Examining the Impact of Contextual Ambiguity on Search Advertising Keyword Performance: A Topic Model Approach

Vibhanshu Abhishek
Jing Gong
Beibei Li

Information Systems and Management
Heinz College, Carnegie Mellon University

Abstract
Sponsored search advertising offers a more targeted way of marketing than traditional advertising. However, the context of consumer search is often unobserved and the prediction of it can be nontrivial. Consumer search contexts may vary even when consumers are searching for the same keyword. Due to the ambiguity of a keyword, a large portion of the ads displayed may fall outside a particular consumer’s interest, potentially leading to low click-through rates on search engines. In our study, we propose an automatic way of examining keyword contextual ambiguity based on probabilistic topic models from machine learning and computational linguistics. We quantify the effect of contextual ambiguity on keyword click-through performance using a Hierarchical Bayesian Model that allows for topic-specific effect and nonlinear position effect. We validate our study using a novel dataset from a major search engine that contains information on consumer click activities for 12,790 distinct keywords across multiple product categories from over 4.6 million impressions from August 10, 2007 to September 16, 2007. We find that consumer click behaviors vary significant across keywords, and such variation is significantly affected by the contextual ambiguity of the keywords. More specifically, keyword contextual ambiguity can lead to a higher click-through rate (CTR) on top-positioned ads, but it is also associated with faster decay in CTR with position. Our study has the potential to help advertisers in designing keywords portfolio and bidding strategy by extracting contextual ambiguity and other semantic characteristics of keywords based on large-scale analytics from unstructured data. It can also help search engines improve the quality of displayed ads in response to a consumer search query.

1 Author names are in alphabetic order.
Overview

Sponsored search advertising has become a major effective marketing channel for businesses today. However, consumers arrive at search engines with diverse interests, and their search intent may vary even when they are searching using the same keyword. For advertisers, bidding on ambiguous keywords may increase the risk of ads unrelated to consumer interests. For search engines, it is also important to understand the contextual and semantic characteristics of keywords in order to maximize the total number of clicks, and thus total profit.

In this paper, we focus on how the ambiguity of a keyword’s context might affect keyword performance. More specifically, we propose an automatic way of examining keyword contextual ambiguity based on topic models from machine learning and computational linguistics, and quantify the effect of contextual ambiguity on keyword click-through performance using a Hierarchical Bayesian Model that allows for topic-specific effect and nonlinear position effect. We validate our study using a novel dataset from a major search engine that contains information on consumer click activities for 12,790 distinct keywords across multiple product categories from over 4.6 million impressions from August 10, 2007 to September 16, 2007. We find keyword performance is significantly affected by keyword category and contextual ambiguity. Specifically, contextual ambiguity has two opposing effects on the performance of ads associated with a keyword. Higher contextual ambiguity may lead to a higher baseline click-through rate (CTR); however, it is also associated with a faster decay in CTR with position. Therefore, the overall effect of contextual ambiguity on CTR for ads at various positions is a combination of these opposing effects. Our study has the potential to help advertisers in choosing keywords portfolio and bidding strategy. It can also help search engines improve the quality of displayed ads in response to a consumer search query.

Modeling Contextual Ambiguity

We model the contextual ambiguity of keywords using latent Dirichlet allocation model (LDA; Blei et al. 2003) from topic modeling in machine learning and natural language processing. We first construct a corpus of documents that store the top 50 Google organic search results of keywords. Based on the documents, we estimate the LDA model with a Gibbs sampler and obtain the topic probabilities.

We propose topic entropy to measure keyword ambiguity by quantifying how “noisy” a keyword is in terms of related topics. A keyword with higher entropy tends to relate to a broader range of topics (more ambiguous). Formally, let \( \hat{\theta}_{k,t} \) denote the posterior probability that keyword \( k \) belongs to topic \( t \). We measure topic entropy as follows:

\[
\text{TopicEntropy}_k = - \sum_{t=1}^{T} \hat{\theta}_{k,t} \log(\hat{\theta}_{k,t}).
\]

where \( T \) is the total number of topics.
Hierarchical Bayesian Model

To capture the impact of keyword characteristics on CTR and how CTR decreases with positions, we propose a Hierarchical Bayesian model that allows for topic-specific effect and flexible nonlinear specifications.

Following Abhishek and Hosanagar (2013), we model the CTR for keyword $k$ at ad position $p$ given that keyword $k$ belongs to topic $t$ as:

$$ CTR_{k,p,t} = P(\text{click}_{k,p} = 1 | \text{topic}_k = t) = \alpha_{k,t} \gamma_{k,t}^{p-1} $$  \hspace{1cm} (2)

where $\alpha_{k,t}$ captures the baseline CTR, and $\gamma_{k,t}$ captures the change of CTR with positions. We assume that $\alpha_{kt} = \frac{\exp(\tilde{\alpha}_{kt})}{1+\exp(\tilde{\alpha}_{kt})}$ and $\gamma_{kt} = \frac{\exp(\tilde{\gamma}_{kt})}{1+\exp(\tilde{\gamma}_{kt})}$.

The total number of clicks is assumed to follow a Binomial distribution:

$$ P(\text{clicks}_{k,p} | \text{impressions}_{k,p}, \text{topic} = t) $$

$$ = \left( \begin{array}{c} \text{impressions}_{k,p} \\ \text{clicks}_{k,p} \end{array} \right)^{CTR_{k,p,t}} \cdot (1 - CTR_{k,p,t})^{\text{impressions}_{k,p} - \text{clicks}_{k,p}} $$  \hspace{1cm} (3)

To incorporate keyword heterogeneity, we assume that $\tilde{\alpha}_{kt}$ and $\tilde{\gamma}_{kt}$ follow a normal distribution:

$$ \begin{pmatrix} \tilde{\alpha}_{kt} \\ \tilde{\gamma}_{kt} \end{pmatrix} \sim \text{MVN}(\mu_{kt}, \Phi), $$  \hspace{1cm} (1)

where $\mu_{kt} = (\mu^{(a)}_{kt}, \mu^{(y)}_{kt})'$. To capture the impact of keyword characteristics, we further assume that the mean effects $\mu^{(a)}_{kt}$ and $\mu^{(y)}_{kt}$ are as follows,

$$ \mu^{(a)}_{kt} = \beta^{(a)}_{0t} + X'_{k} \beta^{(a)}, $$

$$ \mu^{(y)}_{kt} = \beta^{(y)}_{0t} + X'_{k} \beta^{(y)}, $$

where $\beta^{(a)}_{0t}$ and $\beta^{(y)}_{0t}$ denote the topic specific intercept terms and $X'_{k}$ is a vector of characteristics for keyword $k$ including $\text{TopicEntropy}$, $\text{NoWord}$, $\text{Brand}$, $\text{Location}$, $\log(\text{Trans})$, $\text{AvgAdQuality}$, $\text{AvgAdFreq}$, $\text{AvgNumAd}$, and $\log(\text{Impressions})$. We assume that the intercept terms are drawn from a multivariate normal distribution as follows,

$$ \begin{pmatrix} \beta^{(a)}_{0t} \\ \beta^{(y)}_{0t} \end{pmatrix} \sim \text{MVN}(\mu_{0}, \Omega_{0}), $$  \hspace{1cm} (2)

We have multivariate normal priors on $\beta^{(a)}$, $\beta^{(y)}$, and inverse-Wishart priors on $\Phi$ and $\Omega_{0}$.

To capture the topic-level effect, we incorporate the topic distribution associated with each keyword estimated from LDA. We have the log-likelihood function as follows:
\[ LL = \sum_k \sum_p \sum_t \hat{\theta}_{kt} \log \left( \sum_t \hat{\theta}_{kt} P(\text{Clicks}_{k,p} | \text{Impressions}_{k,p}, \text{topic} = t) \right) \]

**Major Findings**

We estimated the model using 70% of the keywords chosen at random. We used Markov Chain Monte Carlo (MCMC) method for estimation and ran two MCMC chains, each with 70,000 iterations. The results are presented in Table 1. Firstly, the coefficient of topic entropy on \( \alpha \) is positive and statistically significant, indicating that keywords with higher topic entropy are associated with higher overall CTR. Second, topic entropy has a negative and significant impact on the decay parameter \( \gamma \), suggesting that keywords that have higher topic entropy witness a larger decrease in CTR with position. This indicates that on average although consumers are more likely to click ads associated with more ambiguous keywords on the top positions, they are less likely to click ads that are positioned lower on the screen once they start the search. Figure 1 demonstrates how CTR of keywords with different topic entropy changes with positions for sample keywords.

In addition, other keyword characteristics also significantly predict the baseline CTR. Specifically, brand-related keywords are associated with higher overall click propensity, while location-specific and popular keywords are associated with lower overall click propensity. Meanwhile, the impact of position tends to be smaller for longer and transaction-related keywords. However, position tends to have a stronger effect for brand-related, location-related, and popular keywords.

**Conclusion and Current Status**

Our findings have potential for advertisers in designing keywords portfolio and bidding strategy by extracting contextual ambiguity and other semantic characteristics of keywords based on large-scale analytics from unstructured data. Moreover, our results suggest that the effectiveness of sponsored search advertising varies across categories. A firm may take into account this difference to better allocate their advertising budget across product categories. Our results can also help search engines improve the quality of displayed ads in response to a consumer search query, to maximize total clicks and to increase overall profits from sponsored search advertisements.

Currently we are finalizing the robustness tests and model comparisons. We expect to have our complete manuscript finished in a couple of weeks.

---

\(^2\) We plan use the other 30% of the keywords for out-of-sample prediction analysis.
Table 1 Estimation Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>Posterior SD</th>
<th>Posterior Mean</th>
<th>Posterior SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha_{k,t}$ (Baseline Effect)</td>
<td>$\gamma_{k,t}$ (Position Interaction Effect)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TopicEntropy</td>
<td>0.064***</td>
<td>(0.021)</td>
<td>-0.145***</td>
<td>(0.023)</td>
</tr>
<tr>
<td>NoWord</td>
<td>0.019</td>
<td>(0.014)</td>
<td>0.055***</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Brand</td>
<td>0.064**</td>
<td>(0.025)</td>
<td>-0.161***</td>
<td>(0.030)</td>
</tr>
<tr>
<td>Location</td>
<td>-0.041*</td>
<td>(0.025)</td>
<td>-0.193***</td>
<td>(0.030)</td>
</tr>
<tr>
<td>log(Trans)</td>
<td>-0.012</td>
<td>(0.010)</td>
<td>0.047***</td>
<td>(0.011)</td>
</tr>
<tr>
<td>AvgAdQuality</td>
<td>-0.019**</td>
<td>(0.009)</td>
<td>-0.059***</td>
<td>(0.011)</td>
</tr>
<tr>
<td>AvgAdFreq</td>
<td>0.372***</td>
<td>(0.009)</td>
<td>0.256***</td>
<td>(0.010)</td>
</tr>
<tr>
<td>AvgNumAd</td>
<td>28.722***</td>
<td>(0.412)</td>
<td>-7.170***</td>
<td>(0.540)</td>
</tr>
<tr>
<td>log(Impressions)</td>
<td>0.064***</td>
<td>(0.021)</td>
<td>-0.145***</td>
<td>(0.023)</td>
</tr>
</tbody>
</table>

Figure 1 CTR by Entropy for Sample Keywords
References

Available at http://www.contrib.andrew.cmu.edu/~jingg/WCBI_References.htm