents a new paradigm to activity-travel behavior analysis by analyzing this large-scale geo-location data, recording human movements over many days, to understand long-term individual mobility patterns, infer short-term individual activity and lifestyle patterns and estimate network traffic congestion.

The overarching goal of this dissertation is to *develop data analytics to understand urban dynamics and user behaviors using large-scale geo-located social media data*. It develops some novel statistical estimation techniques, motivated from machine learning approaches, for understanding the spatiotemporal patterns of urban activities. *First*, long-term patterns of human mobility are described by some statistical properties obtained through data analysis. A physics-based model is developed to explain these properties which contradicts sharply with the existing mobility models. *Second*, in this thesis, data-driven techniques for modeling users' activity patterns and

life-style choices are developed. *Third*, a novel method is developed to infer individual activity type, its duration and location, and the sequence of the activities from incomplete trajectory data. When aggregated these activity-location sequences are indicators of travel demand within a region. The potential of geo-location data to derive dynamic traffic patterns in transportation networks is investigated. *Fourth*, this thesis explores the social dimension of activity participation behaviors and measures the role of social influence in user activity location choices.

This thesis provides a holistic understanding of urban mobility and activity patterns by analyzing large-scale geo-location data which has limitations including absence of socio-demographic characteristics of the users, missing activities and unknown activity duration. It develops some novel methods motivated by machine learning approaches addressing these limitations of geo-location data for inferring individual-level patterns. These methodologies may become more useful and relevant for transportation analysis in near future with the availability of mobility data collected from cell phone call recordings, smart phone GPS recordings or check-ins from location-based applications in social media. Given the lack of appropriate methodologies to characterize this information, this thesis expects to make a novel and significant contribution to activity behavior and transportation system analysis.