

DISSERTATION DEFENSE

Xiao Liu

"Marketing Solutions Enabled by Big Data Using Structural Models"

Wednesday, April 22, 2015

11:00 am

147 Posner Hall

Digital marketing has brought in enormous capture of consumer data. In quantitative marketing, researchers have adopted structural models to explain the (dynamic) decision process and incentives of consumers. However, due to computational burdens, structural models have rarely been applied to big data. Machine learning algorithms that perform well on big data, however, hardly have any theoretical support. Equipped with both economics perspective and machine learning techniques, in this dissertation, I aim to combine rigorous economic theory with machine learning methods and apply them to data on a large scale.

First of all, I am deeply rooted in economic theory in understanding the behaviors of consumers, firms and the entire market. In my first essay "An Empirical Analysis of Consumer Purchase Behavior of Base Products and Add-ons Given Compatibility Constraints", I model consumers as forward-looking utility maximizers who consider the product bundle of base products and add-ons jointly. I derive the add-on-to-base effect from the solution of the dynamic programming problem and then quantified it by model estimates. The underlying theory of switching cost caused by incompatibility constraints helps me rationalize the high market share of Sony's despite its high price. Without this theoretical foundation, I could not explain the data puzzle and might have fallen into false causality.

Doing empirical research in this time of an explosion of the quantity and quality of data is fortuitous. With the help of "Big Data", I am able to explore new areas of research, like social media. In my second essay "A Structured Analysis of Unstructured Big Data Leveraging Cloud Computing", I conduct analysis on a staggering volume of nearly two billion tweets to predict TV show ratings. Different from the traditional economics data that can be easily stored in a spreadsheet, the unstructured format and sheer volume of the tweet data challenges traditional information extraction and selection methods. But after careful sifting the data through a combination of methods from cloud computing, machine learning and text mining, the rich content information imbedded in the text provides us with much better explanatory and predictive power.

The beauty of structural models comes with a cost of computational burden as manifested in estimating dynamic choice models. The problem of curse of dimensionality is exacerbated when I estimate the model at the individual level on a large sample, as I do in the third essay "Overhaul Overdraft Fees: Creating Pricing and Product Design Strategies with Big Data". In this project I build a model where I assume consumers perform financial planning and spend their money rationally subject to monitoring cost and heavy discounting tendency. As consumers exhibit significantly different behavior patterns, I need to estimate the model at the individual level in order to design targeted strategies. But to do this with the standard estimation methods takes prohibitively large amount of time. To solve this problem I borrow a new parallel computing algorithm to conduct parallel MCMC to speed up the estimation. This facilitates the marriage of applying the structural model on Big Data.