On Aggregation Bias in Sponsored Search Data: Existence and Implications

VIBHANSHU ABHISHEK, The Wharton School, University of Pennsylvania
KARTIK HOSANAGAR, The Wharton School, University of Pennsylvania
PETER S. FADER, The Wharton School, University of Pennsylvania

There has been significant recent interest in studying consumer behavior in sponsored search advertising (SSA). Researchers have typically used daily data from search engines containing measures such as average bid, average ad position, total impressions, clicks and cost for each keyword in the advertiser's campaign. A variety of random utility models have been estimated using such data and the results have helped researchers explore the factors that drive consumer click and conversion propensities. However, virtually every analysis of this kind has ignored the intra-day variation in ad position. We show that estimating random utility models on aggregated (daily) data without accounting for this variation will lead to systematically biased estimates – specifically, the impact of ad position on click-through rate (CTR) is attenuated and the predicted CTR is higher than the actual CTR. First, we prove that the average daily position of an ad is less in convex order than the actual position of the ad for an impression. Using this result, we analytically demonstrate the existence of the aggregation bias. Second, using a large disaggregate dataset from a major search engine containing 8 million impressions, we empirically validate our findings for both the traditional logit model and the Hierarchical Bayesian models that are commonly used in the SSA literature. Third, we build a game-theoretic model to analyze the effect of the bias on the equilibrium of the SSA auction. We find that advertisers bid lower in SSA auctions as a result of the bias, which always leads to lower search-engine revenue. We also find that an advertiser can always increase his payoff when he unilaterally switches to complete data from aggregate data. Finally, we empirically quantify the losses experienced by the search engine and the advertisers and find that the search engine loses over 17% of its revenue on average. We also observe that an advertiser loses around 6% of his payoffs due to data aggregation. Our findings raise serious concerns for SSA practitioners and also question the adequacy of the data standards that have become common in SSA. Finally, we provide recommendations for aggregate datasets that do not suffer from the bias.

Categories and Subject Descriptors: K.6.0 [MANAGEMENT OF COMPUTING AND INFORMATION SYSTEMS]: Economics
General Terms: Economics, Measurement
Additional Key Words and Phrases: Sponsored search, aggregation bias, consumer behavior modeling, bayesian statistics


Author's addresses: V. Abhishek and K. Hosanagar, OPIM Department, The Wharton School, University of Pennsylvania; P.S. Fader, Marketing Department, The Wharton School, University of Pennsylvania.

Copyright is held by the author/owner(s).
EC’12, June 4–8, 2012, Valencia, Spain.
ACM 978-1-4503-1415-2/12/06.

1