"Estimating Demand for Hotels by Mining User-Generated and Crowd-Sourced Content on the Internet"

Anindya Ghose
Leonard N. Stern School of Business
New York University

Abstract

User-Generated Content (UGC) is changing the way consumers shop for goods. It is increasingly being recognized that the textual content of product reviews is an important determinant of consumers' choices, over and above any numeric information. Similarly, websites that facilitate the creation of social tags by users can influence the desirability of a product or service. Moreover, one can harness the collective wisdom of the crowds by eliciting consumer opinions through on-demand user-contributed surveys. Based on a unique dataset of hotel reservations over a 3-month period from Travelocity.com, we estimate the demand for hotels using a structural model that incorporates information from different kinds of UGC. Data on UGC is obtained from three sources: (i) text of hotel reviews from two well-known travel search engines, Travelocity.com and Tripadvisor.com, (ii) social geo-tags identifying the different location-based attributes of hotels from Geonames.org, and (iii) on-demand user-contributed opinions on the most important hotel characteristics from Amazon Mechanical Turk. These data sources are merged with satellite images of the different hotel locations to create one comprehensive dataset summarizing the location and service characteristics of the hotels in our sample. We use text-mining techniques to incorporate textual information from user reviews in our estimation. We supplement these methods with image classification techniques and on-demand user-generated annotations. We estimate a two-step random coefficient structural model to infer the weights that consumers place on different location and service-related features of hotels. We also quantify how the extent of subjectivity, readability, complexity and other stylistic features of user-generated reviews affect hotel room sales. Based on the estimation of consumer surplus, we propose a new ranking system for displaying hotels in response to a search query on a travel search engine. By doing so, one can provide customers with the "best-value" hotels early on, thereby improving the quality of online hotel search compared to existing systems. Several field experiments with users suggest that our ranking system does better than existing systems.