MIS Quarterly

RESEARCH ARTICLE

EXAMINING THE IMPACT OF KEYWORD AMBIGUITY ON SEARCH ADVERTISING PERFORMANCE: A TOPIC MODEL APPROACH

Jing Gong
Department of Management Information Systems, Fox School of Business, Temple University,
Philadelphia, PA 19103 U.S.A. {gong@temple.edu}

Vibhanshu Abhishek
Paul Merage School of Business, University of California – Irvine,
Irvine, CA 92697 U.S.A. {vibs@uci.edu}

Beibei Li
Heinz College, Carnegie Mellon University,
Pittsburgh, PA 15213 U.S.A. {beibeili@andrew.cmu.edu}

Appendix A

Summary of Empirical Studies on Sponsored Search Advertising

Paper	Goal	Data Source	Industry	Level of Detail	Number of Keywords Examined
Agarwal et al. (2011)	Impact of position on click-through and conversion	Advertiser	Pet products	Aggregate	68
Agarwal et al. (2015)	Impact of organic competition on click-through and conversion	Advertiser	Pet products	Aggregate	36
Chan et al. (2011)	Measuring the value of customers acquired from sponsored search	Advertiser	Lab supplies	Individual	90-208
Chan and Park (2015)	Advertiser valuation of consumer search activities	Search engine	Sporting goods	Individual	1
Ghose and Yang (2009)	Impact of keyword attributes on click-through and conversion	Advertiser	Retail	Aggregate	1,878
Goldfarb and Tucker (2011)	Online and offline advertising channel substitution	Advertiser	Legal service	Aggregate	139
Jerath et al. (2014)	Impact of keyword popularity on click performance	Search engine		Individual	1,200
Jeziorski and Segal (2015)	Quantifying rational user experience and externalities among ads	Search engine		Individual	4
Rutz et al.	Impact of ad position on	Advertiser	Hotel	Aggregate	301

Paper	Goal	Data Source	Industry	Level of Detail	Number of Keywords Examined
(2012)	conversion performance				
Rutz and Bucklin (2011)	Spillover from generic to branded keywords	Advertiser	Hotel	Aggregate	Several hundred
Rutz and Trusov (2011)	Effects of ad attributes on ad performance	Advertiser	Ringtone	Aggregate	80
Rutz et al. (2011)	Quantifying indirect effects of paid search	Advertiser	Automotive	Aggregate	3,186
Yang and Ghose (2010)	Relationship between organic and sponsored search	Advertiser	Retail	Aggregate	426
Yang et al. (2014)	Impact of ad competition on click performance and cost per click	Advertiser	Digital camera and video products	Aggregate	1,573
Yao and Mela (2011)	Modeling user, advertiser, and search engine interaction	Search engine	Music management	Individual	

Appendix B

Latent Dirichlet Allocation

The most widely used topic model is the 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. Based on our discussion on corpus construction, we construct a document for each keyword that best reflects the contextual information of the keyword. We now discuss how we use LDA to infer the topics from the corpus of documents.

Formally, let T be the number of topics related to the corpus, let D be the number of documents in the corpus, and let W be the total number of words in the corpus. We assume that each document in the corpus is generated according to the following process:

Step 1. For each topic t, choose $\varphi_t = (\varphi_{t1}, \ldots, \varphi_{tW}) \sim Dirichlet(\psi)$, where φ_t describes the word distribution of topic t over the vocabulary of words.

Step 2. For each document d, choose $\theta_d = (\theta_{d1}, \ldots, \theta_{dT}) \sim Dirichlet(\omega)$, where θ_{dt} is the probability of topic t to which document d belongs.

Step 3. For each word n in document d, (1) choose a topic $t_{dn} \sim Multinomial(\theta_d)$, and (2) choose a word $w_{dn} \sim Multinomial(\varphi tdn)$.

ψ and ω are hyper-parameters for the two prior distributions, $Dirichlet(\psi)$ as the prior distribution of φ (word distribution in a topic) and $Dirichlet(\omega)$ as the prior distribution of θ (topic distribution in a document). We use the values suggested by Steyvers and Griffiths (2007) ($\psi = 0.01$ and $\omega = 50/T$).

Based on the generative process described above, we use a Markov chain Monte Carlo (MCMC) algorithm to estimate φ and θ. Specifically, we use a collapsed Gibbs sampler to sequentially sample the topic of each word token in the corpus conditional on the current topic assignments of all other word tokens (for details, see Griffiths and Steyvers 2004). We run a collapsed Gibbs sampler using *MALLET* (McCallum 2002) with 2,000 iterations.

Appendix C

Topic Distribution of Sample Keywords

Figure C1 illustrates the topic distribution of some sample keywords. In Figure C1, topics are labeled on the horizontal axis, and keywords are labeled on the vertical axis. The size of each bubble indicates a posterior topic probability, with larger bubbles representing higher probabilities. For example, the top-left bubble represents the posterior probability that the keyword "judges gavels" belongs to the topic "music," which is much smaller than the posterior probability that "judges gavels" belongs to "government," represented by the eighth bubble on the first row. Meanwhile, the keyword "marriage records" has a much larger posterior probability of belonging to the topic "government," which suggests "marriage records" is most likely related to government affairs rather than other topics.

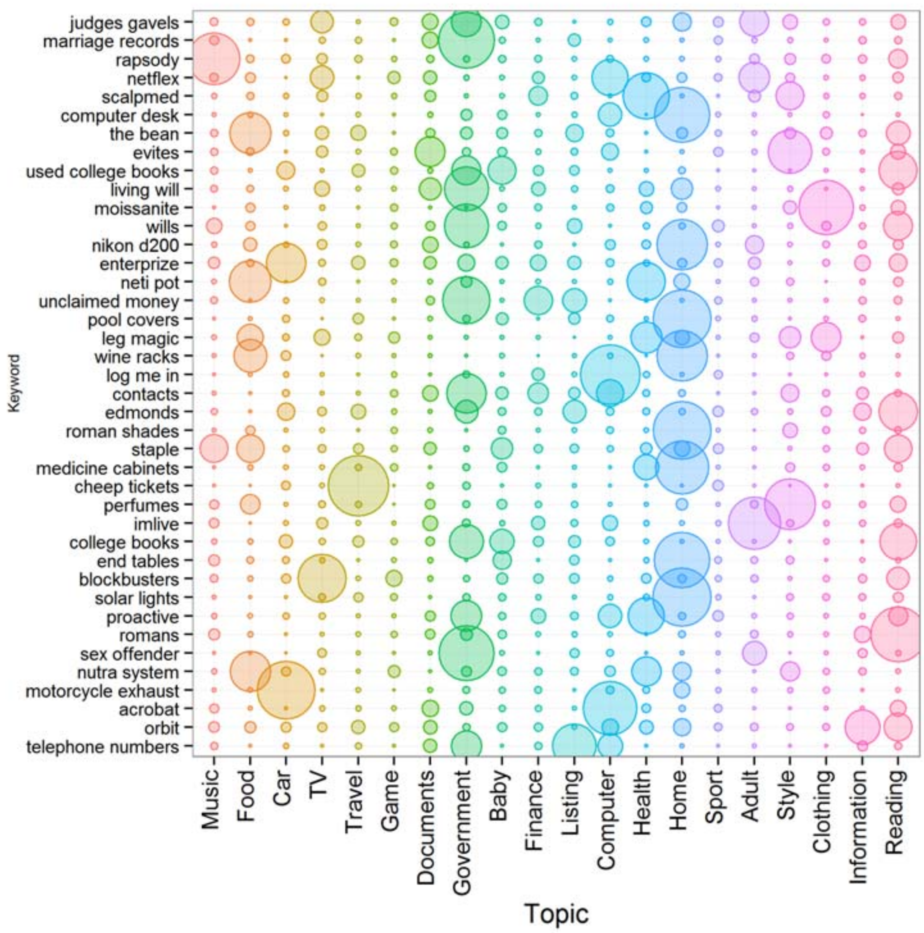

Figure C1. Topic Distribution of Sample Keywords

Appendix D

Extracting Brand and Location Information

We use a rule-based method to identify whether a keyword contains brand information. First, we obtain a list of brand names from namedevelopment.com, and use a fuzzy string matching algorithm to match each keyword against the list of brand names. In particular, we use *Levenshtein distance* (also called *edit distance*; Levenshtein 1966) to measure string similarity.[1] Using partial matching, we allow substrings of a keyword to match against brand names. For example, we want to match the keyword "ikea store" to the brand "ikea." For each keyword, we identify the brand name that gives the longest partial string match. We classify the keyword as containing brand information if one of the following conditions is met: (1) if the highest full-string similarity (i.e., Levenshtein distance computed from our model) is greater than 0.85; (2) if the highest partial-string similarity is greater than 0.85, and the brand name is a complete word in the keyword other than a substring of a word. We choose a Levenshtein distance of 0.85 as the cut-off point to allow for a moderate level of mis-spelling. For example, we match the keyword "chipo**lt**e" with the brand "chipo**tl**e," and "walmart" with "wal mart."[2]

We use a similar approach to extract whether a keyword contains location information. We obtain a list of U.S. city and state names, and match each keyword against the list of locations. For each keyword, we find the location name that gives the longest partial string match. We classify the keyword as containing location information if the highest partial string similarity is 1, which means an exact match is found, and the location name is a complete word in the keyword.[3]

Appendix E

Extracting Transactional Intent

In this study, we are interested in learning how likely consumers are to engage in a transaction when they search for a keyword. Therefore, we focus on detecting transactional intent from keywords. 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." The augmented 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. These transactional words are listed in Table E1. Then, for each search keyword, we count the frequency of transactional words in the corresponding Google organic results. We use *LOG_TRANS*, the natural log of the frequency of transactional words, to measure keyword's transactional intent.

Table E1. Transactional Words

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

[1]As a robustness check, we also use the "n-grams" method for string matching, where we define n = 2, 3, and 4. We find the final results remain consistent.

[2]To choose the optimal cut-off distance, we first manually identified whether a smaller number of keywords contain brand names. We then tried different cut-off values and chose the one (e.g., 0.85) that minimizes classification errors on the small keyword set.

[3]Similar to the process of identifying brand names, we chose the optimal cut-off distance based on a smaller set of keywords to minimize the classification error.

Appendix F

Correlation among Variables

Table F1. Correlation among Variables

Variable	*TOPIC_ENTROPY*	*NUM_WORDS*	*BRAND*	*LOCATION*	*LOG_TRANS*	*LOG_IMP*
TOPIC_ENTROPY	1.00					
NUM_WORDS	-0.38	1.00				
BRAND	-0.03	0.06	1.00			
LOCATION	0.00	0.16	0.04	1.00		
LOG_TRANS	0.07	-0.04	0.22	0.01	1.00	
LOG_IMP	-0.03	-0.17	0.19	-0.02	0.08	1.00

Appendix G

The Gibbs Sampling Procedure

We estimate our hierarchical Bayesian model using a Gibbs sampling procedure, which samples parameters iteratively from their conditional distributions given the data and all other parameters. Note that we model CTR conditional on the topic t_i related to impression i. As t_i is not observed, we use a data augmentation approach by simulating topic assignment based on membership probabilities $\hat{\theta}_{k_i} = \left(\hat{\theta}_{k_i,1}, \cdots, \hat{\theta}_{k_i,T}\right)'$, which is estimated from a topic model. In addition, U_{iat} is also a latent variable that involves data augmentation.

For simplification, we assume that

$$\beta^{kt} = \left(\beta_{0,k,t}, \beta_{1,k,t}\right)'$$

$$\phi^{k} = \left(\phi_{0,k}, \phi_{1,k}\right)'$$

$$u_k^{\beta} = \left(u_{0,k}, u_{1,k}\right)'$$

$$\Delta^{\beta} = \left(\Delta_0^{\beta}, \Delta_1^{\beta}\right)'$$

$$\chi^{k} = \Delta^{\beta} X_k + u_k^{\beta}$$

The hierarchical Bayesian model can be written in the following hierarchical form:

$$t_i | \hat{\theta}_{k_i}$$
$$U_{iat} | POS_{ia}; \beta^{k_i,t}, \beta_2, \tau_a, \eta_{ia}$$
$$POS_{ia} | \phi_{0,k_i}, \phi_{1,k_i}, \phi_2, \epsilon_{ia}$$
$$\eta_{ia}, \epsilon_{ia} | \Lambda$$
$$\tau_a | v^\tau$$
$$\beta^{k_i,t} | \gamma_t, \Delta^\beta, \Omega^\beta$$
$$\phi^{k_i} | \Delta^\phi, \Omega^\phi$$
$$\gamma_t | \Psi$$

We assume the following prior specifications:

$$\beta_2 \sim N(0, v^{\beta_2})$$
$$\phi_2 \sim N(0, v^{\phi_2})$$
$$vecr(\Delta^\beta) \sim MVN\left(\overline{\Delta^\beta}, A^\beta\right)$$
$$vecr(\Delta^\phi) \sim MVN\left(\overline{\Delta^\phi}, A^\phi\right)$$
$$v^\tau \sim IG(m, n)$$
$$\sigma_{12} \sim N(r_0, b_0)$$
$$\sigma^2 \sim IG(v_0, c_0)$$
$$\Psi \sim IW(v^\Psi, V^\Psi)$$
$$\Omega^\beta \sim IW(v^\beta, V^\beta)$$
$$\Omega^\phi \sim IW(v^\phi, V^\phi)$$

We describe the Gibbs sampling procedure below.

Step 1. Draw $t_i \sim Multinomial(\hat{\theta}_{k_i})$ for each impression i.

Step 2. Draw U_{iat} for each observation.

We can draw U_{iat} from the following posterior distribution:

$$U_{iat} \sim TN(\mu_{iat}, \sigma_{1|2})$$

where TN denotes the truncated normal distribution, and U_{iat} is truncated above zero if $Click_{ia} = 1$, and below zero if $Click_{ia} = 0$. Let $\overline{U}_{iat} = \beta_{0,k_i,t} + \beta_{1,k_i,t} POS_{ia} + \beta_2 NUM\_AD_i + \tau_a$, $\tilde{\epsilon}_{ia} = POS_{ia} - \phi_{0,k_i} - \phi_{1,k_i} \overline{POS_{a,-k_i}} - \phi_2 NUM\_AD_i$, then

$$\mu_{iat} = \overline{U}_{iat} + \frac{\sigma_{12}}{\sigma^2 + \sigma_{12}} \tilde{\epsilon}_{ia}$$

$$\sigma_{1|2} = 1 - \frac{\sigma_{12}}{\sigma^2 + \sigma_{12}} = \frac{\sigma^2}{\sigma^2 + \sigma_{12}}$$

Step 3. Draw $\Gamma_k = (\chi^k, \phi^k)'$ for each keyword k.

For each keyword *k*, let N_k be the number of observations such that $k_i = k$. Let $\Gamma_k = (\chi^k, \phi^k)'$, $z_{1ia} = (1, POS_{ia})'$, $z_{2ia} = \left(1, \overline{POS}_{a,-k_i}\right)'$, $y_{1ia} = U_{iat} - z'_{1ia}\gamma_t - \beta_2 NUM\_AD_i - \tau_a$, $y_{2ia} = POS_{ia} - \phi_2 NUM\_AD_i$. Then

$$y_{1ia} = z'_{1ia}\chi^k + \eta_{ia}$$

$$y_{2ia} = z'_{2ia}\phi^k + \epsilon_{ia}$$

where $(\eta_{ia}, \epsilon_{ia})' \sim MVN(0, \Lambda)$. We can write it in matrix version as

$$y_{1k} = Z'_{1k}\chi^k + \eta_k$$
$$y_{2k} = Z'_{2k}\phi^k + \epsilon_k$$

or more compactly, as

$$Y = Z'_k \Gamma_k + E_k$$

where $Y_k = (y_{1k}, y_{2k})'$, $Z_k = \begin{pmatrix} Z_{1k} & 0 \\ 0 & Z_{2k} \end{pmatrix}$, and $E_k = (\eta_k, \epsilon_k)' \sim MVN(0, \Lambda \otimes I_{N_k})$.

We can rewrite $\Delta^\beta = \begin{pmatrix} \Delta^\beta_{11} & \dots & \Delta^\beta_{1r} \\ \Delta^\beta_{21} & \dots & \Delta^\beta_{2r} \end{pmatrix}$ as a vector $\delta^\beta = vecr(\Delta^\beta) = \left(\Delta^\beta_{11}, \dots, \Delta^\beta_{1r}, \Delta^\beta_{21}, \dots, \Delta^\beta_{2r}\right)'$. Similarly, $\delta^\phi = vecr(\Delta^\phi)$.

With prior distribution $\Gamma_k \sim MVN(\overline{\Gamma}_k, \psi_0)$, where $\overline{\Gamma}_k = [I_4 \otimes X'_k]\begin{pmatrix} \delta^\beta \\ \delta^\phi \end{pmatrix}$ and $\psi_0 = \begin{pmatrix} \Omega^\beta & 0 \\ 0 & \Omega^\phi \end{pmatrix}$, we can draw Γ_k from the following posterior distribution:

$$\Gamma_k | all\ other\ parameters \sim MVN(\widetilde{\Gamma_k}, \tilde{\psi})$$

where $\widetilde{\Gamma_k} = \tilde{\psi}\left[Z'_k(\Lambda^{-1} \otimes I_{N_k})Y_k + \psi_0^{-1}\overline{\Gamma_k}\right]$, and $\tilde{\psi} = \left[Z'_k(\Lambda^{-1} \otimes I_{N_k})Z_k + \psi_0^{-1}\right]^{-1}$.

Step 4. Draw γ_t for each topic t.

For each topic t, let N_t be the number of observations with $t_i = t$. Let $Z_{ia} = (1, POS_{ia})'$, $\tilde{\epsilon}_{ia} = POS_{ia} - \phi_{0,k_i} - \phi_{1,k_i} - \phi_2 NUM\_AD_i$, $U^1_{iat} = U_{iat} - Z_{ia}\chi^k - \beta_2 NUM\_AD_i - \tau_a - \frac{\sigma_{12}}{\sigma^2 + \sigma^2_{12}}\tilde{\epsilon}_{ia}$, and $\sigma_{1|2} = \frac{\sigma^2}{\sigma^2 + \sigma^2_{12}}$. Then $U^1_{iat} \sim N(Z'_{ia}\gamma_t, \sigma_{1|2})$. We can write it in matrix version as

$$U^{1t} \sim MVN(Z^t\gamma_t, \sigma_{1|2} I_{N_t})$$

where $U^{1t}: N_t \times 1$ includes all U^1_{iat} such that $t_i = t$, and Z^t is a $N_t \times 2$ matrix. With prior $\gamma_t \sim MVN(0, \Psi)$, we then draw γ_t from the following posterior distribution:

$$\gamma_t | all\ other\ parameters \sim MVN(\widetilde{\gamma_t}, \widetilde{\Psi})$$

where $\widetilde{\Psi} = \left[(\Psi)^{-1} + (Z^t)'Z^t/\sigma_{1|2}\right]^{-1}$ and $\widetilde{\gamma_t} = \widetilde{\Psi}\left[(Z^t)'U^{1t}/\sigma_{1|2}\right]$.

Step 5. Draw τ_a for each ad a.

For each ad a, let n_a be the number of observations. We define $\tilde{\epsilon}_{ia} = POS_{ia} - \phi_{0,k_i} - \phi_{1,k_i} - \phi_2 NUM\_AD_i$, $U^2_{iat} = U_{iat} - (\beta_{0,k_i,t} + \beta_{1,k_i,t} POS_{ia}) - \beta_2 NUM\_AD_i - \frac{\sigma_{12}}{\sigma^2 + \sigma^2_{12}}\tilde{\epsilon}_{ia}$, and $\sigma_{1|2} = \frac{\sigma^2}{\sigma^2 + \sigma^2_{12}}$. Then $U^2_{iat} \sim N(\tau_a, \sigma_{1|2})$. With prior $\tau_a \sim N(0, v^\tau)$, the posterior distribution of τ_a is

$$\tau_a | all\ other\ parameters \sim MVN(\widetilde{\tau_a}, \widetilde{v^\tau})$$

where $\widetilde{\tau_a} = \frac{n_a v^\tau U^2_{iat}}{n_a v^\tau + \sigma_{1|2}}$ and $\widetilde{v^\tau} = \frac{\sigma_{1|2} v^\tau}{n_a v^\tau + \sigma_{1|2}}$.

Step 6. Draw β_2.

Let $\tilde{\epsilon}_{ia} = POS_{ia} - \phi_{0,k_i} - \phi_{1,k_i} - \phi_2 NUM\_AD_i$, $U_{iat}^3 = U_{iat} - \left(\beta_{0,k_i,t} + \beta_{1,k_i,t} POS_{ia}\right) - \beta_2 NUM\_AD_i - \frac{\sigma_{12}}{\sigma^2 + \sigma_{12}^2} \tilde{\epsilon}_{ia}$, $X_i = NUM\_AD_i$, then $U_{iat}^3 \sim N\left(\beta_2 X_i, \sigma_{1|2}\right)$. With prior $\beta_2 \sim N\left(0, v^{\beta_2}\right)$, the posterior distribution of β_2 is

$$\beta_2 | all\ other\ parameters \sim N\left(\widetilde{\beta_2}, \widetilde{v^{\beta_2}}\right)$$

where $\widetilde{v^{\beta_2}} = \left[\sigma_{1|2}^{-1} X'X + \left(v^{\beta_2}\right)^{-1}\right]^{-1}$, and $\widetilde{\beta_2} = v^{\beta_2}\left[\sigma_{1|2}^{-1} X'U^3\right]$.

Step 7. Draw ϕ_2.

Let $\tilde{\eta}_{ia} = U_{iat} - \left(\beta_{0,k_i,t} + \beta_{1,k_i,t} POS_{ia}\right) - \beta_2 NUM\_AD_i - \tau_a$, $w_{ia} = POS_{ia} - \left(\phi_{0,k_i} + \phi_{1,k_i} \overline{POS_{a,-k_i}}\right) - \sigma_{12} \tilde{\eta}_{ia}$, $X_i = NUM\_AD_i$, then $w_{ia} \sim N(\phi_2 X_i, \sigma^2)$. With prior $\phi_2 \sim N\left(0, v^{\phi_2}\right)$, the posterior distribution of ϕ_2 is

$$\phi_2 | all\ other\ parameters \sim N\left(\widetilde{\phi_2}, \widetilde{v^{\phi_2}}\right)$$

where $\widetilde{v^{\phi_2}} = \left[\sigma^{-2} X'X + \left(v^{\phi_2}\right)^{-1}\right]^{-1}$, and $\widetilde{\phi_2} = v^{\phi_2}[\sigma^{-2} X'w]$.

Step 8. Draw Δ^β.

Let K be the number of keywords. With $\Delta^\beta = \begin{pmatrix} \Delta_{11}^\beta & \cdots & \Delta_{1r}^\beta \\ \Delta_{21}^\beta & \cdots & \Delta_{2r}^\beta \end{pmatrix}$, we have $\delta^\beta = vecr\left(\Delta^\beta\right) = \left(\Delta_{11}^\beta, \dots, \Delta_{1r}^\beta, \Delta_{21}^\beta, \dots, \Delta_{2r}^\beta\right)'$. Therefore, $\chi^k = \begin{pmatrix} X_k' & 0 \\ 0 & X_k' \end{pmatrix} \delta^\beta + u_k^\beta$, where $u_k^\beta \sim MVN\left(0, \Omega^\beta\right)$. We can rewrite this in matrix format as

$$\begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix} = \begin{pmatrix} X & 0 \\ 0 & X \end{pmatrix} \delta^\beta + E$$

where $\chi_0 = (\chi_{01}, \dots, \chi_{0K})'$, $\chi_1 = (\chi_{11}, \dots, \chi_{1K})'$, $X = (X_1, \dots, X_K)'$, and $E \sim MVN\left(0, \Omega^\beta \otimes I_K\right)$. More compactly, we can write

$$\chi = (I_2 \otimes X)\delta^\beta + E$$

With prior $\delta^\beta \sim MVN\left(\overline{\Delta^\beta}, A^\beta\right)$, we can draw δ^β from the following posterior distribution:

$$\delta^\beta | all\ other\ parameters \sim MVN\left(\widetilde{\Delta^\beta}, \widetilde{A^\beta}\right)$$

where $\widetilde{A^\beta} = \left[(I_2 \otimes X)'\left(\left(\Omega^\beta\right)^{-1} \otimes I_K\right)(I_2 \otimes X) + \left(A^\beta\right)^{-1}\right]^{-1}$ and $\widetilde{\Delta^\beta} = \widetilde{A^\beta}\left[(I_2 \otimes X)'\left(\left(\Omega^\beta\right)^{-1} \otimes I_K\right)\chi + \left(A^\beta\right)^{-1}\overline{\Delta^\beta}\right]$.

Step 9. Draw Δ^ϕ.

Similar to the previous step, with prior $\delta^\phi \sim MVN\left(\overline{\Delta^\phi}, A^\phi\right)$, we can draw δ^ϕ from the following posterior distribution:

$$\delta^\phi | all\ other\ parameters \sim MVN\left(\widetilde{\Delta^\phi}, \widetilde{A^\phi}\right)$$

where $\widetilde{A^{\phi}} = \left[(I_2 \otimes X)' \left((\Omega^{\phi})^{-1} \otimes I_K\right)(I_2 \otimes X) + \left(A^{\phi}\right)^{-1}\right]^{-1}$ and $\widetilde{\Delta^{\phi}} = \widetilde{A^{\phi}}\left[(I_2 \otimes X)' \left((\Omega^{\phi})^{-1} \otimes I_K\right)\chi + \left(A^{\phi}\right)^{-1}\overline{\Delta^{\phi}}\right]$.

Step 10. Draw v^{τ}.

With prior $v^{\tau} \sim IG(m, n)$, we can draw v^{τ} from its posterior distribution $IG\left(m + \frac{A}{2}, n + \frac{\tau'\tau}{2}\right)$, where A is the total number of unique ads.

Step 11. Draw σ_{12}.

With prior $\sigma_{12} \sim N(0, b_0)$, we can draw σ_{12} from its posterior distribution $N(\tilde{r}, \tilde{b})$, where $\tilde{r} = (\sigma^{-2}\tilde{\eta}'\tilde{\epsilon})$ and $\tilde{b} = (\sigma^{-2}\tilde{\eta}'\tilde{\eta} + b_0^{-1})^{-1}$.

Step 12. Draw σ^2.

With prior $\sigma^2 \sim IG(v_0, c_0)$, we can draw σ^2 from its posterior distribution $IG(\tilde{v}, \tilde{c})$, where $\tilde{v} = v_0 + \frac{N}{2}$ and $\tilde{c} = c_0 + \frac{(\tilde{\epsilon}-\tilde{\eta}\sigma_{12})'(\tilde{\epsilon}-\tilde{\eta}\sigma_{12})}{2}$. N is the number of observations.

Step 13. Draw Ω^{β}.

With prior $\Omega^{\beta} \sim IW(v^{\beta}, V^{\beta})$, we can draw Ω^{β} from its posterior distribution $IW\left(v^{\beta} + K, V^{\beta} + \sum_{k=1}^{K}(\chi^k - \Delta^{\beta}X_k)(\chi^k - \Delta^{\beta}X_k)'\right)$, where K is the number of keywords.

Step 14. Draw Ω^{ϕ}.

With prior $\Omega^{\phi} \sim IW(v^{\phi}, V^{\phi})$, we can draw Ω^{ϕ} from its posterior distribution $IW\left(v^{\phi} + K, V^{\phi} + \sum_{k=1}^{K}(\phi^k - \Delta^{\phi}X_k)(\phi^k - \Delta^{\phi}X_k)'\right)$, where K is the number of keywords.

Step 15. Draw Ψ.

With prior $\Psi \sim IW(v^{\Psi}, V^{\Psi})$, we can draw Ψ from its posterior distribution $IW(v^{\Psi} + T, V^{\Psi} + \sum_{t=1}^{T}\gamma_t\,\gamma_t')$, where T is the number of topics.

Appendix H

Analysis on Time Before First Click

Although we do not directly observe consumers' click behavior on organic search results, we are able to infer the time on the organic links by observing when a consumer starts a search session by entering a keyword, and when the consumer clicks on the first ad. Therefore, we focus our analysis on the subset of consumers who have clicked on at least one sponsored ad. We denote *DURATION* as the time between the consumer starting a search session and making the first click, and we use *DURATION* as a proxy for the time spent on the organic listing. We run a linear regression of *DURATION* on the keyword attributes of interests. As shown in Table H1, higher keyword ambiguity is associated with less time spent on organic search results, while more precise keywords tend to attract more attention on organic search results. This result provides partial evidence that a more ambiguous keyword may reduce the attractiveness of organic search results, and consumers may turn to sponsored ads for finding an alternative that meets their needs.

Table H1. Regression Results: Time Before First Click

Variable	Estimates	
Intercept	56.925***	(0.675)
TOPIC_ENTROPY	-4.643***	(0.232)
NUM_WORDS	0.335**	(0.155)
BRAND	-2.068***	(0.210)
LOCATION	-1.200***	(0.301)
LOG_TRANS	-2.895***	(0.090)
LOG_IMP	-2.674***	(0.057)
Observations	551,239	

***, **, and * indicate a 99%, 95%, and 90% significance level.

Appendix I

Boxplots of Topic-Specific Effects

To present the heterogeneous impact on CTR across topics, we overlay boxplots of the topic specific intercepts (i.e., $\gamma_{0,t}$) in Figure I1 and the topic specific effects of position (i.e., $\gamma_{1,t}$) in Figure I2. Because we have mean-centered all keyword characteristics when estimating the hierarchical Bayesian model, $\gamma_{0,t}$ and $\gamma_{1,t}$ can be interpreted as estimates of $\beta_{0,k,t}$ and $\beta_{1,k,t}$ for a typical keyword in topic *t* of which the covariates are set to mean values. As we can see from Figure I1, the means of the posterior distribution of the topic specific intercepts are highest for topics "travel" and "health," suggesting that consumers who are interested in those topics may be more likely to click on ads at top positions. In contrast, for topics "sport" and "adult," CTR is lower at top positions.

As we can see from Figure I2, the means of the posterior distribution of the topic specific effects of position are highest for topics "home" and "documents," suggesting that consumers who are interested in those topics may be more likely to click on ads at lower positions. In contrast, for topics "music" and "clothing," CTR decreases faster with position.

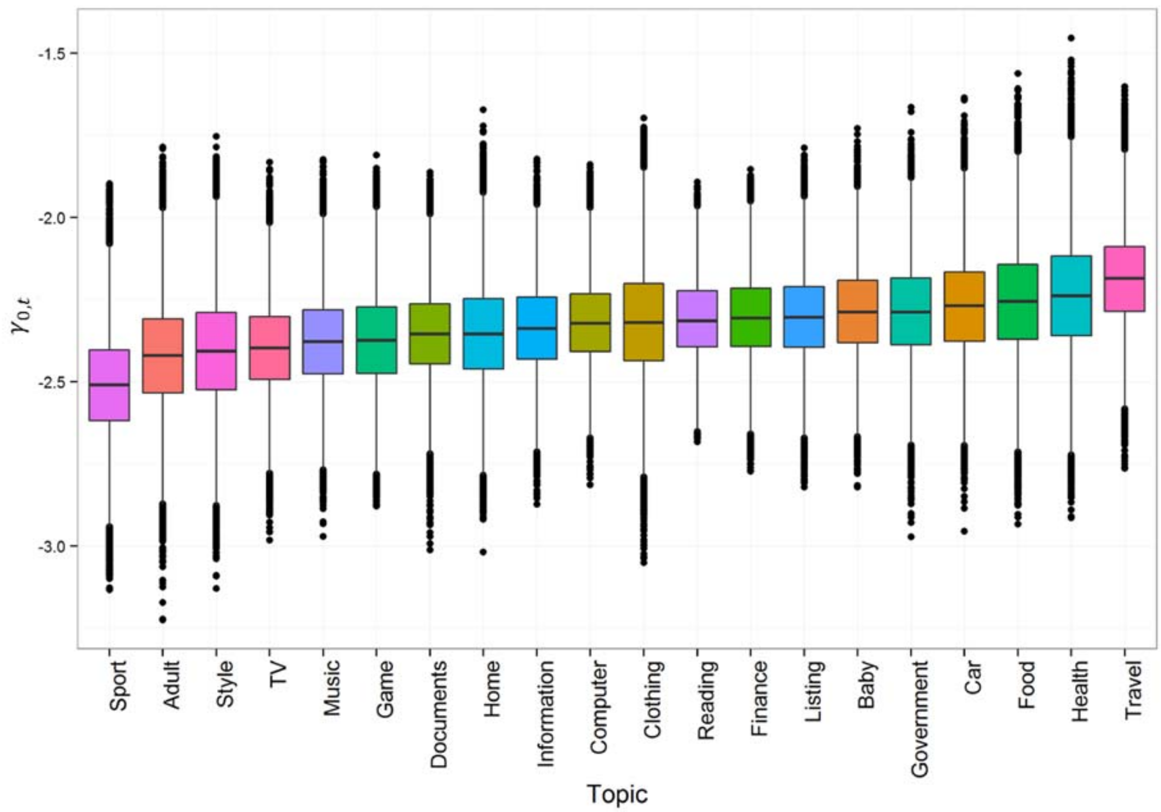

Figure I1. Boxplots of Topic-Specific Intercepts ($\gamma_{0,t}$)

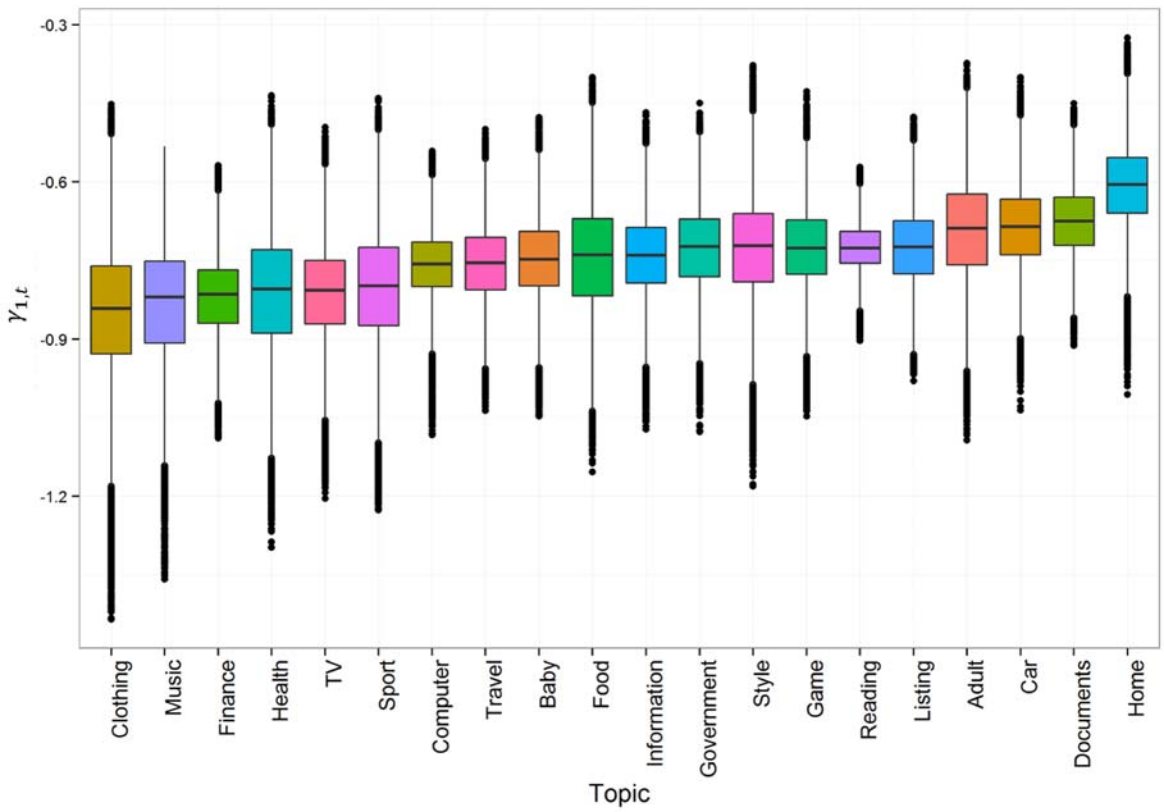

Figure I2 Boxplots of Topic-Specific Effects of Position (γ_{1t})

Appendix J

Number of Organic Results for Corpus Construction

We have compared the topic entropy values and empirical estimation results based on different numbers of Google results (i.e., top 50, top 60, top 80, and top 100), and present the comparisons below.

Comparing topic entropy. In Table J1, we present the summary statistics for the computed topic entropy of the full data set (12,790 keywords) based on different numbers of organic search results. The high correlations among entropy values derived based on different numbers of organic search results suggest that entropy values seem to be fairly robust to the number of organic search results used to construct the corpus for topic modeling.

Table J1. Entropy Values Based on Different Number of Google Organic Search Results

					Correlation			
	Mean	**SD**	**Min**	**Max**	**Top 50**	**Top 60**	**Top 80**	**Top 100**
Top 50	1.60	0.45	0.34	2.99	1	0.87	0.86	0.85
Top 60	1.97	0.41	0.44	3.00	0.87	1	0.97	0.96
Top 80	1.99	0.40	0.44	3.00	0.86	0.97	1	0.97
Top 100	2.00	0.40	0.44	3.00	0.85	0.96	0.97	1

Comparing empirical results. We have further reestimated the hierarchical Bayesian model using the entropy values and topic probabilities we now obtained based on different numbers of organic search results. We present the main results for CTR in Table J2. As can be seen, the estimation results are fairly consistent across different columns, suggesting that our main results are robust to the number of organic search results used to cover the topics related to each keyword.

Table J2. Estimation Results for CTR Based on Different Number of Google Organic Search Results

	Variable	Top 50		Top 60		Top 80		Top 100	
Baseline (β_{0kt})	Intercept	-2.308***	(0.122)	-2.336***	(0.126)	-2.316***	(0.108)	-2.318***	(0.115)
	TOPIC_ENTROPY	0.192***	(0.069)	0.223***	(0.074)	0.184**	(0.076)	0.165**	(0.072)
	NUM_W ORDS	0.040	(0.045)	0.049	(0.045)	0.038	(0.046)	0.033	(0.045)
	BRAND	0.033	(0.064)	0.042	(0.064)	0.040	(0.064)	0.040	(0.064)
	LOCATION	-0.208**	(0.105)	-0.217**	(0.100)	-0.212**	(0.098)	-0.207**	(0.098)
	LOG_T RANS	0.147***	(0.029)	0.150***	(0.030)	0.152***	(0.029)	0.153***	(0.029)
	LOG_IMP	-0.010	(0.019)	-0.004	(0.020)	-0.004	(0.019)	-0.005	(0.019)
POS (β_{1kt})	Intercept	-0.726***	(0.045)	-0.749***	(0.066)	-0.727***	(0.059)	-0.724***	(0.055)
	TOPIC_ENTROPY	-0.134***	(0.049)	-0.115**	(0.048)	-0.114**	(0.048)	-0.113**	(0.049)
	NU M_W ORDS	-0.035	(0.032)	-0.035	(0.032)	-0.034	(0.030)	-0.034	(0.031)
	BRAND	-0.050	(0.045)	-0.056	(0.047)	-0.051	(0.044)	-0.047	(0.045)
	LOCATION	0.070	(0.064)	0.075	(0.065)	0.072	(0.063)	0.075	(0.063)
	LOG_TRANS	-0.014	(0.019)	-0.013	(0.019)	-0.015	(0.019)	-0.014	(0.019)
	LOG_IMP	-0.018	(0.013)	-0.019	(0.013)	-0.021*	(0.013)	-0.021	(0.013)
NUM_ADS (β_2)	*NUM_ADS*	0.166***	(0.017)	0.168***	(0.018)	0.167***	(0.015)	0.165***	(0.016)

***, **, and * indicate a 99%, 95%, and 90% significance level.

References

Agarwal, A., Hosanagar, K., and Smith, M. D. 2011. "Location, Location, Location: An Analysis of Profitability of Position in Online Advertising Markets," *Journal of Marketing Research* (48:6), pp. 1057-1073.

Agarwal, A., Hosanagar, K., and Smith, M. D. 2015. "Do Organic Results Help or Hurt Sponsored Search Performance?," *Information Systems Research* (26:4), pp. 695-713.

Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. "Latent Dirichlet Allocation," *Journal of Machine Learning Research* (3), pp. 993-1022.

Chan, T. Y., and Park, Y. H. 2015. "Consumer Search Activities and the Value of Ad Positions in Sponsored Search Advertising," *Marketing Science* (34:4), pp. 606-623.

Chan, T. Y., Wu, C., and Xie, Y. 2011. "Measuring the Lifetime Value of Customers Acquire from Google Search Advertising," *Marketing Science* (30:5), pp. 837-850.

Dai, H. K., Zhao, L., Nie, Z., Wen, J.-R., Wang, L., and Li, Y. 2006. "Detecting Online Commercial Intention (OCI)," in *Proceedings of the 15th International Conference on World Wide Web*, New York: ACM, pp. 829-837.

Ghose, A., and Yang, S. 2009. "An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets," *Management Science* (55:10), pp. 1605-1622.

Goldfarb, A., and Tucker, C. 2011. "Search Engine Advertising: Channel Substitution When Pricing Ads to Context," *Management Science* (57:3), pp. 458-470.

Griffiths, T. L., and Steyvers, M. 2004. "Finding Scientific Topics," *Proceedings of the National Academy of Science* (101:Supplement 1), pp. 5228-5235.

Jerath, K., Ma, L., and Park, Y.-H. 2014. "Consumer Click Behavior at a Search Engine: The Role of Keyword Popularity," *Journal of Marketing Research* (51:4), pp. 480-486.

Jeziorski, P., and Segal, I. 2015. "What Makes Them Click: Empirical Analysis of Consumer Demand for Search Advertising," *American Economic Journal: Microeconomics* (7:3), pp. 24-53.

Levenshtein, V. I. 1966. "Binary Codes Capable of Correcting Deletions, Insertions and Reversals," *Soviet Physics Doklady* (10), pp. 707-710.

McCallum, A. K. 2002. "MALLET: A Machine Learning for Language Toolkit" (http://mallet.cs.umass.edu).

Rutz, O. J., and Bucklin, R. E. 2011. "From Generic to Branded: A Model of Spillover in Paid Search Advertising," *Journal of Marketing Research* (48:1), pp. 87-102.

Rutz, O. J., Bucklin, R. E., and Sonnier, G. P. 2012. "A Latent Instrumental Variables Approach to Modeling Keyword Conversion in Paid Search Advertising," *Journal of Marketing Research* (49:3), pp. 306-319.

Rutz, O. J., and Trusov, M. 2011. "Zooming In on Paid Search Ads—A Consumer-Level Model Calibrated on Aggregated Data," *Marketing Science* (30:5), pp. 789-800.

Rutz, O. J., Trusov, M., and Bucklin, R. E. 2011. "Modeling Indirect Effects of Paid Search Advertising: Which Keywords Lead to More Future Visits?," *Marketing Science* (30:4), pp. 646-665.

Steyvers, M., and Griffiths, T. 2007. "Probabilistic Topic Models," *Handbook of Latent Semantic Analysis* (427:4), pp. 424-440).

Yang, S., and Ghose, A. 2010. "Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence?," *Marketing Science* (29:4), pp. 602-623.

Yang, S., Lu, S., and Lu, X. 2014. "Modeling Competition and Its Impact on Paid-Search Advertising," *Marketing Science* (33:1), pp. 134-153.

Yao, S., and Mela, C. F. 2011. "A Dynamic Model of Sponsored Search Advertising," *Marketing Science* (30:3), pp. 447-468.