Bayesian Customer Profiling: Applications to Age and Political Partisanship Estimation

Arnaud De Bruyn
ESSEC Business School

Thomas Otter
Goethe University

Abstract
Firms use external data sources (e.g., Census Bureau data, Google, Axciom) to infer the most likely characteristics of their customers and thus better predict their profiles and needs, unobtrusively. For instance, a firm may infer the unobserved income (S) of a customer list (L) based on their observed zip codes (X). We demonstrate that the simple count method most commonly used in this effort is based on the implicit assumption that X causes L. In the more likely situation where S causes L, and X merely serves as an indicator of S, the simple count method provides biased estimates. We develop a Bayesian profiling method that correct for these biases, and use simulations to show that in most managerially-relevant settings, the Bayesian method will outperform the simple count method, often by an order of magnitude. We then compare both methods in two case studies. The first example estimates customers’ age on the basis of their first names; prediction errors decrease sixfold compared with the simple count method. In the second example, the approach identifies 99.9% of people’s political affiliations based on their zip codes (cf. 35.6% with simple count method). The proposed approach offers more precise estimates that can enable marketers to profile their customers and tailor their marketing strategies to match customer characteristics.