NBER Digitization Tutorial Application

Yingjie Zhang

Contact information

Ph.D. Candidate in Information Systems and Management

Heinz College, Carnegie Mellon University

Address: Hamburg Hall 3041, 5000 Forbes Ave, Pittsburgh PA, 15206

Email: yingjie@cmu.edu

Graduate-level Coursework

The courses I took during my PhD focus on the economic modeling and quantitative methods. The following lists the major courses I took (from 2013-2016) in two streams:

Doctoral Level Economic and Econometrics

PhD Microeconomics
Computational Methods for Economics
Introduction to Econometric Theory
Econometric Theory and Methods
Game Theory and Applications
Advanced Economic Analysis (Structural Model)

Advanced Quantitative Method and Statistics

Empirical Models in Marketing
Statistics Methods for Social and Policy Research
Analytical and Structural Marketing Models
Bayesian Statistics in Marketing
Intermediate Statistics
Seminars on Estimating Structural Models

Research

I apply quantitative methods to answer questions related to information technologies. A broad agenda of my research is to use real-world *trajectory* data to analyze users' *digital* and *physical* behavior under a social-cyber-physical system. I achieve my research goals with interdisciplinary approaches, combining econometrics, structural modeling, Bayesian modeling, with multiple machine learning and temporal-spatial data mining techniques. A social-cyber-physical system merges society, computing, with physical systems to create better quality of life. Within this system, users' *digital* behavior is presented on multiple online web services and mobile

platforms, while users' *physical* behavior fastens on the traditional offline world. I listed three examples of my complete papers within this stream in the following:

My paper, *Modeling User Engagement in Mobile Content Consumption with Tapstream Data and Field Experiment*, studies the user behavior on the mobile platform. Specifically, I proposed a structural econometric framework for modeling of consumer latent engagement stages. This model accounts for both the time-varying nature of engagement and consumer forward-looking consumption behavior. The estimated results show that with our detected engagement stages, a personalized engagement-based pricing strategy leads, simultaneously, to lower average prices for consumers and higher overall business revenues for the app. To further evaluate the effectiveness of our method, I conducted a randomized field experiment on this mobile reading application platform. The experimental results provide more causal evidence that a personalized promotion strategy targeting at different user engagement stages can both decrease costs and enhance the overall business performance.

Second, I am also interested in the analytics of users' physical behavior by taking advantages of sensor technologies, data management and analytic models. In the paper, *Learning Individual Behavior Using Sensor Data: The Case of GPS Traces and Taxi Drivers*, I study the decision-making behavior of taxi drivers to find passengers when the taxis are vacant. With a Bayesian learning framework, we modeled three signals (i.e., pick-up, drop-off and drive-by signals) in drivers' learning process of taxi demand over space and time. To empirically study this, I parsed a large-scale dataset containing 11,196 taxi drivers' GPS trace and trip records from a large Asian city. Interestingly, the policy simulations indicate information that is noisy at the individual level becomes valuable after being aggregated across various spatial and temporal dimensions. This work allows us not only to explain driver decision-making behavior using these detailed behavioral traces, but also to prescribe information sharing strategy for the firm in order to improve the overall market efficiency.

Rich literature can be found respectively in users' digital and physical behavior, but few has been done with regards to the *links* between these two. This is crucial because nowadays, users' digital and physical behavior are inherently connected within the social-cyber-physical system. One simple example is the mobile check-in behavior in restaurants. My first attempt explores this area from the economic perspective. The paper, *Understanding User Economic Behavior in the City Using Large-scale Geotagged and Crowdsourced Data*, examines the economic value of the urban cities from crowdsourced and geotagged data. Specifically, I extracted multiple traffic and human mobility features from publicly available data sources (e.g., Twitter, Foursquare, and city planning websites) using natural language processing and geo-mapping techniques, and examined the effects of both static and dynamic features on economic outcomes of local businesses. This study demonstrates the potentials of utilizing such crowdsourced and geotagged data to create matrices to predict local economic demand in a manner that is fast, cheap, accurate, and meaningful.