Using Large Scale Taxicab Data to Estimate Link Travel Time, Predict Demand and Measure System Efficiency

Xianyuan Zhan, Research Assistant, Purdue University, Xinwu Qian Research Assistant, Purdue University
Satish V. Ukkusuri, Professor, Purdue University

Introduction
- The era of big data
  - Advance in sensing technologies
  - Development of large scale pervasive computing infrastructure
- Big data and transportation engineering
  - Reconsider traditional research problems
  - Make infeasible problems feasible
- In this work
  - Using large scale taxi data from NYC
  - Taxi Ridership analysis
  - Link travel time estimation
  - Taxi system efficiency

Key Findings
- Urban form has significant impact on ridership
- GWR explains up to 90% of the variance and achieves good prediction
- Both coefficients and t-stats of determinants vary over space
- Failing to consider spatial variation will result in erroneous estimations of determinants

Spatial Variation of Taxi Ridership

Motivation
- Statistical analysis of taxi ridership
- Trips are varying spatially
- The effects of determinants is nonhomogeneous

Methodology
- Geographically weighted regression
  \[ w_i = \exp\left(-\frac{d_i^2}{\sigma^2}\right) \quad \text{for} \quad \sigma = 0.4 \]
- Dependent variable: Taxi ridership
- Independent variables: commuting time, population, land use, median income, road density, subway accessibility

Efficiency of Taxi Service System

Motivation
- Vacant taxi trips lead to unnecessary externalities
- How to quantify the efficiency of the system performance and how far is the current system from the theoretically optimal one?

Key Findings
- Algorithm converges rapidly and entire estimation takes less than 15 minutes
- Robust estimation results: MAPE controlled under 30%
- Can be extended easily as a Bayesian mixture model by making use of historical data

References