Perils of Uncertainty? The Impact of Contextual Ambiguity on Search Advertising Keyword Performance

Completed Research Paper

Jing Gong
Carnegie Mellon University
5000 Forbes Avenue,
Pittsburgh, PA 15213
jingg@andrew.cmu.edu

Vibhanshu Abhishek
Carnegie Mellon University
5000 Forbes Avenue,
Pittsburgh, PA 15213
vibs@andrew.cmu.edu

Beibei Li
Heinz College
Carnegie Mellon University
5000 Forbes Avenue,
Pittsburgh, PA 15213
beibeili@andrew.cmu.edu

Abstract

In this paper, we explore how the contextual ambiguity of a search can affect a keyword’s performance. We propose an automatic way of categorizing keywords and examining keyword contextual ambiguity based on 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, and validate our study using a novel dataset from a major search engine containing information on click activities for 12,790 keywords across multiple categories from over 4.6 million impressions. We find that consumer click behaviors vary significantly across keywords, and keyword category and contextual ambiguity significantly affect such variation. Specifically, higher contextual ambiguity can lead to higher click-through rate (CTR) on top-positioned ads, but the CTR tends to decay faster with position. Our study has the potential to help advertisers design keyword portfolios, and help search engines improve the quality of sponsored ads.

Keywords: Sponsored search advertising, topic models, contextual ambiguity, machine learning, keyword selection

Introduction

With the growing pervasiveness of consumer search for relevant information and products via search engines, sponsored search advertising (also known as “paid search” or “search advertising”) has become an important marketing channel for businesses. In the first half of 2013, sponsored-search advertising generated revenues of $8.7 billion and accounted for 43% of online advertising (Interactive Advertising Bureau 2013). Most online advertising techniques offer a more effective way of targeting customers as compared to traditional advertising. However, search advertising considerably outperforms other forms of online advertising such as display or social media advertising on metrics such as return on investments (ROI), click-through-rate (CTR), and conversion rates. The effectiveness of search
advertising is attributed to the fact that search engines match the ads shown to consumers with their current search intent derived from the keyword she uses.

When a consumer issues a query on a search engine using a keyword, the search engine identifies a list of advertisers who are bidding on the keyword. It subsequently presents an appropriate list of ads based on factors such as bids placed by the advertisers and their historical performance. The ability to present consumers ads tailored to their search context (as indicated by the keywords) considerably increases the likelihood that they will click on these ads.

However, even though a search keyword provides an indication of a consumer's search context, consumers from varied contexts might use the same keyword for searching. For example, a consumer who searches for the keyword “Mars” may be interested in astronomy and the planet Mars, or may be interested in buying chocolates and candies from the confectionery company Mars, or may be looking for a local chain of grocery stores in metropolitan Baltimore, Maryland. Therefore, the search engine faces ambiguity in predicting the consumer's search context. In comparison, some keywords are specific and do not have a variety of meaning, for example, “antivirus.” Consumers who search for “antivirus” and advertisers who bid on “antivirus” are likely to refer to the same product. Because keywords can have either a narrow or broad context, they might have varying appeal to consumers, and their performance might depend on the ambiguity in their context.

A keyword’s contextual ambiguity refers the uncertainty in using the keyword to predict consumers' search contexts. If a keyword is related to a broader range of contexts, a particular consumer’s search context becomes more ambiguous or uncertain to search engines and advertisers.

In this paper, we want to understand the interplay between a keyword’s context and consumers’ search behavior. More specifically, we wish to ascertain how the breadth of a keyword’s context might affect consumer behavior and keyword performance. On the one hand, prior literature in search theory suggests that as the uncertainty in the quality of search results increases, users are more likely to search (e.g., Weitzman 1979), because higher uncertainty leads to higher variance and more diversity in the alternatives, and users believe they will be more likely to find an alternative with a good fit or high value during their search. On the other hand, consumer psychology theories suggest that as the alternatives become less relevant, users are more likely to abandon their search (e.g., Jacoby et al. 1974; Dhar and Simonson 2003; Kuksov and Villas-Boas 2010), because users tend to get overwhelmed and discouraged by the complexity of information, and therefore lose their interest or trust in the search results. In reality, keyword contextual ambiguity can result in both higher diversity in ad quality and higher probability of ad irrelevancy. Therefore, how keyword contextual ambiguity would affect consumer click behavior is unclear. To explore this question, in this paper, we use a rich dataset from a major search engine to perform a cross-category analysis and examine which of these two opposing effects dominates in the context of search advertising.

In this study, we propose an automatic way of categorizing keywords and 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 that consumer click behaviors vary significantly across keywords, and keyword category and the contextual ambiguity of the keywords significantly affect such variation. Specifically, higher contextual ambiguity can lead to a higher click-through rate (CTR) on top-positioned ads, but the CTR tends to decay faster with position. Therefore, the overall effect of contextual ambiguity on CTR varies across positions. Moreover, we also find significant interplay between keyword category and screen position. In particular, the distribution of CTR among different screen positions varies significantly across keyword categories. For example, the CTR of keywords in certain categories, such as “adult,” “home,” or “style,” is more evenly distributed across different positions compared to other categories, such as “baby products,” “finance,” and “travel.” These results suggest that position effects appear to be more significant for certain product categories than others.

One major advantage of our study is that we are able to examine a large variety of keywords across multiple product categories. Such cross-category analysis allows us to distinguish our study from the
previous research, which mostly focused on a single product category from a particular retailer (e.g., Ghose and Yang 2009; Agarwal et al. 2011; Rutz and Bucklin 2011; Yao and Mela 2011). This cross-category analysis helps us provide broad generalizations as well as focus on category specific effects. Because the effectiveness of sponsored search advertising may vary across categories, a firm may take these industry-specific insights into account while crafting its search advertising strategy. Another advantage of this dataset is the presence of all competing ads for an impression, which helps us build a richer model of consumer search and derive new insights.

Web-scale data have commonly been used in computer science research, but there is an increasing need for such analysis in marketing research and practice, as exemplified by Archak et al. (2011), Netzer et al. (2012), and Ghose et al. (2012). Due to the extensive nature of our data, we resort to novel machine learning techniques to extract textual characteristics such as keyword ambiguity, semantic category, and transactional intent. We propose an automatic method of categorizing keywords and examining keyword contextual ambiguity based on topic models from machine learning and computational linguistics. Specifically, we construct a new semantic characteristic of a keyword, topic entropy, which is derived from the results of a topic model and measures the uncertainty in predicting consumers’ buying interest.

This paper makes the following contributions. First, we demonstrate how machine learning tools such as topic models can be used to extract semantic characteristics of keywords based on large-scale text analytics. Second, we expand the search advertising literature by examining how keyword attributes affect keyword performance in multi-category search advertising. Most previous studies obtain data from a single advertiser and thus can only measure keyword performance from the perspective of one specific firm (e.g., Ghose and Yang 2009; Agarwal et al. 2011), and the results may lack generalizability. The availability of detailed consumer search data for keywords across multiple advertisers and categories allows us to measure click performance at position level and identify keywords that have the potential to generate clicks – even at lower ad positions. Third, we find that consumer click behaviors vary significantly across keywords, and such variation can be explained by keyword category and the contextual ambiguity of keywords. Our empirical analysis offers new insights into consumer search behavior in the context of multi-category sponsored search, which can help advertisers select and evaluate keywords more effectively. Interestingly, our analysis suggests that consumers follow a two-step process while clicking on search ads. The effect of different keyword characteristics, such as entropy, differs in the propensity to start clicking and traversing the list of search ads. As a result, we reconcile the two opposing theories on consumer search and pave the path for richer consumer search models that can incorporate our findings.

**Literature Review**

Our paper is closely related to three different streams of literature -- (i) search advertising, (ii) machine learning and text mining, and (iii) search theory.

First, our study is related to the sponsored search advertising literature. During the past decade, the increasing popularity of sponsored search advertising has motivated research from multiple areas. Theoretical analysis on sponsored search include Edelman et al. (2007), Varian (2007), Athey and Ellison (2011), Katona and Sarvary (2010), and so on. Most of these studies focus on auction design and bidding strategies of firms. Empirical research on search advertising is also growing rapidly (Ghose and Yang 2009; Yang and Ghose 2010; Agarwal et al. 2011; Animesh et al. 2011; Rutz et al. 2011). Most of these studies primarily use keyword-level aggregate data provided by advertisers in particular industries to study ad performances. Exceptions are Jerath et al. (2014), Jeziorski and Segal (2014), and Yao and Mela (2011), who use individual-level data provided by search engines. The keyword characteristics the previous studies examined include keyword length, brand and retailer name (e.g., Ghose and Yang 2009; Yang and Ghose 2010; Jerath et al. 2014), geographic location (Rutz et al. 2012), specificity (Agarwal et al. 2011; Jerath et al. 2014), and product categories (Animesh et al. 2010). However, in these papers, keyword characteristics are handpicked on a small scale. Rutz et al. (2011) is to our best knowledge the most relevant to our study. The authors use a top-down approach to identify the key area of business related to each keyword (“keyword cluster”) in the automobile industry. However, this process relies on human input (e.g., interviews, questionnaires, and/or other communications with the firm’s management) to define keyword clusters. In our study, the use of topic modeling, namely, Latent Dirichlet Allocation model (LDA; Blei et al. 2003), allows us to characterize the topical content and category of a
keyword automatically on a large scale using unstructured text data. Moreover, most of the aforementioned studies focus on only a small set of keywords using data from a particular advertiser. In our study, we use a dataset that contains consumer click-through information for a large number of keywords across multiple product categories from different types of advertisers. By analyzing the contextual information extracted from each keyword, we are able to examine keyword performance across multiple product categories under different search contexts.

Second, our study is also related to literature in machine learning and text mining in marketing. A stream of research has recently emerged in marketing and information systems that applies machine learning and text mining techniques in examining online markets (e.g., Archak et al. 2011; Netzer et al. 2012; Ghose et al. 2012). Ghose and Ipeirotis (2011) and Archak et al. (2011) use text mining to extract multiple aspects of online review texts to identify important text-based features and to study their impact on review helpfulness (Ghose and Ipeirotis 2011) and product sales (Archak et al. 2011). Netzer et al. (2012) combine text mining and semantic network analysis to understand the brand associative network and the implied market structure. Decker and Trusov (2010) use text mining to estimate the relative effects of product attributes and brand names on the overall evaluation of the products. Ghose et al. (2012) use text mining and image classification to analyze the economic effects of user-generated content and crowd-sourced content, and to design a novel utility-based ranking system for products that can lead to an increase in consumer surplus. In our study, we propose applying topic modeling (i.e., LDA; Blei et al. 2003; Blei 2012; Steyvers and Griffiths 2007) from machine learning and natural language processing that allows us to extract the topical content of each keyword. To the best of our knowledge, our work is the first study that introduces probabilistic topic modeling in marketing research.

Third, our research is related to research on consumer search behavior. Traditional economic theories model consumer search behavior in an expected utility framework, where consumers search sequentially and stop searching when the marginal cost of search outweighs the marginal gain from search (Weitzman 1979). Many recent empirical studies adopt the sequential search framework to analyze online consumer search behavior (e.g., Kim et al. 2010; Ghose et al. 2013). This literature concludes that as the uncertainty in the quality of the alternatives becomes higher, consumers are more likely to search. On the other hand, work by Iyengar and Lepper (2000), Dhar and Simonson (2003), and Kuksov and Villas-Boas (2010) suggests that as the number of alternatives increases, consumers are less likely to search. Too many choices may present a cognitive burden on the consumers, reducing the likelihood of a purchase. Our study provides a rich and large dataset to shed light on the manner in which consumers perform search activities in a search advertising environment and reconcile these opposing viewpoints.

**Data**

The dataset used in this study is provided by one of the largest search engines in the United States. It consists of a random sample of close to 8 million search impressions conducted in the United States between August 10, 2007 and September 16, 2007. For every impression, the dataset comprises the keyword a user searches and a list of search ads shown to the consumer. The maximum number of ads shown per impression is eight. Although we do not have individual identifiers, which may restrict our ability to track individuals over time, every ad in our dataset has a unique identifier. For each ad displayed, we observe whether it was clicked during an impression. Note that although we have a unique ad identifier and we can track an ad across impressions, we do not have any ad-specific information. Because our data are derived from the search engine, we do not have post-click information, unlike some previous papers (Ghose and Yang 2009; Agarwal et al. 2011), which use data provided by advertisers.

We apply the following steps to pre-process the data: (i) we focus on keywords that receive at least one click during the entire five-week period and (ii) remove keywords that are domain names. We remove the low-performing keywords for two specific reasons. First, these keywords are not as relevant, as ads related to them never get clicked in our sampling period. Second, applying our machine learning techniques to all the keywords would take a substantial amount of time. Hence, we restrict our analysis to a smaller but more relevant dataset. We choose to ignore keywords containing domain names, because users who use these keywords know exactly which website they wish to visit, and these keywords are unlikely to lead to additional traffic for the website.
The subset of the dataset we examine in our analysis includes 12,790 distinct keywords from more than 4.6 million impressions. More than 0.17 million unique ads are displayed when consumers search for these keywords, resulting in 5.19 ads per impression. Twelve percent of the impressions receive at least one click. Table 1 presents the distribution of the number of clicks. Overall, there are about 640,000 clicks, and the average number of clicks per impression is 0.14. This observation is in agreement with prior research by Jerath et al. (2014) that suggests very few searches lead to clicks on search ads because the user needs might be met by organic ads.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMP</td>
<td>362.919</td>
<td>3736.077</td>
<td>2</td>
<td>292,692</td>
</tr>
<tr>
<td>CLICKS</td>
<td>49.660</td>
<td>197.835</td>
<td>1</td>
<td>6,984</td>
</tr>
<tr>
<td>AVG_CTR</td>
<td>0.028</td>
<td>0.027</td>
<td>0</td>
<td>0.555</td>
</tr>
<tr>
<td>NUM_WORDS</td>
<td>1.741</td>
<td>0.683</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>AVG_NUM_ADS</td>
<td>6.599</td>
<td>1.215</td>
<td>1.139</td>
<td>8</td>
</tr>
<tr>
<td>AVG_AD_QUALITY</td>
<td>0.026</td>
<td>0.022</td>
<td>0.001</td>
<td>0.451</td>
</tr>
<tr>
<td>BRAND</td>
<td>0.172</td>
<td>0.378</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>LOCATION</td>
<td>0.130</td>
<td>0.336</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>LOG_TRANS</td>
<td>2.173</td>
<td>1.075</td>
<td>0</td>
<td>5.509</td>
</tr>
</tbody>
</table>

Table 1. Summary Statistics

Table 1 presents summary statistics for the data. Note that we use the impression-level data to generate the summary statistics at a keyword level. IMP denotes the total number of times consumers search for a particular keyword in the dataset, which is analogous to the popularity measure used in Jerath et al. (2014). CLICKS measures the total number of clicks a particular keyword receives. AVG_CTR measures the average CTR of ads across different impressions for a keyword. NUM_WORDS denotes the number of words in the keyword. Previous studies on sponsored search report inconsistent findings regarding the effect of keyword length on CTR. Ghose and Yang (2009) suggest a negative relationship between keyword length and CTR, whereas Rutz et al. (2012) find that as the keyword becomes longer, the CTR increases. In this study, we try to resolve this conflict. AVG_NUM_ADS measures the average number of competing advertisers during an impression, which denotes the competitive intensity for a keyword. An important factor that determines keyword performance is the quality of ads, which is captured in AVG_AD_QUALITY. Because we do not directly observe ad attributes, we proxy for the quality of ads associated with a particular keyword by the average CTR that these ads receive for other keywords. Search engines commonly use this measure to predict ad quality. BRAND and LOCATION denote the presence of brand and location information in the keyword string.

We use LOG_TRANS to measure a keyword’s transactional intent. Some keywords may contain explicit transactional words, such as “cheap hotels” and “cruise deals,” but most keywords don’t contain explicit transactional indicators in the keywords, such as “airline tickets” and “honda parts.” Google organic search results, on the other hand, provide a better picture in terms of consumer search intent. If the keyword has a transactional intent, the Google organic search results are likely to contain transactional indicators such as “buy,” “discount,” “promotion,” and “check out.” Therefore, we propose to infer transactional intent using the keyword’s corresponding Google organic results. First, we compose a list of transactional words based on Dai et al. (2006) and general knowledge. The terms are listed in Table 2. Then, for each keyword, we count the frequency of transactional words in the corresponding Google organic results. We use the natural log of the frequency of transactional words, to measure keyword’s transactional intent.
Modeling Contextual Ambiguity

The major challenge in examining the impact of a keyword’s contextual ambiguity on consumer click behavior is how to quantify such ambiguity. We model the contextual ambiguity of each keyword based on probabilistic topic models from machine learning and natural language processing (Blei et al. 2003; See Blei 2012 for general introduction to topic models, and Steyvers and Griffiths 2007 for a more technical overview of topic models).

Topic models are unsupervised algorithms from machine learning and natural language processing that aim to extract hidden topics from unstructured text data. The intuition behind topic models is that a topic is a cluster of words that frequently occur together, and that documents, consisting of words, may belong to multiple topics with different probabilities. A probabilistic topic model tries to discover the underlying topic structure in a statistical framework. Topic models have been applied to many research areas, including the analysis of social network textual data (e.g., Paul and Dredze 2011), opinion mining, and information retrieval (e.g., Titov and McDonald 2008; Aral et al. 2011).

In this study, we model the contextual ambiguity of each keyword based on latent Dirichlet allocation model (LDA; Blei et al. 2003), which is a hierarchical Bayesian model that describes a generative process of document creation. The goal of LDA is to infer topics as latent variables from the observed distribution of words in each document. In particular, a topic is defined as a multinomial distribution over a vocabulary of words, a document is a collection of words drawn from one or more topics, and a corpus is the set of all documents.

To estimate the LDA model, we first construct a corpus of documents that describe the information content conveyed by the keywords. Keywords are usually words or short phrases. Obtaining adequate semantic information based on only keywords is usually difficult. To solve this challenge, we use Google organic search results to augment the keyword data and better understand the semantic meaning of each keyword (Dai et al. 2006; Abhishek and Hosanagar 2007). Obtaining the true contextual meaning(s) of a keyword can be difficult, but Google organic search results generated based on the classical theory of document relevancy provide a reasonable approximation.

Therefore, for all 12,790 keywords in our dataset, we extract the title and textual content of the brief description from each of the top-50-ranked Google organic search results, to construct the corresponding keyword-specific document. The results produce a total of 12,790 documents, each containing the most relevant information describing the corresponding keyword. After constructing the corpus of keyword documents, we pre-process the data following a standard procedure (e.g., Aral et al. 2011). We first remove annotations and tokenize the sentence into distinct terms. Then we remove stopping words using a standard dictionary.

Based on the documents, we estimate the LDA model with a Gibbs sampler and obtain the topic probabilities. In our study, we estimate the LDA model with a different number of topics, \( T = 20, 50, \) and 100. The most frequent words identified for the 20-topic model are presented in Figure 1, where topics are

| advertise | check out | get | promotion | sell |
| auction   | clearance | gift | product   | service |
| bidding   | consumer  | lease | purchase | ship |
| bill      | cost      | market | rebate    | shop |
| book      | coupon    | offer | rent      | store |
| buy       | customer  | pay   | reserve   | ticket |
| brand     | deal      | payment | retail   | order |
| cart      | delivery  | price | sale     |     |
| cheap     | discount  | promo | saving   |     |

Table 2. Transactional Words

E-Business
For convenience, we assign a label to each topic (e.g., “Sport,” “Music,” “Food”) based on its high-frequency words. For example, documents related to “style” often contain words such as “dress,” “party,” “woman,” and “fashion,” and so on.

We propose using topic entropy to measure keyword ambiguity. It captures the uncertainty of a keyword/document’s topic distribution (Hall et al. 2008). In information theory, entropy measures the unpredictability of a random variable. In our context, each keyword is associated with its own topic distribution inferred from the keyword-specific document. Therefore, we treat the topic assignment as a multinomial random variable, and use topic entropy to quantify how “noisy” a keyword is in terms of underlying topics. A keyword with higher entropy tends to relate to a broader range of topics (more ambiguous), whereas keywords with lower entropy tend to relate to fewer dominant topics (less ambiguous). Formally, let $\theta_k$ denote the posterior probability that keyword $k$ belongs to topic $t$. We measure topic entropy as follows:

$$\text{TOPIC ENTROPY}_k = -\sum_{t=1}^{T} \theta_k \log(\hat{\theta}_k).$$

**Figure 1. Frequent Words in Each Topic**

We propose using topic entropy to measure keyword ambiguity. It captures the uncertainty of a keyword/document’s topic distribution (Hall et al. 2008). In information theory, entropy measures the unpredictability of a random variable. In our context, each keyword is associated with its own topic distribution inferred from the keyword-specific document. Therefore, we treat the topic assignment as a multinomial random variable, and use topic entropy to quantify how “noisy” a keyword is in terms of underlying topics. A keyword with higher entropy tends to relate to a broader range of topics (more ambiguous), whereas keywords with lower entropy tend to relate to fewer dominant topics (less ambiguous). Formally, let $\theta_k$ denote the posterior probability that keyword $k$ belongs to topic $t$. We measure topic entropy as follows:

$$\text{TOPIC ENTROPY}_k = -\sum_{t=1}^{T} \theta_k \log(\hat{\theta}_k).$$
where $T$ is the total number of topics.

Figure 2. Topic Entropy: A Demonstration

Figure 2 illustrates the posterior topic probabilities and computed entropy for two keywords, “free anti virus” and “express.” For the keyword “free anti virus,” the estimated probability that it is related to topic “computer” is extremely high (0.93), and low for other topics, which means “free anti virus” is highly likely to relate to a single dominant topic -- “Computer.” As a result, the computed topic entropy for “free anti virus” is relatively small (0.44). By contrast, the keyword “express” has a fairly even topic distribution, resulting in relatively high topic entropy (2.64). Consequently, predicting what consumers are looking for when they search for the keyword “express” is difficult. We present the summary statistics for the estimated topic entropy in Table 3. As can be seen in Table 3, the maximum entropy value depends on the number of topics chosen. Simple calculation also shows that with $T$ topics, entropy ranges from 0 to $\log(T)$. The high correlations among entropy values derived based on a different number of topics also suggest entropy seems to be fairly robust to the number of topics specified in the LDA model.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 Topics</td>
<td>362.919</td>
<td>3736.077</td>
<td>2</td>
<td>292,692</td>
</tr>
<tr>
<td>50 Topics</td>
<td>49.660</td>
<td>197.835</td>
<td>1</td>
<td>6,984</td>
</tr>
<tr>
<td>100 Topics</td>
<td>0.028</td>
<td>0.027</td>
<td>0</td>
<td>0.555</td>
</tr>
</tbody>
</table>

Table 3. Summary Statistics for Topic Entropy
A Hierarchical Bayesian Model of Keyword Performance

To capture the impact of keyword characteristics on CTR and how CTR decreases with positions, we propose a hierarchical Bayesian model. Hierarchical Bayesian models have been widely used in marketing (Rossi and Allenby 2003) and sponsored search literature (e.g. Ghose and Yang 2009; Yang and Ghose 2010; Agarwal et al. 2011). One key advantage of a Bayesian approach is its convenience in estimating complex models. Feng et al. (2007) analyze a large dataset from a major search engine and suggest the exponential decay function provides the best fit for the variation in CTR with position. Following Feng et al. (2007) and Abhishek and Hosanagar (2013), we assume the CTR for a keyword follows an exponential decay function. More specifically, the CTR for an ad at position p for keyword k that belongs to topic t is as follows:

\[ CTR_{kpt} = P(click_{k} = 1|topic_{k} = t) = \alpha_{kt}\gamma_{kt}^{p-1}, \]  

where \( \alpha_{kt} \) captures the baseline CTR, and \( \gamma_{kt} \) captures the change of CTR with positions. Both \( \alpha_{kt} \) and \( \gamma_{kt} \) are constrained to be between 0 and 1, and differ across topics as consumers who are searching for a specific topic might demonstrate different kinds of behavior due to the nature of the product category or the information provided by organic search results. 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.

\[ P(CLICK_{kpt} | IMP_{kpt}, topic_{k} = t) = \binom{IMP_{kpt}}{CLICK_{kpt}} CTR_{kpt}^{CLICK_{kpt}} (1-CTR_{kpt})^{IMP_{kpt}-CLICK_{kpt}}, \]

where IMP denotes the total number of times consumers search for a particular keyword, and CLICKS measures the total number of clicks a particular keyword receives.

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

\[ \begin{pmatrix} \tilde{\alpha}_{kt} \\ \tilde{\gamma}_{kt} \end{pmatrix} \sim MVN(\mu_{\alpha}, \Sigma), \]

where \( \mu_{\alpha} = (\mu_{\alpha}^{(\alpha)}, \mu_{\alpha}^{(\gamma)})' \).

To capture the impact of keyword characteristics, we further assume that the mean effects \( \mu_{\alpha}^{(\alpha)} \) and \( \mu_{\alpha}^{(\gamma)} \) are as follows,

\[ \begin{align*}
\mu_{\alpha}^{(\alpha)} &= \beta_{0(\alpha)}^{(\alpha)} + X_{k}^{(\alpha)} \beta^{(\alpha)}, \\
\mu_{\alpha}^{(\gamma)} &= \beta_{0(\gamma)}^{(\gamma)} + X_{k}^{(\gamma)} \beta^{(\gamma)},
\end{align*} \]

where \( \beta_{0(\alpha)}^{(\alpha)} \) and \( \beta_{0(\gamma)}^{(\gamma)} \) denote the topic specific intercept terms and \( X_{k} \) is a vector of characteristics for keyword \( k \) including TOPIC_ENTROPY, NUM_WORDS, BRAND, LOCATION, LOG_TRANS, AVG_AD_QUALITY, AVG_NUM_AD, and LOG_IMP.

We assume that the intercept terms are drawn from a multivariate normal distribution as follows,
We have multivariate normal priors on $\beta^{(a)}_t$, $\beta^{(r)}_t$, and $\beta_0$, and inverse-Wishart priors on $\Phi$ and $\Omega_0$.

$$\beta^{(a)}_t \sim MVN(0, \Sigma^{(a)}_t), \quad \beta^{(r)}_t \sim MVN(0, \Sigma^{(r)}_t),$$

$$\mu_0 \sim MVN(0, \Sigma_0),$$

$$\Phi \sim InverseWishart(V^\Phi, \nu^\Phi),$$

$$\Omega_0 \sim InverseWishart(V^\Omega, \nu^\Omega),$$

where $\Sigma^{(a)}_t = \Sigma^{(r)}_t = 1000I_k$, $\Sigma_0 = V^\Phi = V^\Omega = 1000I_2$, and $\nu^\Phi = \nu^\Omega = 2$.

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:

$$LL = \sum_k \sum_p \log(\sum_t \hat{\alpha}_{tp} P(CLICKS_{kp} | IMP_{kp}, topic_k = t)).$$

### Major Findings

We estimated the model using 70% of the keywords chosen at random. We mean-centered all keyword characteristics so that the intercepts $\beta^{(a)}_0$ and $\beta^{(r)}_0$ can be interpreted as estimates of $\bar{\alpha}_{ki}$ and $\bar{\gamma}_{ki}$ for a typical keyword of which the covariates are set to mean values. We used Markov Chain Monte Carlo (MCMC) method for estimation. We use MCMC because MCMC methods work particularly well with models with multiple levels and mixtures. We also run robustness checks by estimating linear and logit regressions using CTR as the dependent variable, and the results are qualitatively similar. We compare the prediction performance of our focal model against the alternative models, and our model outperforms other models. We ran two MCMC chains, each with 70,000 iterations. The results are presented in Table 4.

<table>
<thead>
<tr>
<th></th>
<th>Posterior Mean</th>
<th>Posterior SD</th>
<th>Posterior Mean</th>
<th>Posterior SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{ki}$ (Baseline Effect)</td>
<td>0.064***</td>
<td>(0.021)</td>
<td>-0.145***</td>
<td>(0.023)</td>
</tr>
<tr>
<td>$\gamma_{ki}$ (Position Interaction Effect)</td>
<td>-0.041*</td>
<td>(0.025)</td>
<td>-0.193***</td>
<td>(0.030)</td>
</tr>
<tr>
<td><strong>TOPIC_ENTROPY</strong></td>
<td>0.019</td>
<td>(0.014)</td>
<td>0.055***</td>
<td>(0.016)</td>
</tr>
<tr>
<td><strong>BRAND</strong></td>
<td>0.064**</td>
<td>(0.025)</td>
<td>-0.161***</td>
<td>(0.030)</td>
</tr>
<tr>
<td><strong>LOCATION</strong></td>
<td>-0.101</td>
<td>(0.025)</td>
<td>-0.193***</td>
<td>(0.030)</td>
</tr>
<tr>
<td><strong>LOG_TRANS</strong></td>
<td>-0.012</td>
<td>(0.010)</td>
<td>0.047***</td>
<td>(0.011)</td>
</tr>
<tr>
<td><strong>LOG_IMP</strong></td>
<td>-0.019**</td>
<td>(0.009)</td>
<td>-0.059***</td>
<td>(0.011)</td>
</tr>
<tr>
<td><strong>AVG_NUM_AD</strong></td>
<td>0.372***</td>
<td>(0.009)</td>
<td>0.256***</td>
<td>(0.010)</td>
</tr>
<tr>
<td><strong>AVG_AD_QUALITY</strong></td>
<td>28.722***</td>
<td>(0.412)</td>
<td>-7.170***</td>
<td>(0.540)</td>
</tr>
</tbody>
</table>

***, **, and * indicate a 99%, 95%, and 90% significance level.
One potential explanation is that a consumer who searches with a higher entropy keyword may face higher uncertainty about alternatives of a consideration set, and may be more likely to explore the ads associated with the ambiguous keyword to formulate a better understanding of alternatives. In this regard, our finding also concurs with extant literature on consumer search (e.g., Weitzman 1979; Kim et al. 2010). Classical consumer search theory suggests that as the uncertainty in the search results increases, users are more likely to search (e.g., Weitzman 1979), because higher uncertainty reflects higher variance in the distribution of the utility of the alternatives, and users believe they will be more likely to find an alternative with an extreme value in the utility distribution (e.g., good fit or high value) during their search. Therefore, the expected marginal benefit of search is likely to be high when the choice uncertainty is high, which explains the higher probability of consumer search on average. Another explanation for the positive effect of ambiguity on $\alpha$ can be provided by the organic results shown for the search. The search for ambiguous keywords leads to a varied set of results by construction, which might force users to click on ads to fulfill their search intent.

Second, we observe that topic entropy has a negative and significant impact on the decay parameter $\gamma$. That is, keywords that have higher topic entropy seem to have a lower $\gamma$ and witness a larger decrease in CTR with position. This finding indicates that although on average, consumers are more likely to click ads associated with more ambiguous keywords, especially from the top positions, they are less likely to click ads that are positioned lower on the screen once they start the search. In other words, during the sequential search process, consumers are more likely to lose their search interests and stop exploring the choice set further under more ambiguous keywords. Reduced consumer search activity associated with ambiguous ads may be attributed to search costs theory (e.g., Stigler 1961) and information overload theory (e.g., Iyengar and Lepper 2000, Dhar and Simonson 2003, and Kuksov and Villas-Boas 2010). The well-established search costs literature has demonstrated that higher search costs can lead to lower search intensity (e.g., Stigler 1961; Weitzman 1979). Meanwhile, previous research has shown both theoretical and empirical evidence that information overload can discourage consumers and lead to consumer search termination due to loss of interests in the search results (e.g., Iyengar and Lepper 2000, Kuksov and Villas-Boas 2010, and Ghose et al. 2013). In the context of sponsored search, once consumers start clicking ads from top to bottom, they start to learn the diverse quality of the ads in sequence. Moreover, they also learn their search costs of searching each of these ads. Due to the high heterogeneity in the quality of the ads generated by the ambiguous keywords, consumers might find it cognitively costly to process these ads to pick the relevant ones (hence high search costs). In addition, a highly diverse set of ads might introduce an increasing amount of new information to consumers during the search process. Such information may lead to an overload for consumers due to their cognitive limitation and non-negligible search costs. Therefore, when keywords are more ambiguous, we notice a significant reduction in the probability of clicking on the ads as consumers proceed down the list. This finding is also consistent with the effects of knowledge uncertainty found by Urbany et al. (1989). One potential explanation is that keywords with high topic entropy, and hence high ambiguity, may occur in search contexts where consumers’ knowledge uncertainty is high. The lack of knowledge can in turn increase the barrier for consumers to evaluate choices, leading to much higher search costs for consumers during the search process. This interpretation is in line with the search costs theory where higher search costs can lead to lower search intensity.

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.

The analysis presented here suggests that consumers may follow a two-step process while engaging in sponsored search. Their behavior while deciding whether to click on search ads differs considerably from their behavior once they have started search, as evidenced by the differing effect of keyword characteristics on $\alpha$ and $\gamma$. For example, keyword entropy has two opposing effects on the performance of ads associated with a keyword. Our finding suggests that keyword contextual ambiguity can lead to a higher baseline CTR; however, it also implies the CTR decreases more sharply with position. Urbany et al. (1989) provide a potential explanation of the dual effect of entropy on keyword performance. Urbany et al. (1989) divides the uncertainty consumers face into knowledge uncertainty and choice uncertainty. Choice
uncertainty captures a consumer’s uncertainty in comparing alternatives and leads to more search, whereas knowledge uncertainty reflects customer uncertainty about the dimensions of comparison and leads to less search. In our case, higher entropy may be related to higher choice uncertainty overall, hence increasing overall consumer search. However, once consumers start a search, higher entropy also may also to related to higher knowledge uncertainty from each position as the consumers become confused by incoherent ads, resulting in deceased search activity and thus a faster decay in CTR with position. The overall effect of keyword entropy on CTR for ads at various positions is a combination of these opposing effects. In our analysis, we observe that the decrease in the decay with position always dominates the increase in the baseline CTR. For example, the ads at positions 2–8 for a high entropy keyword tend to get a smaller number of clicks as compared to ads for a keyword with lower entropy. This finding indicates that firms should either advertise on less ambiguous keywords or bid aggressively on the top position for ambiguous keywords if they are interested in increasing traffic to their websites.

Figure 3 illustrates how CTR changes with position by topic. Darker red colors represent higher CTR values and light yellow colors represent lower CTR values. Our results suggest the position effect on CTR is heterogeneous across different topics. For example, baby-related keywords tend to attract higher CTR than listing-related keywords at the top position, but the CTR decreases quickly at lower positions. This pattern we observe for baby-related keywords might be driven by the fact that consumers are extremely interested in baby-related ads, which leads to a high CTR at the first position, but their search intent is satisfied after the first few clicks. On the other hand, searches related to “adult,” “home,” and “style” start out with fewer clicks on ads in the top position but continue to receive clicks on ads in lower positions. In fact, we observe that the search depth for “adult” is the largest among all the categories considered in our analysis.

The large-scale, cross-category analysis presented here enables us to generate insights that depart from well-established results in the sponsored search literature. First, we can leverage data from multiple advertisers to generate insights that cannot be derived using data from a single advertiser. For example, we observe that ads associated with branded keywords tend to have smaller CTR at lower positions than generic keywords. At first, this finding might seem contrary to the results shown by Rutz and Bucklin (2011), who use data from a single advertiser to show that the presence of a brand name always increases CTR. To illustrate why data from a single advertiser might lead to such a result, consider the following example. An advertiser (Hilton) bids on two keywords, “hotel NYC” and “Hilton NYC”, and shows an identical ad for both the keywords. A consumer who uses the keyword “Hilton NYC” is looking for a Hilton property and is more likely to click on an ad by Hilton. On the other hand, a consumer searching for “hotel NYC” is looking for hotels in New York City and might not have a strong preference for Hilton,
reducing the likelihood of clicking on an ad by Hilton. Ceterius paribus, the advertiser will see more clicks for the keyword “Hilton NYC” than “hotel NYC,” and analyzing data from only one advertiser might lead to the conclusion that branded keywords always lead to more clicks. Our analysis, on the other hand, is based on a large sample of keywords across different advertisers, and we believe our findings complement the earlier findings. We conclude that, in general, keywords that contain brand names perform better than keywords that do not; however, a sharp decrease occurs in the CTR (with position) for branded keywords. This insight is manifested in the positive effect of BRAND on $\alpha$ but a negative effect on $\gamma$, resulting in an overall lower CTR for branded keywords. Second, data from all the advertisers for a keyword help us estimate the effect of keyword characteristics on the baseline CTR and the decay parameter separately. Such a model allows us to measure the differential effect of factors such as LOG_TRANS and the topic membership on $\alpha$ and $\gamma$, and gain a better insight of customer search behavior in the context of sponsored search.

**Conclusion and Discussion**

In this paper, we study the effect of semantic characteristics on keyword click-through rate and try to provide insights into consumer behavior in the context of search advertising. We analyze data from a major US search engine for 12,790 distinct keywords spanning 4.6 million impressions. Our paper is unique in several ways. First, to our knowledge, this dataset is the most extensive one used in the sponsored search literature that includes individual-level data across multiple product categories and advertisers. Most of the prior research in this area has used data from a single advertiser or a few keywords from search engines. Second, we are able to exploit machine learning techniques to generate semantic characteristics of keywords given the large volume of keywords available to us. The large number of keywords helps us create meaningful categorization without any manual intervention. Using text mining techniques, we are able to generate keyword characteristics such as presence of brand name and location, length of keywords, and whether a keyword is transactional, and study the impacts of these factors on consumer click behavior. Third, we introduce a new keyword characteristic -- topic entropy -- that measures the ambiguity in the semantic meaning of a keyword. We find that topic entropy has a significant impact on the CTR. Our results suggest topic entropy and keyword categories are significant predictors of keyword performance and potentially affect consumer click behavior.

Our results indicate that when consumers search using ambiguous keywords, they are more likely to click on sponsored ads in top positions. However, consumers who click on ads associated with these ads tend to search less aggressively and click fewer ads in lower positions. More specifically, CTR decays faster with position for ambiguous keywords as compared to precise keywords. We also observe that consumer behavior varies significantly across categories. Categories such as TV and sports receive fewer clicks, whereas categories such as style and travel receive significantly more clicks. We extend the prior understanding of brand- and location-specific keywords. Our results suggest consumers are “in general” less likely to click on ads when searching for location related information, and are less likely to conduct intensive search when searching for brand information. These counterintuitive results (as compared to previous literature) that shed further light on consumer behavior are derived because we are able to capitalize on data across different advertisers, whereas most prior studies use data from a single advertiser. We also observe that as the search volume of the keywords increases, the number of clicks decreases. In addition, as the number of ads for an impression increases, consumers are likely to click on more ads.

Our research has several managerial implications for advertisers and search engines. First, advertisers who are interested in driving traffic through clicks should be cautious about adding ambiguous keywords to their portfolios. Furthermore, if they need to choose some ambiguous keywords, they might be better off bidding aggressively and trying to attain higher positions. The search engines should also be cognizant of the topic ambiguity while designing their sponsored search strategies. For example, search engines may incorporate keyword characteristics while evaluating ad quality. Second, given that consumers behave differently across categories, advertisers can design their advertising strategies by allocating resources appropriately across products. Third, advertisers can use insights about how consumers respond to different keyword characteristics in designing better campaigns.
This paper incorporates new machine learning techniques in the literature that are instrumental in deriving new insights. We hope this research can pave the way for interesting future research that addresses the limitation of our current work. One major limitation of our analysis is the lack of ad information. For example, we don’t know ad characteristics such as ad copy, bid, and landing page. Although we try to compute some ad measures from our data and use these measures as proxies to control for ad-level differences in our analysis, providing richer insights with ad-level data is possible. Another limitation of this paper is the lack of post-click or conversion activity in the data. Therefore, our measurement of keyword performance is limited only to CTR. From the perspective of a search engine, CTR is a more relevant metric, because advertisers only pay when a click-through occurs. However, advertisers are not only interested in CTR, but they are also interested in conversion rates as well as other non-transactional benefits such as increased awareness. A richer dataset can be used to address this issue. We believe that sponsored search advertising continues to evolve as an interesting area of marketing research and this paper can contribute both methodologically and theoretically to this growing literature.

References


