Marketing Modeling of Large-scale Sales Data

Nobuhiko Terui
Graduate School of Economics and Management
Tohoku University

Abstract

I present our research on statistical modeling of sales data for store management in marketing. We propose models to analyze two kinds of POS (point of sales) data in aggregated and disaggregated manners. They are both characterized as high dimensional sparse data.

First, we consider the aggregated sales and introduce models accommodating marketing concept when measuring market response and forecasting. Specifically, we first employ topic model for constructing market baskets well as reducing dimensionality of sample space. It is next connected to regression with hierarchical factor model to estimate the structure in reduced dimensional space and it is decompressed for discovering effective covariates and measuring their market responses in original sample space. We show unusual findings on the sets of product sale and its covariates obtained by scanning whole database beyond usual practice of category based analysis.

Second, we extend the topic analysis to disaggregated sales, called ID-POS data, which identify individual customers in purchase records. We explore factors that explain customer's purchasing contexts by the topic model with hierarchical structure whose topic is decomposed into demographic and marketing effects. Then store manager could recognize which and why customer prefers specific topics, for example, he is likely to purchase "lunch and dinner(topic)" for "himself(demographic)", and he prefers this topic more when "End promotion" is applied, and he is also likely to purchase "drink and snack(topic)" for "his family(demographic)" and he prefers this topic more when "Flyer promotion" is applied.