

„Bayesian Customer Profiling: Applications to Age and Political Partisanship Estimation.“

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 determines list membership L . In the situation where unobserved S determines list membership the simple count method provides biased estimates. We develop a Bayesian profiling method that corrects for these biases, and use simulations to show that in 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 by 27.5% compared with the simple count method, and critical errors decrease from 6.6% to less than 0.1%. In the second example, the approach identifies 99.4% of people's political affiliations based on their ZIP codes (cf. 30.3% with simple count method).