

Exploiting Contextual Factors for Click Modeling in Sponsored Search

Dawei Yin^{†*} Shike Mei^{†*} Bin Cao[§] Jian-Tao Sun[§] Brian D. Davison[†]

[†]Dept. of Computer Science and Engineering, Lehigh University, Bethlehem, PA, USA

[‡]Dept. of Computer Science, University of Wisconsin-Madison, Madison, WI, USA

[§]Microsoft Research Asia, Beijing, China

[†]{day207, davison}@cse.lehigh.edu [‡]mei@cs.wisc.edu [§]{bincao, jtsun}@microsoft.com

ABSTRACT

Sponsored search is the primary business for today’s commercial search engines. Accurate prediction of the Click-Through Rate (CTR) for ads is key to displaying relevant ads to users. In this paper, we systematically study the two kinds of contextual factors influencing the CTR: 1) In micro factors, we focus on the factors for mainline ads, including ad depth, query diversity, ad interaction. 2) In macro factors, we try to understand the correlations of clicks between organic search and sponsored search. Based on this data analysis, we propose novel click models which harvest these new explored factors. To the best of our knowledge, this is the first paper to examine and model the effects of the above contextual factors in sponsored search. Extensive experiments on large-scale real-world datasets show that by incorporating these contextual factors, our novel click models can outperform state-of-the-art methods.

Categories and Subject Descriptors: H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval;

Keywords: CTR prediction, Click Through Rate, Ad CTR prediction, Sponsored search, Click model

1. INTRODUCTION

Sponsored search is the primary business for today’s commercial search engines. Pay-per-click (PPC) is an advertising model that has been adopted by the sponsored search markets. Under the PPC model, advertisers are charged when their advertisements are clicked by search engine users. More clicks bring more revenue to search engine companies [6]. However, since showing too many ads on search result pages may hurt the user experience, search engine companies have to be conservative in the number of ads shown and try to display those that earn the most money which is a function of click-through rate (e.g., $\text{bid} \times \text{CTR}$). Therefore, estimating the click-through rate of a search ad is an important problem for commercial search engines.

To better understand the challenges and difficulties of estimating click-through rate in sponsored search, examine the results of the

search query “iPad deal” in Figure 1. We see that when a user issues a query, both a list of web pages (organic search) and two lists of ads (sponsored search) are displayed. Clearly, to estimate the click-through rate of a target ad, there are many factors to explore. Focusing on the mainline ads (top ads), we have the following the micro factors: ad depth (the total number of the ads displayed), the interaction between the target ad and query, and the interaction between the target and the adjacent ads. For macro factors, there may be interactive influences between sponsored search and organic search results [13].

As previous studies have shown, user clicks contain various biases [22, 8, 17, 12], which is both true in organic search and sponsored search [39]. But unlike organic search, sponsored search has additional biases and characteristics. For example, many people tend to skip ads and they are reluctant to click ads even if they are related to the query. Previous Cascade Click Models [12] assume that the top position is certainly examined by user ($P(E_1) = 1$) and is thus not suitable for modeling user click behavior in sponsored search. Intuitively, the probability of a user examining the sponsored search results is related to two factors: the depth of the ads (the number of ads displayed) and the commercial intent of a query. For the depth of the ads, the more ads that are shown at the same time, the larger the ad area is, the more attention from users is attracted. However, the search engine usually should not simply show too many ads due to the poor user experience. For the second feature, we notice that for some queries with a high commercial intent, ads usually are more likely to be interesting to users. This indicates that the examination probability is likely to be query dependent. The effects of both factors on examination and click-through rate are still not clear and have not been systematically investigated. Another key factor is interactive influence between the target ad and its context (e.g., query and adjacent ads). Previous research into mutual influence between similar ads [33] shows that there exists a mutually exclusive relationship between similar ads. However, in general, the mutual relationship between ads is not necessarily exclusive and negative, but could also be positive [34]. Unfortunately, the relational click model [33] fails to model positive effects between the ads. Xu et al. [34] study the cases of the relationship between two ads displayed together, but the case of more ads (e.g., 3 or 4 ads) is still unknown. No previous research has investigated the interaction between the target ad and its context, including both query and adjacent ads.

Beyond the factors within ads, the relationship between organic search and sponsored search is also a major factor influencing CTR and yet has not been systematically explored. Statistical analysis shows that about 80% of clicks go to organic search while approx-

*Equal contribution

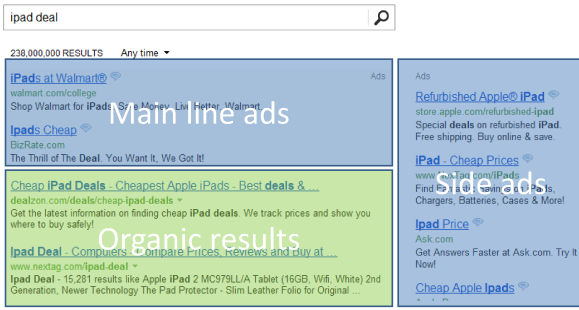


Figure 1: Example: The returned results for a query include ads and search results. The ads can be displayed on the left side (mainline-ads) as well as the right side (side-ads)

imately 5% go to ads [9], which is an order of magnitude smaller.¹ Moreover, the clicks also follow a power law distribution with respect to queries and ads. That is to say, most queries and ads fall into the “long tail” region and thus suffer from a limited number of user clicks. The sparsity problem and the long tail distribution introduce intrinsic difficulties to obtaining accurate CTR prediction results. Previous studies tried to solve the sparsity problem by exploring the rich features of ads, queries, and their relationships [18, 14, 33]. The hierarchical information on ads is leveraged and collaborative filtering techniques are utilized to smooth the model [27]. However, most current CTR prediction models are still limited by using the data within the ad domain. However, we have observed that exclusively depending on the ad domain is not ideal since most users seldom click on ads. Also, the information reflected by user behavior in organic search results is relatively richer and can potentially help ad click prediction. This phenomenon has also been observed in a recent study in [13], which shows there is a competitive relationship between search and ads in navigational queries. However, the question of how to leverage user behavior in organic search to address the sparsity problem in CTR prediction is still an under-explored research problem.

In this paper, we focus on analyzing mainline ads and systematically study the contextual factors which have not been well explored. We start with data analysis to understand the user clicks in sponsored search. For micro factors, we first investigate a number of contextual factors including ad depth, query diversity, ad interaction. Based on our analysis, we incorporate these factors into a novel click model. Experimental results show that the model with these new contextual factors can outperform three state-of-the-art methods. We then perform data analysis to understand the click correlations between organic search and ads (i.e., the macro factors), which illustrates the possibility of transferring the knowledge from the rich data in organic search to the ad CTR prediction problem. Our method jointly models user responses to sponsored ads and organic search results. The knowledge contained in the organic search log is utilized and transferred to the ad domain. We further discovered that the correlations between clicks in ads and organic search are query-dependent. To the best of our knowledge, this is the first paper to examine the effects of the above contextual factors in sponsored search together.

2. RELATED WORK

There are two major assumptions in click models: the Examination Hypothesis and the Cascade Model. The examination hypothesis assumes that if a displayed URL is clicked, it must be

¹The summation is not 1 because clicks could also go to other links such as query suggestions.

both examined and relevant [30, 37, 8, 12, 17, 26, 25, 32]. They have been compared by O’Chapelle [8] and experimentally found to be outperformed by the cascade model [12], which assumes that the user views search results from top to bottom and decides whether to click each url. Once a click is issued, documents below the clicked result are not examined regardless of the position. Then the dependent click model [20] generalizes the cascade model to allow multiple clicks within a single session. The click chain model [19] and dynamic Bayesian network [8] (which is inferred through Infer.NET [28]) provide a further generalization by allowing the user to abandon examination of more results. The general click model [39] treats all relevance and examination effects in the model as random variables. In addition to the above methods, there have been several click models following the above two assumptions but apply them in different contexts, such as the task centric click model [38], and the position normalized click model [10]. However, the above methods are mainly designed for understanding user click behavior in organic search. They fail to explore the specific properties in sponsored search as discussed earlier.

Recently, Xu et al. investigated relational click behavior [34], but their method can only be applied in cases with two ads. They fail to model the more general cases (e.g., 3, 4 or more ads) or handle new ads and queries. Xiong et al. [33] showed the mutually exclusive influence between similar ads, but their model fails to include positive effects between the ads. Also, they do not take examination bias into account. Additionally, there has been some research that addresses the relationship between organic search and sponsored advertising [35, 13]. However, Danescu-Niculescu-Mizil et al. [13] only considered the global correlation between clicks in search and advertising. Yang and Ghose [35] only consider this problem on a very restricted set of ads.

Generally speaking, previous studies [15, 18, 1, 14, 30, 29, 37, 11] mainly focus on building a linear model based on features extracted from a query and its ads. On the other hand, the ad CTR prediction problem can be also formulated as a recommendation problem and collaborative filtering (CF) techniques can be applied [27]. As shown in several recent competitions (e.g., the Netflix Prize and KDDCUP 2011), collaborative filtering (CF) techniques, especially matrix factorization models, are very effective for recommendation problems [24, 4, 16]. In contrast to general recommendation tasks such as movie or music recommendations, the data of ad CTR prediction is far more sparse. Menon et al. [27] proposed using a matrix factorization model for ad CTR prediction and achieved state-of-the-art performance. However, we argue that, even with the help of auxiliary information, the ad CTR prediction problem still cannot be well-solved due to the limited number of user clicks we obtain from users. By incorporating other related domains into the model, we can exploit the inherent relationships across domains and help either or both domains.

Recently, Ashkan et al. [3, 2] conduct a preliminary study on the factors for click analysis in sponsored search: in [2], they study query commercial intent effects in sponsored search; in [3], they introduce query-bias into the click model. However, effects of other contextual factors still remain unknown and have not been modeled effectively, such as ad depth, interactive influence between ads and correlations between organic search and sponsored search. In this paper, we will focus on analyzing and modelling these factors.

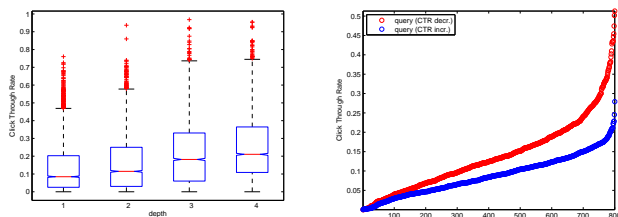
3. MICRO FACTORS ANALYSIS

In this section, we verify our intuitions and conjectures for mainline ads in sponsored search. We collected one month of data from a commercial search engine. In total, there are 127,230,609 query

Table 1: Average ad CTR at different positions and depths.

	pos. 1	pos. 2	pos. 3	pos. 4	Depth Avg.
depth 1	0.1298				0.1298
depth 2	0.1568	0.0450			0.1009
depth 3	0.2158	0.0511	0.0306		0.0992
depth 4	0.2479	0.0711	0.0430	0.0341	0.0990
Pos. Avg.	0.1586	0.0547	0.0382	0.0341	

sessions, where a session consists of an input query, a list of sponsored search results and a list of clicked positions with time stamps.



(a) CTR on position 1 as a function of depth (b) Depth changed from 1 to 3

Figure 2: Representative analyses

3.1 Depth Analysis

We first try to answer the following question: is ad CTR related to the number of ads displayed? Table 1 shows the average CTR at different positions and depth of mainline ad block. From the bottom line, we can see that the relationship between CTR and position is consistent with previous studies [30, 12]. Ads in the top positions usually receive higher CTR. We can also make some interesting observations about the relationship between CTR and depth: the CTR at the same position increases along with the depth. That is, as Figure 2(a) shows, the CTR of position 1 and depth 4 is significantly larger than the CTR of position 1 and depth 3, and CTR of position 1 and depth 3 is also larger than the CTR of position 1 and depth 2, etc. And these patterns also exist for position 2 and position 3.

We conjecture that these observations, shown in Table 1 and Figure 2(a), may be the result of four factors: first, greater ad depth typically means a larger ad area, which attracts more user attention. Therefore, ads in a larger ad area have a higher probability of user examination and higher CTRs should be expected. Second, the probability of user examining higher position might be greater than the lower position. Third, after a user examines all the ads, that user tends to click on the most relevant one. More relevant ads absorb more user clicks. The bottom ads actually play a supporting role, which has positive effects on the top ads. Finally, larger depth may imply higher commercial intent of query and higher relevance of the ads in the top position. To verify the conjectures, in Section 4, we incorporate these factors into the click model and the experiments in Section 5 show their contributions on CTR prediction.

We also notice that the average CTR decreases as the depth increases (from the right column of Table 1). This indicates that more ads being displayed may dilute the CTR of specific ads. However, note that the total click yield (the CTR summation of all ads for a given depth) becomes larger: for instance, the click yield of depth 4 is 0.099×4 , which is much larger than depth 1.

3.2 Query Analysis

Query commercial intent has been well explored elsewhere [3, 2]. Here we conduct some experiments which [3, 2] does not cover.

In Section 3.1, we considered the effect of depth on click through rate: From Table 1, the average CTR over all positions (also aver-

Table 2: CTRs of ads appearing in different depths

No. of ads	depth 1	depth 2	depth 3	depth 4
588,619	0.0679	0.0736	0.0895	0.1161

aged over ads and queries) slightly decreases when the depth increases. If we consider micro factors on queries, this may not be always true. For example, for some queries, the average CTR of the larger depth is greater than of the shorter depth. To find the properties of these queries, here we treat the average CTR of the query as the commercial intent of the query (higher average CTR means higher commercial intent), and investigate the relationship between the depth and the query’s commercial intent. In our experiments, we change the depth from 1 to 3, and collected queries that achieve higher CTRs and queries that achieve lower CTRs. In Figure 2(b), we plot the queries whose average CTRs increase in blue and the queries whose average CTRs decrease in red. We see that queries with an increasing CTR have lower commercial intent while queries with a decreasing CTR have higher commercial intent. We perform a two sample t-test on the two groups of CTRs, and the p -value is $3.4710e-43$ which is highly significant. This indicates that for queries with higher commercial intent, a single ad has a higher probability of matching user intent, and the user will more likely click this ad. However, when we show more ads, the average CTR decreases. This is consistent with the analysis by Xiong et al. [33]. Similar ads shown together will have exclusive effects and the CTR can be diluted by adjacent ads. On the other hand, for queries with lower commercial intent, if we show more ads at the same time, we achieve higher average CTRs. We conjecture that for these queries, the effects across ads are positive rather than exclusive or negative. In [33], the authors neglect to consider and model the positive cases.

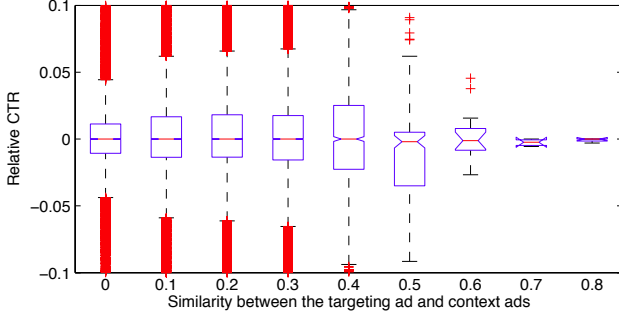
3.3 Ad Analysis

In this section, we analyze the CTR from the perspective of ads. For ads that appear at different depths, we calculate their average CTR at different depths (dropping the ads that only appear in one depth). Table 2 shows that in our data sets, there are 588,619 ads that appear in at least two different depths. As the depth increases, the CTR (averaged over queries) also grows. This observation is consistent with the results in Section 3.1. We then further investigate the averaged CTR changes in specific positions and depths for these ads. The results are shown in Table 3. Take the first line as an example. There are 464,633 ads that appear at both depth 1 and depth 2. For these ads, if we show them alone, the click through rate was 0.0589. When they are shown at depth 2, their CTRs are 0.0676 and 0.0469 at position 1 and position 2, respectively. Similarly, the results of changing the depth from 1 to 3 and 4 are also shown. We can see that if the ad is in position 1, it achieves a higher CTR than when shown alone, but for position 2 and 3, a lower CTR occurs on these ads. However, we also notice that when the depth changes from 1 to 4, the ad CTR of all four positions at depth 4 are greater than the CTR at depth 1, even for the bottom position.

Next, to find ads with mutual influences, we perform a similarity analysis for ads that are shown together. Figure 3 presents the results. The similarity is calculated the same as Xiong et al. [33]. The X-axis indicates the similarity buckets. The Y-axis measures the difference $CTR_{sim} - CTR_{avg}$. In this box plot, the results are similar to [33]. As the similarity becomes higher, it shows negative effects. However, we also notice that for the majority of cases (similarity between 0 and 0.4), this negative effect is almost trivial. The median shows that the positive influence was also in half of the cases. We cannot only consider the exclusive influence between the targeted ad and the adjacent ads, even for highly similar ads. In

Table 3: CTRs of ads appearing in multiple positions

Depth	No. ads	shown alone	pos 1	pos 2	pos 3	pos 4
1→2	464,633	0.0589	0.0676	0.0469		
1→3	117,625	0.0609	0.0862	0.0558	0.0441	
1→4	80,198	0.0547	0.1056	0.0778	0.0628	0.0555


Figure 3: Δ CTR against Similarity

Section 5, we will clearly see the improvement of our interaction model, which can model both cases.

4. CLICK MODEL FOR MAINLINE ADS

We have seen some properties of ad context for predicting click through rate in sponsored search. In this section, we propose the Click Model for Mainline Ads to verify their effectiveness. Let us first define the notation. The records are presented by $\mathbf{S} = (q_m, \mathbf{a}_m, \mathbf{c}_m)_{m=1}^M$. For the single click log $(q, \mathbf{a}, \mathbf{c}) \in \mathbf{S}$, q means the query in the record, \mathbf{a} means the list of ads $\mathbf{a} = \{a_1, a_2, \dots, a_{|\mathbf{a}|}\}$ and $|\mathbf{a}|$ is the ad depth of this record. i is the position of the list of ads \mathbf{a} , and \mathbf{c} means the corresponding clicks $\mathbf{c} = \{c_1, c_2, \dots, c_{|\mathbf{a}|}\}$, where $c_i = 1$ for clicked position i , otherwise $c_i = 0$.

We have already seen some hints that for some queries, the probability of users being attracted by a specific number of ads is different. Here we introduce a new variable, $E_{(q, \mathbf{a})}$ to model the event that given a query q and the displayed ads list \mathbf{a} , users will examine the ad area. Then following the classical examination hypothesis, we have the model assumption: for a record $(q, \mathbf{a}, \mathbf{c}) \in \mathbf{S}$,

$$\begin{aligned} c_i = 1 &\Leftrightarrow E_i = 1, E_{(q, \mathbf{a})} = 1, N_{i, (q, \mathbf{a})} = 1 \\ E_{(q, \mathbf{a})} = 0 &\Rightarrow E_i = 0 \\ E_i = 0 &\Rightarrow c_i = 0 \end{aligned}$$

where E_i is the event that the user will examine the ad at position i and $N_{i, (q, \mathbf{a})}$ is the event that the user is interested in the i th ad, given the context (q, \mathbf{a}) . $N_{i, (q, \mathbf{a})}$ can also be considered as the variable of relevance. For a record $(q, \mathbf{a}, \mathbf{c}) \in \mathbf{S}$, the click is generated as

$$\begin{aligned} P(c_i = 1) &= P(c_i = 1 | E_i = 1, E_{(q, \mathbf{a})} = 1) \\ &P(E_i = 1 | E_{(q, \mathbf{a})} = 1) P(E_{(q, \mathbf{a})} = 1) \end{aligned}$$

The probability of examining position i $P(E_i = 1 | E_{(q, \mathbf{a})} = 1)$ is well studied and can be formulized as the existing versatile position bias model, such as cascade models [19, 20, 12] and the examination hypothesis [17, 32]. In this work, we will not focus on investigating position bias and employed Position Model [32, 31]. We will however study the context-related hypotheses: the examination model for ad area $P(E_{(q, \mathbf{a})})$ and the relevance model $P(c_i = 1 | E_i = 1, E_{(q, \mathbf{a})} = 1)$.

4.1 Examination Models

In this section, we discuss several possible types of examination hypothesis in modeling sponsored search clicks.

Constant Examination (CE). A simple method for modeling the probability of examining ad area $P(E_{(q, \mathbf{a})})$ is assuming $P(E_{(q, \mathbf{a})})$ is independent and does not depend on either query or ad depth. Then we have $P(E_{(q, \mathbf{a})} = 1) = \delta$, where δ is a corpus level parameter and shared by all queries and ad lists. Under this assumption, we might notice that $P(E_i = 1 | E_{(q, \mathbf{a})} = 1) P(E_{(q, \mathbf{a})} = 1) = \delta \cdot \gamma_i$. Clearly, the two parameters can be merged. That is, the effects of the new parameter δ can be absorbed by γ_i without changing prediction performance. Actually, this model is equivalent to the position model [32, 31]. $P(E_i = 1, E_{(q, \mathbf{a})} = 1) = \beta_i \equiv \delta \cdot \gamma_i$

Depth Dependent Examination (DDE). In Section 3, we analyzed the effects of ad depth on click through rate. As the analysis showed, a longer ad list displayed usually meant there was a higher probability of the ad area being examined. Following the intuition that the probability of examining the ad area is highly related to the depth of the ad list, we assume that the probability of examining the ad area $P(E_{(q, \mathbf{a})} = 1)$ is a depth dependent variable. That is $P(E_{(q, \mathbf{a})} = 1) = \delta_{|\mathbf{a}|}$, where $\delta_{|\mathbf{a}|}$ is discriminated by each different depth. Then we have the joint probability: $P(E_i = 1 | E_{(q, \mathbf{a})} = 1) P(E_{(q, \mathbf{a})} = 1) = \delta_{|\mathbf{a}|} \cdot \gamma_i$. Alternatively, as in Constant Examination, we can also merge the two parameters. $P(E_i = 1, E_{(q, \mathbf{a})} = 1) = \beta_{i|\mathbf{a}|} \sim \delta_{|\mathbf{a}|} \cdot \gamma_i$. The merged model, which can be considered as a depth-position bias model with parameter $\beta_{i|\mathbf{a}|}$, actually has freer properties than the original one. However, this depth dependent examination model still cannot capture the query effects on the probability of examining the ad area.

Query-Depth Dependent Examination (QDDE). As in Section 3, we discussed that like depth effects, for queries with different commercial intent, the ad area may also receive different examination probability. Following this intuition, we propose a more delicate model, where we assume the $P(E_{(q, \mathbf{a})} = 1)$ is related to both queries and the depth of ad list. Similarly, $P(E_i = 1 | E_{(q, \mathbf{a})} = 1) P(E_{(q, \mathbf{a})} = 1) = \delta_{q|\mathbf{a}|} \cdot \gamma_i$. Alternatively, as in the above section, we could also merge the two parameters. $P(E_i = 1, E_{(q, \mathbf{a})} = 1) = \beta_{iq|\mathbf{a}|} \sim \delta_{q|\mathbf{a}|} \cdot \gamma_i$. Similarly to the Depth Dependent Examination, the merged model with parameter $\beta_{iq|\mathbf{a}|}$ has freer properties than the original $\delta_{q|\mathbf{a}|} \cdot \gamma_i$. For every query q , we have a depth-position bias model. Although the two factors, query and depth are modeled here, one potential problem of this model is overfitting, due to the large number of parameters behind this model and sparseness of the ad data. We will test this method in the following experiments on the real data.

4.2 Relevance Models

In this section, we will describe the possible methods for the relevance model $P(c_i = 1 | E_i = 1, E_{q|\mathbf{a}|} = 1) = \alpha_{i, (q, \mathbf{a})}$.

Non-Informational Relevance

As with previous methods [17, 32, 12, 39] for the click model, the relevance model can be trivially set to informational constraint on $\alpha_{i, (q, \mathbf{a})}$, that is, $\alpha_{i, (q, \mathbf{a})} \sim \mathcal{U}(0, 1)$, where $\alpha_{i, (q, \mathbf{a})}$ can be valued from uniform distribution in the range zero to one, in order to serve as the Bernoulli parameter of $P(c_i = 1 | E_i = 1, E_{q|\mathbf{a}|} = 1)$. The major drawbacks of this model are three factors: first, since there is no informational constraints on α , the model will more easily become overfitted to training data. Second, it cannot determine the hidden interactions across query, user, and ads. Finally, it cannot handle cold starts.

Informational Relevance

To overcome the problems in the non-information relevance model, we could add some informational constraints on the Bernoulli parameter α . Let $f_{i,(q,\mathbf{a})}$ be the informational constraints, calculated from multiple resources of the context. By placing Gaussian noise, we have $\alpha_{i,(q,\mathbf{a})} \sim \mathcal{N}(\Phi(f_{i,(q,\mathbf{a})}), \sigma^2)$, where $\alpha_{i,(q,\mathbf{a})}$ follows univariate gaussian distribution with mean $\Phi(f_{i,(q,\mathbf{a})})$ and variance σ^2 and the link function $\Phi(x)$ could be Sigmoid function or Gaussian cumulative distribution function. Next, we will discuss several possible models for informational constraints $f_{i,(q,\mathbf{a})}$.

Latent Bias (LBM): One straightforward intuition is that whether the ad is clicked or not depends on the average click rate: $f_{i,(q,\mathbf{a})} = \mu_{a_i}$ where μ_{a_i} is the click through rate of the target ad across all contexts. We can extend this basic estimation by incorporating a wider bias from the context: Query bias μ_q , relevance position bias μ_i , ad list bias $\mu_{\mathbf{a}}$ and global average bias μ_0 . Then we have $f_{i,(q,\mathbf{a})}^{(1)} = \mu_0 + \mu_{a_i} + \mu_q + \mu_i + \mu_{\mathbf{a}}$

Feature Model (FM): Another relevance model is linear estimation, which predicts relevance by a linear combination of features. For a search session, we could collect the features of the ads and their contexts: Let \mathbf{x}_{a_i} be the feature vector of the target ad a_i , which can be extracted from the terms of the shown title and body of the ad, click over expected clicks (COEC) [37] or latent topic distribution of the ad content. Let $\mathbf{x}_{(q,\mathbf{a})}$ be the feature vector of the context of the target ad a_i , which can be the terms of query and the titles of the adjacent ads, user profiles, the depth of the ad list, etc. $f_{i,(q,\mathbf{a})}^{(2)} = \mathbf{b}_1^T \mathbf{x}_{a_i} + \mathbf{b}_2^T \mathbf{x}_{(q,\mathbf{a})}$, where \mathbf{b}_1 and \mathbf{b}_2 are coefficients to be learned from the training set. Here, we further place a zero mean Gaussian prior or Laplace prior on the coefficient \mathbf{b} , corresponding to the L_1 and L_2 regularization respectively.

Interactive Influence (IM): Although linear models are efficient, they are usually over simplified and cannot capture interaction between the ad and the elements (e.g., query and adjacent ads) of the context. To model the interactions between the target ad and the context, we use two interaction matrices $\mathbf{Q} \in \mathbb{R}^{n \times m}$ and $\mathbf{A} \in \mathbb{R}^{n \times n}$, where m is the number of queries and n is the number of ads. The entry $Q_{a_i q}$ of \mathbf{Q} represents the interaction between the target ad a_i and query q , and the entry A_{a_i, a_j} represents the interaction between the target ad a_i and its adjacent ad a_j . $f_{i,(q,\mathbf{a})}^{(3)} = Q_{a_i q} + \sum_{j \neq i, j \leq |\mathbf{a}|} A_{a_i a_j}$ The interaction matrix \mathbf{Q} and \mathbf{A} are unknown and their entries could be either positive or negative. We treat them as latent variables to be learned from the data set. However, the model has two serious problems: the observed pair (a_i, q) and (a_i, a_j) in the training data is extremely sparse, and the majority of entries in the interaction matrix cannot be learned effectively. The second problem is that without any constraints on the interaction matrix, this model may cause overfitting. To avoid these two problems, we place low-rank constraints on the interaction matrix \mathbf{Q} and \mathbf{A} . The low-rank approximation is widely used in recommender systems [23, 7, 36, 5]: $\mathbf{A} \approx \Theta^T \tilde{\Theta}$ and $\mathbf{Q} \approx \Theta^T \Psi$. Let k be the dimensionality of the latent factor vector. $\Theta \in \mathbb{R}^{k \times n}$ is the latent factor matrix for the target ads, $\tilde{\Theta} \in \mathbb{R}^{k \times n}$ is the latent factor matrix for the adjacent ads. Because the ads can act as both the target ad and the adjacent ad, for each ad, there will be two latent factor vectors representing the ad's two roles (the target ad and adjacent ad) respectively. Similarly, $\Psi \in \mathbb{R}^{k \times m}$ is the latent factor matrix for queries. Plug Equation 4.2 back into the interaction model, and we get

$$f_{i,(q,\mathbf{a})}^{(3)} = \theta_{a_i}^T \psi_q + \sum_{j \neq i, j < |\mathbf{a}|} \theta_{a_i}^T \tilde{\theta}_{a_j}$$

Table 4: Data sets

	April 2012	May 2012	total
Impressions	29,722,684	13,160,289	42,882,973
Context (q, \mathbf{a})	112,084	52,814	135,445
Ads	62,423	34,214	72,959
Queries	49,483	26,104	53,730

As with the coefficients in the Feature Model, the latent factor vectors of ads and queries $\theta_{a_i}, \psi_q, \tilde{\theta}_{a_j}$ can be assumed to be generated from Gaussian or Laplace priors, corresponding to L_2 or L_1 regularization respectively. In this paper, we assume the latent factors follow zero mean multi-variate gaussian distribution.

Combined Models (CM): It is straightforward to consider combining the three models LBM, FM and IM. Thus in the combined model, different parts of the model will explain a variety of the ads and contexts. The combined model could be simply: $f_{i,(q,\mathbf{a})} = f_{i,(q,\mathbf{a})}^{(1)} + f_{i,(q,\mathbf{a})}^{(2)} + f_{i,(q,\mathbf{a})}^{(3)}$ Obviously, the combination could also be any two of the three aspects.

4.3 Inference

To learn the parameters, we adopt Maximum a Posteriori (MAP). Let $\Lambda = \{\alpha, \beta, \mu, \mathbf{b}, \Theta, \tilde{\Theta}, \Psi\}$ represent all model parameters. Assuming independent contexts, we get

$$\begin{aligned} p(\mathbf{S}|\Lambda) &= \prod_{(q,\mathbf{a},\mathbf{c}) \in \mathbf{S}} p(\mathbf{c}|\Lambda) = \prod_{(q,\mathbf{a},\mathbf{c}) \in \mathbf{S}} \prod_i p(c_i|\Lambda) \\ &= \prod_{(q,\mathbf{a},\mathbf{c}) \in \mathbf{S}} \prod_i [(\alpha_{i,(q,\mathbf{a})} \cdot \beta_{iq|\mathbf{a}|})^{c_i} \cdot (1 - \alpha_{i,(q,\mathbf{a})} \cdot \beta_{iq|\mathbf{a}|})^{1-c_i}] \end{aligned}$$

With the assumption of independent α , we also have $p(\alpha|\mu, \mathbf{b}, \Theta, \tilde{\Theta}, \Psi) = \prod_{i,(q,\mathbf{a})} p(\alpha_{i,(q,\mathbf{a})}|\mu, \mathbf{b}, \Theta, \tilde{\Theta}, \Psi)$. As we describe above, we placed zero mean Gaussian prior on the parameter \mathbf{b}, Ψ, Θ and $\tilde{\Theta}$ (L_2 regularization). We employ EM algorithms to solve the model, treating examination E as a latent variable.²

5. MAINLINE AD EXPERIMENTS

We use the data of Section 3 as training data, and further capture an additional two weeks of traffic as our test data. The data statistics are shown in Table 4.

Since the train/test data are split temporally, this evaluation is more instructive for real applications. The major measurement for performance evaluation is log likelihood, which is widely used in the CTR prediction problem [11]. Additionally, we also report Relative Information Gain (RIG) [18]. In our experiments, the dimension of latent factors for queries and ads is set at 10.

5.1 Analysis of Click Model for Mainline Ads

We conduct experiments to analyze the performance of the variations of our context aware CTR prediction model. First, we check the examination hypotheses by setting the relevance to Non-Informational Relevance (NIR). The results are shown in Table 5. We can see that Depth-Dependent Examination (DDE) outperforms the Constant Examination (CE) as we expected. However, we also notice that the performance of the query-dependent model QDDE is slightly worse than DDE. After examining the relevant data, we find that more than half of the queries do not hold complete depth records, and so, for these queries, some specific depths data are not applicable in the training data while they may appear in test

²We use SGD in M-step, so the learning algorithm is very efficient and the training time is linear in data size. Materials showing how to solve the model are available upon request.

Table 5: Model Analysis

Examin.	Relev.	LL	RIG
CE	NIR	-0.4206	0.0250
DDE	NIR	-0.4081	0.0539
QDDE	NIR	-0.4123	0.0442
DDE	LBM	-0.3789	0.1215
DDE	LBM+FM	-0.3783	0.1229
DDE	CM	-0.3543	0.1787
CE	CM	-0.3633	0.1578
DDE	CM	-0.3543	0.1787
QDDE	CM	-0.3712	0.1395

data. In these cases, QDDE cannot make a better prediction and a more advanced model that can effectively handle query dependent parameters is necessary.

By using Depth Dependent Examination as our examination model, we can try different relevance models. From Table 5, we can see that if we only use the Latent Bias Model (LBM), we get better results than NIR. After adding features to the model (LBM+FM), the performance gets slightly better in both tasks too, but the improvement is negligible. By using more delicate features such as Click Over Expected Clicks (COEC) [37] or latent topic distribution of the ad, a larger improvement may be possible. Next, we use Combined Model, which combines the three models: LBM, FM, and IM. As expected, the model achieve the best performance. The improvement over LBM is much larger than FM. We also test different examination models with the combined relevance model, and the experiments show similar results, the Depth Dependent Model is always better than CE and QDDE. We also check convergence of the EM algorithms and it converges quickly, achieving a saddle point within 100 iterations. Next, we will employ the combined relevance model (CM) and Depth Dependent Examination model (DDE) to compare with other baselines.

5.2 Comparison with Existing Methods

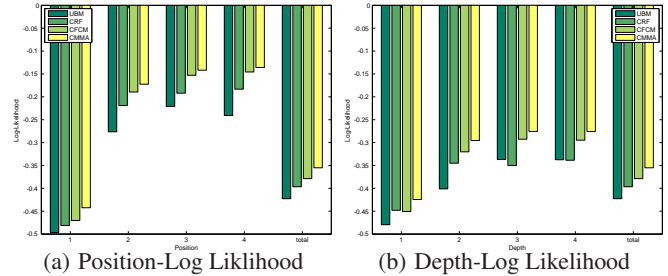
We name our method CMMA for Click Model for Mainline Ads and compare our model with three existing methods on both Description Oriented Evaluation and Prediction Oriented Evaluation: **User Browsing Model (UBM)** is the baseline method and a classical click model [17]. We adopt EM inference process for parameter α and β . To avoid infinite values in log-likelihood, α is truncated into the range (0.0001, 0.9999). **Matrix Factorization Click Model (MFCM)** is proposed by Shen et al. [31].³ The dimensionality of the latent factor vector is also set at 10 for a fair comparison. **Relational click model (CRF)** is proposed by Xiong et al. [33]. They adopt Conditional Random Field (CRF) to model the exclusive effects between similar ads that are shown together in the same impression. The reason we choose these three methods is that UBM is a classical click model while MFCM and CRF are very recent state-of-the-art methods.

Table 6: Comparing with the baselines

	UBM	CRF	MFCM	CMMA
Log Likelihood	-0.4227	-0.39654	-0.38464	-0.3543
RIG	0.0201	0.08086	0.10844	0.1787

The overall results are shown in Table 6. We see that CRF achieves much better results than UBM. As discussed, UBM cannot handle new query, ads or even the new pair query-ads with the existing query and ad. Thus, CRF which is designed for prediction

³In this paper, we do not focus on the personalized model. Personalization could be a factor of the ads contexts, and easily plugged into our framework.


Figure 4: Comparing CMMA with baselines

tasks, outperforms UBM. For MFCM, we also consistently find that MFCM also outperforms UBM significantly as shown in [31]. Our method outperforms all three baselines. Similarly, we also show the results on positions in Figs. 4(a) and the results on depths in Figs. 4(b). For the cases of Position 1 and depth 1, the performance of all four methods is worse than other cases. Because the number of the cases of Position 1 and depth 1 are always much larger than other cases, they have a greater variance on CTR which increases the difficulty of predicting CTR. In all cases, our methods outperform all three other methods. The improvement on RMSE over MFCM is more than 10%.

6. MACRO FACTORS ANALYSIS

In previous section, we have already seen the effects of micro factors, including ad depth and interactive influence between query and ads and among ads. However, whether there are correlations or co-effects between ads and webpages is still not clear. In this section, our major objective is to leverage the relationship between ads and organic results to obtain better performing CTR prediction for ads. To study the relationship between their CTRs, we first introduce two factors. One is *Externality*, which means how competitive the sponsored search and organic search are. For example, when an ad is shown with a more attractive webpage, the attention given to the ad will be distracted [33]. The other is *Similarity*, which means how similar the sponsored search and organic search are in content. When a set of ads and webpages are too similar in content, the attention given to each individual ad/webpage may all be decreased or increased.

Since we target the relationship between mainline ads and organic results, we only consider search results displaying mainline ads. We examine data from a commercial search engine containing about 65 million search result page views (SERPs), about 200,000 unique queries, and 200,000 mainline ads. Each SERP can be represented by a triple consisting of (query, ad, organic results), where organic results is a ranked list of organic search results.

6.1 Overall Externality between Mainline Ads and Organic Results

To study the externality between mainline ads and organic results, we first define a *SERP group* as SERPs with the following criteria. 1. They contain the same ad in the same position. 2. They contain the same query. 3. They contain the same organic results. Note that different SERP groups may have overlapping SERPs because one SERP may contain more than one mainline ad. We denote a mainline ad and the organic results in group g as $A_{q,g}$ and $O_{q,g}$, respectively, where q represents the query. To make the statistics meaningful, we only consider the groups shown in at least 15 SERPs. We define the CTR of ad $A_{q,g}$ (denoted as $CTR_{A_{q,g}}$)

as the ratio of total clicks for the ad to the number SERPs in $A_{q,g}$: $CTR_{A_{q,g}} = \frac{\#Click_{A_{q,g}}}{|A_{q,g}|}$. The way to compute the CTR for organic results, denoted as $CTR_{O_{q,g}}$, of set $O_{q,g}$, is similar. After obtaining the CTRs, we can analyze the relationship between the CTR of mainline ads and organic results. To show the externality visually, we plot each SERP group as a point with coordinates of the two CTRs and get a density graph as shown in Figure 5(a).

Figure 5(a) shows that organic results' CTR does correlate to mainline ads' CTR. The increase of organic results' CTR accompanies the decrease of mainline ads' CTR, which shows these two CTRs are negatively correlated. Note that this observation is slightly different from [13], which can be attributed to several reasons. First, the setting of their data analysis is different from ours. They consider the relationship between a single ad and a single organic result while we consider the relationship to the complete organic result set. Second, they only show the average mainline ads' CTR for fixed organic results' CTR while we display the density graph, which shows more detailed information.

6.2 Externality between Mainline Ads and Organic Results on Query Level

The above analysis considers the externality with respect to overall CTRs. To better understand externality, we would like to consider externality on query level, which could help to remove the influence of relevance. We define a query group as a set of SERPs groups with the same ad and query. Within a query group, the relevance is the same and the relationship between the CTR of the ad and organic results can be mainly attributed to externality. For the statistical significance of our analysis, query groups with less than 10 SERP groups are discarded. After the preprocessing, we had 7,961 different query groups that covered about 21 million SERPs (including possible overlaps) for our externality study.

For each query group, we plot the points with coordinate $(CTR_{A_{q,g}}, CTR_{O_{q,g}})$ for each SERP group. To see the correlation between ad clicks and organic result clicks, we use linear function $y = \alpha + \beta x$ to fit the points. The slope β expresses the sign and strength of the externality. To measure the fitness of the linear model, we further estimate the coefficient of determination R^2 , which is the proportion of variability of data which can be explained by the linear regression function. The R^2 value of two random variables X and Y for the linear function model is defined by $R_{XY}^2 = \frac{(\sum_i (x_i - \bar{x})(y_i - \bar{y}))^2}{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}$, where x_i is the i^{th} sample of X and \bar{x} is the mean of the samples for X , similarly for y_i and \bar{y} .

We then plot the distributions of the slope and R value for different query groups⁴ in Figure 5(b) and Figure 5(c). There are a significant number of queries that have nonzero correlations ($|\beta| > 0$) between ad CTR and organic CTR, which is strong evidence that we should not treat them as independent and model them separately. Specifically, a major portion of the correlation coefficients fall into the interval $[-5, 0]$, indicating that a negative correlation dominates the relationship between ad clicks and organic clicks. This is also consistent with our previous analysis from a global perspective. At the same time, we can still observe a significant number of query groups (29.4%) that have positive correlations. The standard deviation of β is 1.1282, showing that the slope varies widely between different query groups. The R value can be seen as a confidence value of the linear dependency. We consider query groups with $R \in [0.5, 1]$ as high confidence groups. We find that

⁴We plot R instead of R^2 value because the approximately linear relationship of R is easier to observe.

21.8% query groups have high confidence and 9.8% of these high confidence query groups have positive correlation.

To better understand the results, we also conducted some case studies. We randomly drew two query groups with positive correlation and two query groups with negative correlation from the high confidence groups. We present the raw data and the fitting line, as well as the query of the query group in Figure (6). The two negative query groups both have navigational queries and the two positive query groups both are commercial but non-navigational queries. A non-navigational query means the user usually does not have strong intent when searching the query. This result is also consistent with [13]. However, they divided the query as navigational and non-navigational without considering the externality at the query level. We can see that the degree of correlations, measured by the slope of the line, is different for different queries.

In conclusion, the overall correlation between ads and organic results is negative. Yet, the correlation varies according to the query and the positive correlation cannot be neglected.

6.3 Similarity Analysis

In the previous section we showed the correlation between the CTR of ads and organic search results. In this section we study how similarity of content between ads and organic results influences their CTRs. We focus on the most clicked ads (denoted as $a_{q,g}^*$) in each set $A_{q,g}$ for query q , which is assumed to be the most popular ads in this group. We then find the most clicked organic results in the same group (denoted as $o_{q,g}^*$). If $a_{q,g}^*$'s URL and $o_{q,g}^*$'s URL share the same URL domain, we assume their content is similar to some extent and regard such a pair $(a_{q,g}^*, o_{q,g}^*)$ as a similar pair. For each similar pair we find the control group where $o_{q,g}^*$ does not appear and the experimental group where they both appear, by which we exclude the influence of other factors to the greatest extent. Then we estimate $\Delta CTR_{q,g}$, defined as, $\Delta CTR_{q,g} = CTR_{o_{q,g}^*, q, g} - CTR_{o_{q,g}^*, q, j} | a_{q,g}^*$, which means the difference in CTR on $o_{q,g}^*$ caused by the existence and absence of $a_{q,g}^*$.⁵ We calculate the average difference for all similar pairs in a query and plot the distribution of the average value for each query in Figure 5(d). The resulting values are mainly positive, which fits the intuition that similar ads to organic results will distract attention from organic results. This phenomenon may be explained in that users have limited attention for one SERP, which makes it necessary to choose between ads and organic results based on their content. When the mainline-ads have similar content, it is likely that some clicks are distracted from organic results. This explanation implies that we should take the similarity of ads and organic results into consideration in CTR prediction. It further motivates us to prefer diversity when displaying ads and organic results.

7. CROSS-DOMAIN CLICK MODEL

We have already seen the correlations between organic search and sponsored search. To further verify the effects of correlations, we build a cross-domain model based on the results of the data analysis. Major parameters for this model are listed in Table 7. Note that for description convenience, we redefine some notations used in this sections. For the ad and organic search domains, we obtain two matrix factorization (MF) models as,

$$\begin{aligned} r_{qa} &\sim \theta_q^{(a)} \cdot \psi_a + \mathbf{b}^{(a)} \cdot \mathbf{x}_{qa} + \epsilon_a, & \epsilon_a &\sim \mathcal{N}(0, (\sigma^{(a)})^2) \\ r_{qo} &\sim \theta_q^{(o)} \cdot \psi_o + \mathbf{b}^{(o)} \cdot \mathbf{x}_{qo} + \epsilon_o, & \epsilon_o &\sim \mathcal{N}(0, (\sigma^{(o)})^2) \end{aligned} \quad (1)$$

⁵It would be better to have the setting the other way around. However, it is hard to do so offline and the current setting is meaningful in that it helps reveal the relationship between ads and organic results.

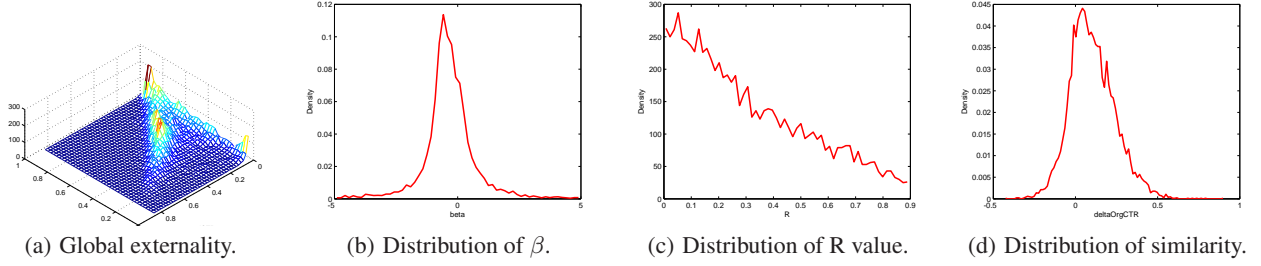


Figure 5: Experimental results

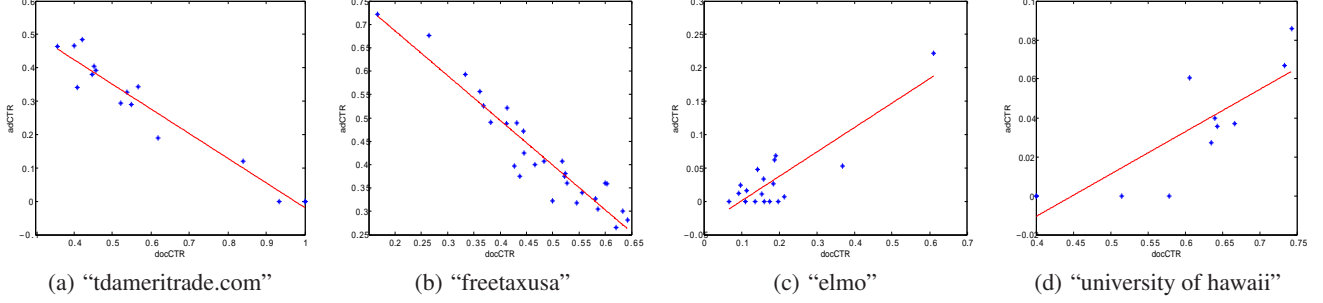


Figure 6: Case studies.

Table 7: Notations used in model

Symbol	Description
K	The dimension of latent feature vector
Ω^q	The correlation between $\theta_q^{(a)}$ and $\theta_q^{(o)}$
$\theta_q^{(a)}$	Latent feature vector representing query q in ad domain
$\theta_q^{(o)}$	Latent feature vector representing query q in organic domain
ψ_{a_i}	Latent feature vector representing ad a_i
ψ_{o_j}	Latent feature vector representing organic result o_j
$\mathbf{b}^{(a)}$	Weights for explicit features in ad domain
$\mathbf{b}^{(o)}$	Weights for explicit features in organic domain
r_{qa}	User click/non-click response for ad a given query q
r_{qo}	User click/non-click response for organic result o given query q

where ψ_a and ψ_o are the latent representations of ad and organic results. $\theta_q^{(a)}$ and $\theta_q^{(o)}$ are the latent representations of the query in the ad and the organic domain. When we constrain the length of the vector to be 1 (then it degenerates to scalar) and fix the ψ_a and ψ_o as the element vector, model (1) becomes a multi-task linear regression model.

The factor for query q appears in both MF models. According to our previous analysis, they should be correlated. To model the relationship between ads and organic search, we put a matrix-variate normal distribution [21] on the $\theta_q^{(a)}$ and $\theta_q^{(o)}$, which represent the latent factor of a query for ads and organic results, respectively. That means we use Ω^q to characterize query-specific relationship between two domains based on the latent factors. Formally:

$$\mathbf{U}_q \sim \mathcal{MN}_{K \times 2}(\mathbf{0}_{K \times 2}, \mathbf{I}_K \otimes \Omega_q) \quad (2)$$

where \mathbf{U}_q is the $K \times 2$ matrix representation for query q and $\mathbf{U}_q = [\theta_q^{(a)}, \theta_q^{(o)}]$.

In Eq (2), $\mathcal{MN}_{a \times b}(\mathbf{M}, \mathbf{A} \otimes \mathbf{B})$ denotes a matrix-variate normal distribution with mean $\mathbf{M} \in \mathbb{R}^{a \times b}$, row covariance matrix $\mathbf{A} \in \mathbb{R}^{a \times a}$ and column covariance matrix $\mathbf{B} \in \mathbb{R}^{b \times b}$.

The probabilistic density function of matrix-variate normal distribution is

$$\mathcal{MN}_{a \times b}(\mathbf{M}, \mathbf{A} \otimes \mathbf{B}) = \frac{\exp(-\frac{1}{2} \text{tr}(\mathbf{A}^{-1}(\mathbf{X} - \mathbf{M})\mathbf{B}^{-1}(\mathbf{X} - \mathbf{M})^T))}{(2\pi)^{ab/2} |\mathbf{A}|^{b/2} |\mathbf{B}|^{a/2}} \quad (3)$$

where $\text{tr}(\cdot)$ and $|\cdot|$ denote the trace and determinant, respectively, of a square matrix. Since \mathbf{B} represents the covariance for columns, Ω^q is the covariance for query q 's latent factors in two different domains. That is, Ω^q is the parameter to describe the relationship between ads and organic search. Due to the scale of our problem, we use stochastic gradient descent (SGD) to solve the problem, which is an efficient algorithm for matrix factorization.

8. CROSS-DOMAIN EXPERIMENTS

In this section, we will present our experimental results based on our large real-world dataset. We try to answer two questions. Firstly, can the click information of organic results benefit the prediction of ad CTR? Secondly, can similarity information improve the performance of cross-domain CTR prediction?

The dataset we used in the following experiments is the same as the one introduced in Section 6. We do not use queries less than 30 SERPs because queries with few SERPs are not enough to build a good model for clicks of organic results, not to mention to transfer the knowledge for ad CTR prediction. For the purpose of this evaluation, we split the dataset into a training set, of which 85% is the data and the remaining 15% is the test set. The data was separated according to time information to avoid using future data to predict the past. We extract 19 features such as query bias, ad bias, ad URL, ad title, ad position, search result position, search result URL, time, as well as user bias and location and browser as the feature set.

We further group the queries in the test set according to the query frequency. The detailed information with the distribution of query frequency is listed in Table 8. The top query set, which consists of queries appearing more than 30,000 times in the training data, is discarded. This is because the top queries are relatively easy and the classical linear model [18] could already handle them well.

8.1 Methods

Here we compare our approach with two baseline methods. The first method is AdOnly, the classical linear regression model does not use the information of organic results. The second method is AdSim, the same algorithm as AdOnly, but some features used

Table 8: Details of Experiment Data

Interval	Frequency	#Query	#SERP	#Entry
Freq 1	$10^{1.5} \sim 10^2$	94,907	7,718,620	13,168,273
Freq 2	$10^2 \sim 10^{2.5}$	48,547	10,376,698	19,184,516
Freq 3	$10^{2.5} \sim 10^3$	16,381	10,936,742	20,062,092
Freq 3	$10^3 \sim 10^{3.5}$	4,220	8,934,949	15,686,684
Freq 3	$10^{3.5} \sim 10^4$	1,307	8,644,470	13,516,733
Freq 3	$10^4 \sim 10^{4.5}$	309	6,340,607	9,031,495

Table 9: Overall Performance of Different Models under Log Likelihood and Relative Information Gain

Model	AdOnly	AdSim	AdOrg	AdOrgSim	AdOrgSimMF
LL	-0.421	-0.418	-0.411	-0.401	-0.391
RIG	-0.330	-0.322	-0.308	-0.279	-0.236

in [13] to describe the relationship between the ad and organic result, such as the edit distance between the ad’s URL domain and the organic result’s URL domain, are added as features. There could be more sophisticated ad CTR prediction baselines that use more information such as [27]. However, we did not include them because of two reasons. Firstly, due to the speciality of the problem, there is no benchmark dataset available and it is not possible to obtain the same data or even same feature set for fair comparisons. Secondly, our focus in this paper is to show the effect of combining organic search and sponsored search, rather than integrating all possible data sources to reach the best performance. Therefore, we only compare to the above two baselines.

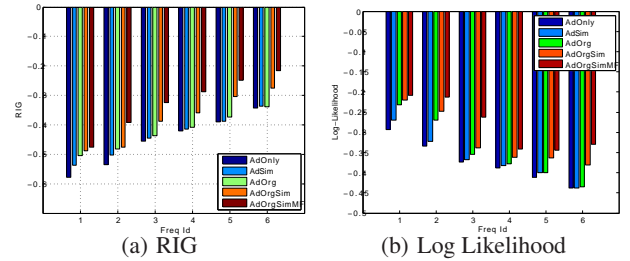
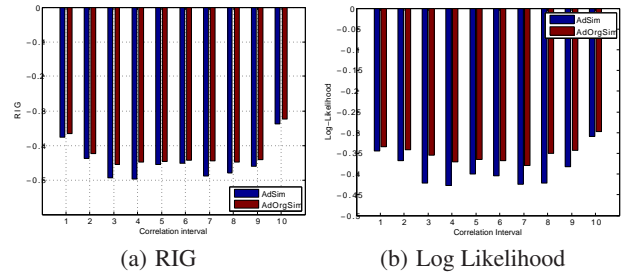
Our proposed models also have three versions. The first version is AdOrg, our multi-task linear model that introduces correlated query bias. This model does not use similarity information used by AdSim. The second version is AdOrgSim, our multi-task linear model that considers the influence of similarity in the same way as in AdSim. The features used are exactly the same as AdSim. The third version is AdOrgSimMF, our multi-task matrix factorization models in section 7 that also consider similarity by utilizing features in AdSim. AdSim is designed to show that the introduction of the similarity features can improve the performance of ad CTR prediction. The AdOrg is expected to show that the multi-task linear regression model is a better way to model the interactions between ads and organic results. The AdOrgSim model is expected to further improvements over the AdOrg model by incorporating similarity. The AdOrgSimMF is expected to improve on the AdOrgSim model by exploiting the correlation between domains at the latent factor level. Like in Section 5, we also set $K=10$ for the AdOrgSimMF model below.

8.2 Results

We evaluated the performance based on queries of all 6 frequency intervals and on queries of different correlation coefficient intervals, which we learned in the multi-task linear model.

8.2.1 Overall performance

The overall performances of the five models are shown in Table 9. Note that we do not report the variance of the experiment results. This is because our dataset is large enough to be highly confident in the results even with one run. Based on the results, we can make the following observations: 1) The AdSim model, which considers the organic information with the similarity between ads and organic results as side information, outperforms AdOnly. This result is consistent with our data analysis that similarity is an important factor influencing the CTR. 2) By comparing AdOrg and


Figure 7: Experiment result for queries with different frequency

Figure 8: Experiment result for different value of ρ_q

AdOnly, which use the same set of explicit features in the ad domain, we find that AdOrg significantly outperforms AdOnly. This observation is consistent with our analysis that AdOrg can predict the ad CTR better than AdSim by exploiting the relationship between ads and organic results. 3) The comparison between AdOrg and AdSim is more interesting. We find that the AdOrg model outperforms the AdSim model, though it uses fewer features than AdSim. This result indicates that only considering organic results as features is not enough. Instead, we should model the organic results in more principled method, as AdOrg achieves better performance through joint modeling. 4) We find that AdOrgSim has better performance than AdOrg and achieves significant improvements over the state-of-the-art AdSim model, which uses the information from organic results as side information. 5) Finally, we find the AdOrgSimMF model achieves the best performance and improves AdOrgSim, which shows that the factorization term in the model can better exploit the relationship among ads and organic results.

8.2.2 Queries with different frequency

To study the influence of query frequency on performance, we present the detailed performance of the five models with different query frequency intervals in Figure 7. The results show that the improvements of AdOrg and AdOrgSim over AdOnly and AdSim are more significant for less frequent queries. On “tail queries,” which appear only a few times (Freq 1~ Freq 3), AdOrg and AdOrgSim achieve the largest improvements over the baselines. The observations support our hypothesis that by exploiting the interaction between ad and organic results via a multi-task linear model, the proposed model alleviates the sparsity problem in ad CTR prediction. By carefully comparing the results of AdSim and AdOrg, we find the improvement is most significant for query intervals of Freq 2 and Freq 3. Queries of Freq 1 may give insufficient information to learn the relationship between ad and organic results. In conclusion, modeling the correlations between ad and organic search is more effective on tail queries and slightly more information gives us greater improvement.

8.2.3 Performance over correlation coefficients

In this section, we check the query-dependent correlations, which are modeled as the correlation parameter ρ_q for each query in our AdOrgSim model, and its influence on the prediction of CTR.

ρ_q can represent the correlation between ads and organic results and is guaranteed to be within the range $[-1, 1]$. A positive value means a positive correlation between ads and organic results and vice versa. The larger the absolute value is, the stronger the correlation is. We divide $[-1, 1]$ to 10 intervals with equal length, and compute the RIG and log likelihood for queries in each interval for model AdSim and AdOrgSim. The results are shown in Figure 8. As we can see, in the middle of Figure 8, which are the results for interval $[-0.2, 0]$ and $[0, 0.2]$, the average improvement is the smallest. However, when the absolute value of ρ_q becomes larger, the improvement is much more significant. The results can be explained by the intuition that, when the correlation is weak, the advantage of jointly modeling the ad and organic result is small. When the correlation is stronger, the improvement becomes more significant. Figure 8 also proves that we can learn the correlations via the multi-task model.

9. CONCLUSION

We have studied the contextual factors of click-through rate for sponsored search in two aspects: for micro factors, we analyze the factors for mainline ads, such as ad depth, query bias and interactive influence between ads displayed together. For macro factors, we explored the correlation factor between sponsored search and organic search in terms of externality and similarity. We have incorporated these factors into click models and found that they play important roles in understanding and predicting user clicks on ads. Our extensive experiments based on large-scale real-world data sets show that our two proposed models (based on micro and macro factors) can outperform state-of-the-art methods.

However, we also notice that although our models verify the effectiveness of these contextual factors and provide some hints for future research in sponsored search, in this paper, we did not incorporate all factors together into a unified model, due to the complexity of modeling. We leave that as future work.

10. REFERENCES

- [1] D. Agarwal, A. Z. Broder, D. Chakrabarti, D. Diklic, V. Josifovski, and M. Sayyadian. Estimating rates of rare events at multiple resolutions. In *KDD*, 2007.
- [2] A. Ashkan and C. L. Clarke. Characterizing commercial intent. In *CIKM*, 2009.
- [3] A. Ashkan and C. L. A. Clarke. Modeling browsing behavior for click analysis in sponsored search. In *CIKM*, 2012.
- [4] R. Bell, Y. Koren, and C. Volinsky. The BellKor solution to the Netflix Prize A factorization model. *KorBell Teams Report to Netflix*, 2007.
- [5] G. Bouchard, D. Yin, and S. Guo. Convex collective matrix factorization. In *AISTATS*, 2013.
- [6] R. Briggs and N. Hollis. Advertising on the web: Is there response before click-through? In *Journal of Advertising Research*, 1997.
- [7] B. Cao, D. Shen, K. Wang, and Q. Yang. Clickthrough log analysis by collaborative ranking. In *AAAI*, 2010.
- [8] O. Chapelle and Y. Zhang. A dynamic bayesian network click model for web search ranking. In *WWW*, 2009.
- [9] W. Chen, Z. Ji, S. Shen, and Q. Yang. A whole page click model to better interpret search engine click data. In *AAAI*, 2011.
- [10] Y. Chen and T. W. Yan. Position-normalized click prediction in search advertising. In *KDD*, 2012.
- [11] H. Cheng and E. Cantú-Paz. Personalized click prediction in sponsored search. In *WSDM*, 2010.
- [12] N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey. An experimental comparison of click position-bias models. In *WSDM*, 2008.
- [13] C. Danescu-Niculescu-Mizil, A. Z. Broder, E. Gabrilovich, V. Josifovski, and B. Pang. Competing for users' attention: on the interplay between organic and sponsored search results. In *WWW*, 2010.
- [14] K. S. Dave and V. Varma. Learning the click-through rate for rare/new ads from similar ads. In *SIGIR*, 2010.
- [15] K. Dembczynski, W. Kotlowski, and D. Weiss. Predicting ads' click-through rate with decision rules. *Workshop on Targeting and Ranking in Online Advertising*, 2008.
- [16] G. Dror, N. Koenigstein, and Y. Koren. The Yahoo! Music Dataset and KDD-Cup'11. *KDD-Cup*, 2011.
- [17] G. E. Dupret and B. Piwowarski. A user browsing model to predict search engine click data from past observations. In *SIGIR*, 2008.
- [18] T. Graepel, J. Q. Candela, T. Borchert, and R. Herbrich. Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft's bing search engine. In *ICML*, 2010.
- [19] F. Guo, C. Liu, A. Kannan, T. Minka, M. Taylor, Y.-M. Wang, and C. Faloutsos. Click chain model in web search. In *WWW*, 2009.
- [20] F. Guo, C. Liu, and Y. M. Wang. Efficient multiple-click models in web search. In *WSDM*, 2009.
- [21] A. K. Gupta. *Matrix Variate Distributions*. Chapman & Hall/CRC, Oct. 1999.
- [22] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay. Accurately interpreting click through data as implicit feedback. In *SIGIR*, 2005.
- [23] Y. Koren. Factor in the neighbors: Scalable and accurate collaborative filtering. *ACM TKDD*, 4(1):1–24, January 2010.
- [24] Y. Koren and R. Bell. Matrix factorization techniques for recommender systems. *IEEE Computer*, 42(8), 2009.
- [25] C. Liu, F. Guo, and C. Faloutsos. Bbm: bayesian browsing model from petabyte-scale data. In *KDD*, 2009.
- [26] C. Liu, F. Guo, and C. Faloutsos. Bayesian browsing model: Exact inference of document relevance from petabyte-scale data. *ACM Trans. Knowl. Discov. Data*, 4(4):19:1–19:26, Oct. 2010.
- [27] A. K. Menon, K.-P. Chitrapura, S. Garg, D. Agarwal, and N. Kota. Response prediction using collaborative filtering with hierarchies and side-information. In *KDD*, 2011.
- [28] T. Minka, J. Winn, J. Guiver, and A. Kannan. A click through model - sample code. <http://research.microsoft.com/en-us/um/cambridge/projects/infernet/docs/Click%20through%20model%20sample.aspx>, 2009.
- [29] M. Regelson and D. Fain. Predicting click-through rate using keyword clusters. In *Proceedings of the Second Workshop on Sponsored Search Auctions*, volume 9623, 2006.
- [30] M. Richardson, E. Dominowska, and R. Ragno. Predicting clicks: estimating the click-through rate for new ads. In *WWW*, 2007.
- [31] S. Shen, B. Hu, W. Chen, and Q. Yang. Personalized click model through collaborative filtering. In *WSDM*, 2012.
- [32] R. Srikant, S. Basu, N. Wang, and D. Pregibon. User browsing models: relevance versus examination. In *KDD*, 2010.
- [33] C. Xiong, T. Wang, W. Ding, Y. Shen, and T.-Y. Liu. Relational click prediction for sponsored search. In *Proceedings of WSDM*, 2012.
- [34] W. Xu, E. Manavoglu, and E. Cantu-Paz. Temporal click model for sponsored search. In *SIGIR*, 2010.
- [35] S. Yang and A. Ghose. Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence? *Marketing Science*, 30(1), Apr. 2010.
- [36] D. Yin, S. Guo, B. Chidlovskii, B. D. Davison, C. Archambeau, and G. Bouchard. Connecting comments and tags: Improved modeling of social tagging systems. In *WSDM*, 2013.
- [37] W. V. Zhang and R. Jones. Comparing click logs and editorial labels for training query rewriting. In *Query Log Analysis: Social And Technological Challenges. A workshop at WWW 2007*.
- [38] Y. Zhang, W. Chen, D. Wang, and Q. Yang. User-click modeling for understanding and predicting search-behavior. In *KDD*, 2011.
- [39] Z. A. Zhu, W. Chen, T. Minka, C. Zhu, and Z. Chen. A novel click model and its applications to online advertising. In *WSDM*, 2010.