DISSERTATION PROPOSAL

Samuel Levy

"Essays on Bayesian Machine Learning in Marketing"

Thursday, December 2, 2021
4:30pm EST
Tepper 4219 & via Zoom

In this dissertation, I present three essays on Bayesian statistical and machine learning methods to infer latent variables either structural or reduced-form for informed marketing decision-making under uncertainty. I develop and apply these methods to different applied contexts across individual-level retail transaction data, clickstream data, and industrial customer transaction-based data.

Marketing managers need to allocate resources in a timely and effective manner under scarce data constraints when dealing with retention of recently acquired customers. As omnichannel interactions become more and more the norm, it is crucial to integrate online and offline consumer activity. In this first essay, I build a novel multimodal machine learning model for customer retention using transaction and clickstream data. The model generates topics of buying and browsing behavior, guided by a censored purchase timing at the customer level. Using a novel dataset of a leading supplier of industrial products and services, I find that clickstream data substantially improves the accuracy of purchase timing predictions. By fusing together different types of structured and unstructured data, I cluster customers into transaction and clickstream segments, which are used to predict customer activity. The model automatically discovers probabilistic relationships between the timing of transactions and clickstream activity, and predicts distributions of time-to-event at the individual level. I plan to apply this framework in the banking industry. Customers often need to interact with their bank for their specific needs and banks receive millions of calls every month. Understanding ahead of time the intent of interactions and their likely future occurrence would help anticipate customers' needs and propose a more customized approach in channel interactions. Marketers would be able to better anticipate when customers are likely to call, allocate the right human agents at the right time, enhancing customer experience.

Outside goods are usually included in marketing choice models to allow for flexible substitution patterns across products. However, restrictive structural assumptions are formulated through specific functional forms of outside good preferences. An imposed functional form of utility

strongly influences the range of estimable substitution patterns. Consequently, demand functions and elasticities may be systematically misspecified. In this second essay, I relax the functional form of utility of the outside good using Gaussian process priors and estimate a nonhomothetic choice model with simultaneous purchases under constrained utility maximization. Using transaction data from a large US retail chain, I capture nonlinear rates of satiation across consumers that traditional models fail to capture, by providing more precise and robust statistical inference. The nonlinearity in the preference for the outside good allows us to better understand the thresholds that certain customers are at, and predict demand more robustly. Modeling nonlinearities in the outside good would allow marketers to exploit threshold effects in consumption of the outside good relative to the inside good. In my future counterfactual analysis, I plan to design promotional offers that will account for sudden changes in satiation rate, as consumers expand their purchases in the category.

Managing industrial customer retention is challenging due to the lack of understanding of why certain customers are more at risk of churning than others, and how to prevent churn. Understanding controllable causes of churn is crucial but has been largely understudied. Pricing varies across buyers and may even vary between purchases of the same buyer, which might simultaneously impact the probability of churn and purchases. Unobserved random shocks, such as supply shortages and demand boosts, may impact pricing decisions, which is a potential source of endogeneity. Scaffolding on my second essay's direct utility framework, the third essay introduces a structural model of purchasing behavior and churn. Industrial consumers maximize their utility of choosing different product categories and/or attributes under unobserved budget constraint. Preference parameters of industrial consumers vary over time to reflect the dynamics of purchasing cycles, seasonality and option to churn. In my proposed work, I will estimate individual preferences nonparametrically with Gaussian process priors, and will control for price endogeneity with a Bayesian instrumental variable approach. Using a novel dataset from a large industrial supplier of products and services, I will investigate the effect of pricing on consumers' propensity to buy the inside goods against an unobserved outside good that characterizes customer inactivity and churn. I intend to run counterfactuals to determine why certain industrial consumers are more at risk of churning than others, and how dynamic pricing can help improve retention.