

# Do Ads Compete or Collaborate? Designing Click Models with Full Relationship Incorporated

Xin Xin  
School of Computer Science  
Beijing Institute of Technology  
xxin@bit.edu.cn

Irwin King  
The Chinese University of  
Hong Kong  
king@cse.cuhk.edu.hk

Ritesh Agrawal  
AT&T Labs–Research  
ragrawal@attinteractive.com

Michael R. Lyu  
The Chinese University of  
Hong Kong  
lyu@cse.cuhk.edu.hk

Heyan Huang  
School of Computer Science  
Beijing Institute of Technology  
hhy633@bit.edu.cn

## ABSTRACT

Traditionally click models predict click-through rate (CTR) of an advertisement (ad) independent of other ads'. Recent researches however indicate that the CTR of an ad is dependent on the quality of the ad itself but also of the neighboring ads. Using historical click-through data of a commercially available ad server, we identify two types (competing and collaborating) of influences among sponsored ads and further propose a novel click-model, Full Relation Model (FRM), which explicitly models dependencies between ads. On a test data, FRM shows significant improvement in CTR prediction as compared to earlier click models.

## Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval

## Keywords

Sponsored Search, Click Models, Collaborating and Competing Influence

## 1. INTRODUCTION

**Influence among ads.** The target of click models [1, 2, 4, 5, 6, 7, 8, 9, 11, 12, 13] is to understand the click-through data by analyzing users' behaviors, in order to get better CTR prediction. Originally, most click models treat each ad independently, while recent work [10] demonstrates that the ads are correlated among each other, and incorporating competing influence within ads into click models would obtain better performance [11]. As shown in Figure 1 (left), consider to estimate the CTR of the first ad, denoted as  $ad_1$ , under three different cases. The quality of the second ad is denoted by an integer from 1 to 5 (1 is the lowest quality

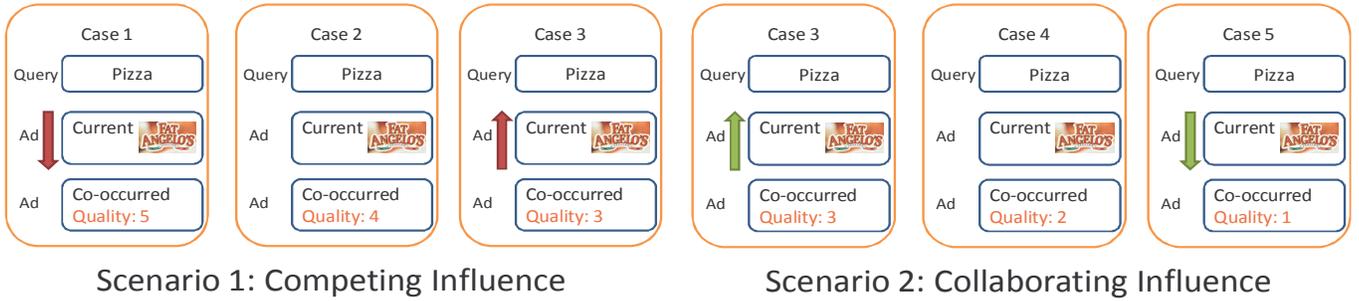
and 5 is the highest). In original click models, the influence among the ads would not be considered. Consequently, the estimation of  $ad_1$ 's CTR should be the same in all the three cases. Nevertheless, if the competing influence among ads is considered, the CTR of  $ad_1$  in Case 1 would be smaller than that in Case 2; and the CTR of  $ad_1$  in Case 3 would be larger than that in Case 2. Because following the competing hypothesis, when an ad is displayed with a high-quality ad, its CTR will be decreased; and when it is displayed with a low-quality ad, its CTR will be increased. The intuitive meaning is that the co-occurred high-quality ad will attract some of the users' attention from  $ad_1$ .

**Limitation of previous work.** Although previous work has identified the effect of competing influence, through further investigation in real-world data, we identify a novel observation that would make click models with only competing influence fail to work. The observation is that the competing influence does not always take effect; and under some cases, its opposite, the collaborating influence takes more effect. Consider Scenario 2 in Figure 1. In this scenario, the qualities of the second ad are all below average. If we follow the competing influence, the CTR of  $ad_1$  in Case 3 will be smaller than the CTR of  $ad_1$  in Case 4. However, from statistics in real-world dataset, the opposite observation is true. We have found the CTR in Case 3 will be larger than that in Case 4 in most cases. This is a kind of collaborating influence, which means when an ad occurs with a high-quality ad, its CTR will be increased and when an ad occurs with a low-quality ad, its CTR will be decreased. An intuitive psychology interpretation could be found from the authority fear [3] phenomenon. It means that users treat the search engines as experts, and do not like to challenge it. However, if an ad is co-occurred with a very low-quality ad, the trust would be destroyed. This will reduce the click chance of the first ad. By further demonstration from data analysis, we will show when the qualities of the co-occurred ads are above average, the competing influence takes more effect; however, when the qualities of the co-occurred ads are below average, the collaborating influence takes more effect. This observation indicates that only incorporating the competing influence will fail to work under the second condition.

**Motivation and Contributions.** To solve the above limitation, it is natural to incorporate both competing and collaborating influences into the click model. This moti-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CIKM'12, October 29–November 2, 2012, Maui, HI, USA.  
Copyright 2012 ACM 978-1-4503-1156-4/12/10 ...\$15.00.



**Figure 1: An example of influence given the same query and different ads in 5 cases. The red arrow denotes the competing influence and the green arrow denotes the collaborating influence.**

vates our work in this paper. The main contributions of this paper lies in: 1) we identify the competing and collaborating influence among ads; and 2) we propose a novel click model, which considers both collaborating and completing influences, and the model outperforms previous ones significantly.

## 2. COMPETING AND COLLABORATING IDENTIFICATION

We will define and identify the competing and collaborating influence in this section. The dataset information would be presented in the later experiment part. The relation between two ads would be discussed first for clearness. The generalization would be designed in the algorithms.

### 2.1 Definitions

As shown in Figure 1, suppose the average CTR of  $ad_1$  in all the cases is  $CTR_{avg}(ad_1)$ , and the CTR of  $ad_1$  when impressed with a special ad  $ad_2$  is  $CTR_{ad_2}(ad_1)$ .

**DEFINITION 1 (NEGATIVE EFFECT).**  $ad_1$  has obtained a **negative effect** from  $ad_2$  if and only if

$$CTR_{ad_2}(ad_1) - CTR_{avg}(ad_1) < 0;$$

**DEFINITION 2 (POSITIVE EFFECT).**  $ad_1$  has obtained a **positive effect** from  $ad_2$  if and only if

$$CTR_{ad_2}(ad_1) - CTR_{avg}(ad_1) > 0.$$

One question is what will result in the negative and positive effects? Do ads compete with each other? Do ads collaborate with each other? Or do competing and collaborating relations occur together but with different strength in different cases? This is the fundamental problem to be discussed in this session.

**DEFINITION 3 (QUALITY).** The **quality** of an ad to a certain query is the probability of clicking the ad on the condition of having examined it, denoted as  $p(C = 1|E = 1)$

This definition is utilized to remove the position-bias in the relationship between a query and an ad. Under this definition, the competing and collaborating influences are defined as follows.

**DEFINITION 4 (COMPETING).**  $ad_2$  would have an **competing influence** to  $ad_1$  if and only if

$$\begin{aligned} CTR_{(quality(ad_2)+\varepsilon)}(ad_1) - CTR_{ad_2}(ad_1) < 0 \\ CTR_{(quality(ad_2)-\varepsilon)}(ad_1) - CTR_{ad_2}(ad_1) > 0. \end{aligned}$$

**DEFINITION 5 (COLLABORATING).**  $ad_2$  would have an **collaborating influence** to  $ad_1$  if and only if

$$\begin{aligned} CTR_{(quality(ad_2)+\varepsilon)}(ad_1) - CTR_{ad_2}(ad_1) > 0 \\ CTR_{(quality(ad_2)-\varepsilon)}(ad_1) - CTR_{ad_2}(ad_1) < 0. \end{aligned}$$

$\varepsilon$  is a small positive value, thus  $CTR_{(quality(ad_2)+\varepsilon)}(ad_1)$  means what the CTR of  $ad_1$  would be if  $ad_2$  is replaced with a little higher-quality ad. Therefore, we could see that if  $ad_2$  has a competing influence to  $ad_1$ , the higher quality  $ad_2$  is, the lower CTR  $ad_1$  would get; and if  $ad_2$  has a collaborating influence to  $ad_1$ , the higher quality  $ad_2$  is, the higher CTR  $ad_1$  would get.

### 2.2 Indirect Identification

The indirect identification is designed as follows. If the ads compete with each other, the higher the quality of the co-occurred ad is, the more chance the current ad gets negative effect; and if the ads collaborate with each other, the higher the quality of the co-occurred ad is, the more chance the current ad gets positive effect. In the statistics, we gather sessions with the same first two ads, making the first ad the current ad and the second one the co-occurred ad. Totally, 700,942 gathered sessions are collected. Then we rank the gathered sessions according to the quality of the second ad, approximated by the CTR of the second ad. We divide the ranked sessions into 15 bins with equal size. In each bin, we make statistics about the percentage of the sessions that have the positive effect, and draw the result in Figure 2.

If the ads compete with each other, the result should be a single trend from top to bottom; and if the ads collaborate with each other, the result should be a single trend from bottom to top. However, as shown in the figure, the result first goes up and then goes down. This means that both collaborating and competing influences exist among ads. When the quality of the co-occurred ad is below average, the collaborating influence takes more effect; and when the quality of the co-occurred ad is above average, the competing influence takes more effect.

### 2.3 Direct Identification

To directly identify the effect of competing and collaborating influences, we conduct the following data analysis. An ad  $ad_1$  is shown in the first position with different ads in the second position. When it is shown with  $ad_{2a}$ , the CTR of  $ad_1$  is  $CTR_{2a}(ad_1)$  and the CTR of  $ad_{2a}$  is  $CTR(ad_{2a})$ ; and when it is shown with  $ad_{2b}$ , the CTR of  $ad_1$  is  $CTR_{2b}(ad_1)$  and

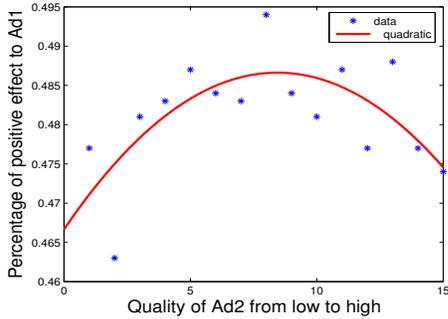


Figure 2: Indirect identification.

the CTR of  $ad_{2b}$  is  $CTR(ad_{2b})$ . If we consider the competing influence, we should have  $(CTR_{2a}(ad_1) - CTR_{2b}(ad_1)) * (CTR(ad_{2a}) - CTR(ad_{2b})) < 0$ . For example, if the quality of  $ad_{2a}$  is better than  $ad_{2b}$ , the CTR of  $ad_1$  would be decreased. In the same way, if we consider the collaborating influence, we should have  $(CTR_{2a}(ad_1) - CTR_{2b}(ad_1)) * (CTR(ad_{2a}) - CTR(ad_{2b})) > 0$ .

In the statistics, we collect 10,356 data samples. Each data sample contains a first ad and two co-occurred second ads. We rank all the data samples according to the quality of the co-occurred ad (the average of the two ads at the second position). In the ranked list, we divide the data samples into 19 bins. The first bin has the lowest quality and the last bin has the highest. For each data sample in a bin, we calculate the value of  $(CTR_{1a} - CTR_{1b}) * (CTR_{2a} - CTR_{2b})$ . For each bin, we make statistics about the percent of positive values divided by the total data sample size in a bin. We draw the percent values of different bins in Figure 3.

If there is no influence among the ads, the value in all these bins should around 0.5; if there is competing influence, the value should below 0.5; and if there is collaborating influence, the value should be above 0.5.

From the figure, it can be observed that the result drops down as the quality of co-occurred ad increases. When the quality of co-occurred ad is above average, the result is below 0.5. This means the ads show more competing influence, which is consistent with previous work [11]. However, when it is below average, the result is above 0.5. This means that the collaborating influence takes more effect under this case. This can demonstrate that the collaborating influence does exist in the ads. It takes more effect when the quality of the co-occurred ad is below average.

These two identifications have shown that both competing and collaborating influences exist among ads. Previous work [11], however, only identifies and utilizes the competing influence. Thus when the quality of the co-occurred ad is below average, the model would fail to work.

### 3. FULL RELATIONAL MODEL

#### 3.1 Incorporating Collaborating Influence

In this paper, we propose the full relational model (FRM) to incorporate both collaborating and competing influences. The user behavior of FRM is shown in Figure 4. Before examining the ads in detail, we assume that the user skims all the ads first. In the skimming process, the user is supposed to do the following two steps.

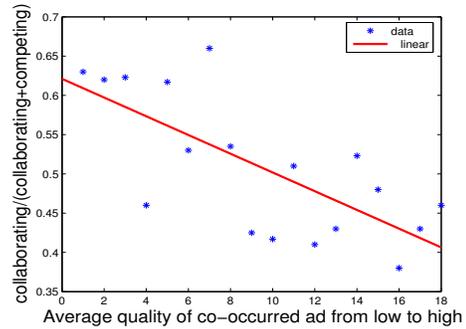


Figure 3: Direct identification.

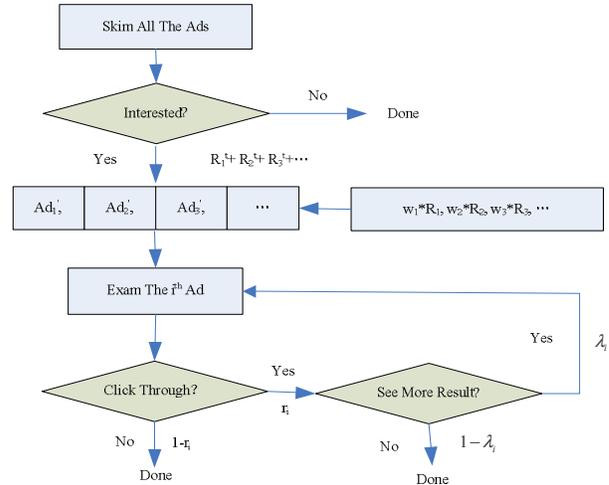


Figure 4: User behavior of the full relational model.

The first is to see whether he/she is interested in examining the ads. If the user is not interested in the ads, he/she would leave at once without any further examination. The interested issue would be determined by the overall relevance of all the ads. We define the probability of “interested” as

$$p(\text{Interested} = 1) \propto \sum_{i=1}^M r_i.$$

Therefore, the collaborating influence could be modeled in this step. If the qualities of all of the ads to the query are good enough, the probability of being interested would be larger; or otherwise, it would reduce the probability for further examination.

The second is to re-ranked the ads according to the value of  $w_i * r_i$ .  $r_i$  is defined as  $r_i = p(C_i = 1 | E_i = 1)$ .  $w_i$  is a constant, which denotes a global weight for the ad at position  $i$ . Intuitively, since the position-bias exists, the ad at the latter position should outperform the previous one by a certain margin to win the priority of being examined. Under this assumption, if an ad is much more relevant than others, it would take the first chance to be examined. Consequently, the competing influence among the ads is modeled by this step.

In designing the model, a fact should be considered that when the quality of the co-occurred ad is below average, the collaborating influence would take more effect; and when it is above average, the competing influence would take more

effect. Thus we propose to utilize a heuristic way to solve this fitting. Consider the function  $f(x) = x^t$ , ( $t < 1, 0 < x < 1$ ), when  $x$  is small, the increasing speed of  $f(x)$  is high; and when  $x$  is larger, the increasing speed becomes lower. This means that if we change the probability of “Interested” to

$$p(\text{Interested} = 1) \propto \sum_{i=1}^M r_i^t,$$

the strength of the collaborating influence would be weakened when the value is above the average; and it would be strengthened then below the average. The competing influence would take more effect in the former case, and would be the opposite in the latter case, which is just as expected.

After these two steps, the ads in the re-ranked list are assumed to be examined as previous click models.

In the learning process, the parameters for estimation include  $r$ ,  $\lambda$ , and  $w$ . For simplicity,  $w$  is set empirically.  $\lambda$  would be estimated based on the maximum log-likelihood method similar to [6]. In the following, we would introduce how to estimate  $r$  in detail.

### 3.2 Estimation of $r$

We utilize the Bayesian estimation to learn  $r$ . Suppose for an query-ad pair with  $r$ , it occurs at different positions in a set of sessions  $X^1, \dots, X^U$ . According to Bayes’s rule, we have

$$p(r|X^{1:U}) = \frac{p(X^{1:U}|r) * p(r)}{p(X^{1:U})} = \frac{p(X^{1:U}|r) * p(r)}{\int_r p(X^{1:U}|r) * p(r) dr}.$$

$$p(X^{1:U}|r) = \prod_{k_1}^{n^{(1)}} r * p(E^{k_1} = 1) * \prod_{k_2}^{n^{(0)}} (1 - r * p(E^{k_2} = 1)).$$

$$\prod_{k_2}^{n^{(0)}} (1 - r * p(E^{k_2} = 1)) = \prod_{k_2^1}^{n^{(0)0}} (1 - p(E^{k_2^1} = 1)) * \prod_{k_2^2}^{n^{(0)1}} p(E^{k_2^2} = 1) * (1 - r).$$

$n^{(0)1}$  denotes the ads that are examined by the user but are not clicked, and  $n^{(0)0}$  denotes the ads that are not examined by the user thus are also not clicked. If the user clicks an ad, he/she must has already examine the ad, thus the probability is  $r * p(E^{k_1} = 1)$ ; if the user does not click an ad, if he/she has examined the ad, the probability is  $p(E^{k_2} = 1) * (1 - r)$ ; or otherwise if the user does not examine the ad, the probability is  $1 - p(E^{k_2} = 1)$ .

In the formula,  $p(E = 1)$  could be calculated by utilizing the current  $r$  in the system. In this paper, for ad that has not enough evidences to estimate  $r$ , we utilize features (such as the category average CTR, etc.) to predict the value. We suppose in the re-ranked list,  $ad_i$  is in position  $i$ . For the  $i^{th}$  position, we could have

$$p(E_i' = 1) = \prod_{j=1}^{i-1} r_j' * \lambda_j.$$

The main problem is to determine  $n^{(0)1}$  and  $n^{(0)0}$ . If the user has not clicked an ad, we do not know whether he/she has examined it. To solve this problem for simplicity, we

utilize a sampling trick to sample these hidden variables in the calculation according to  $p(E = 1)$ . We have

$$p(E_i' = 1|C_i' = 0) = \frac{p(E_i' = 1)(1 - r_i)}{1 - p(E_i' = 1) * r_i}.$$

Thus we sample a random number from 0 to 1. If it is smaller than  $p(E_i' = 1|C_i' = 0)$ , we put it into  $n^{(0)1}$ ; or otherwise, we put it into  $n^{(0)0}$ .

After sampling  $n^{(0)1}$  and  $n^{(0)0}$ , we have

$$\begin{aligned} p(r|X^{1:U}) &= \frac{p(X^{1:U}|r) * p(r)}{\int_r p(X^{1:U}|r) * p(r) dr} \\ &= \frac{r^{n^{(1)}} * (1 - r)^{n^{(0)1}} \frac{1}{B(\alpha, \beta)} * r^{\alpha-1} (1 - r)^{\beta-1}}{Z} \\ &= \frac{r^{[n^{(1)} + \alpha] - 1} (1 - r)^{[n^{(0)1} + \beta] - 1}}{B(n^{(1)} + \alpha, n^{(0)1} + \beta)} \\ &= \text{Beta}(r|n^{(1)} + \alpha, n^{(0)1} + \beta). \end{aligned}$$

Thus,  $p(r|X^{1:U})$  follows beta distribution. By utilizing the expectation of beta distribution, we could have

$$E(r|X^{1:U}) = \frac{n^{(1)} + \alpha}{n^{(1)} + n^{(0)1} + \alpha + \beta}.$$

## 4. EXPERIMENT

### 4.1 Dataset and Experimental Setup

The dataset we utilized is the click-through log from AT&T. It is collected from September 1st 2011 to September the 15th 2011. In the search engine, when a user submit a query, 6 slots are available for ads impression. Totally, there are 2,339,992 sessions with 11,419,160 ad impressions. Over 67% of the sessions have full six ads. There are 197,119 unique queries and 1,408,009 unique ads.

We utilize the area under receiver operating characteristic (ROC) curve as our metric. We calculate the ROC under two scenarios. In the first scenario, we divide the queries into 5 groups, from the most infrequent to the most frequent queries. For each groups, we use the ads in all positions to calculate the ROC value. In the second scenario, we use the ads in all queries, but divide them into 6 groups according to their positions, from the top to bottom.

The baseline method we utilized is an average-based method. In this method, we utilize constants to model the probability of examination at different positions. The competitive methods we utilized for comparison are the CCM model and the TCM model.

### 4.2 Overall Performance

Table 1 shows the overall performances of different algorithms with different query sets. Table 2 shows the overall performances with different positions. The first observation is that TCM constantly outperforms the CCM and the baseline method. This demonstrates that the TCM model is effective in modeling the competing influence among ads and effective in improving the performance. The second observation is that FRM constantly outperforms the TCM. The main reason is that when the quality of co-occurred ad is below average, the collaborating influence, rather than the competing influence, takes more effect. Thus previous TCM model cannot perform well under such cases. For FRM, it considers both collaborating and competing influences, thus

**Table 1: Overall performance with different query sets**

Methods	QrySet1	QrySet2	QrySet3	QrySet4	QrySet5
baseline	0.7337	0.7176	0.6992	0.7010	0.7171
CCM	0.7338	0.7206	0.7009	0.7033	0.7176
TCM	0.7533	0.7315	0.7131	0.7102	0.7113
FRM	0.7547	0.7333	0.7160	0.7159	0.7209
Impr. to baseline	2.86%	2.18%	2.40%	2.12%	0.529%
Impr. to CCM	2.84%	1.76%	2.15%	1.79%	0.459%

**Table 2: Overall performance with different positions**

Methods	Pstn 1	Pstn 2	Pstn 3	Pstn 4	Pstn 5	Pstn 6
baseline	0.6150	0.6168	0.6107	0.5988	0.5957	0.5995
CCM	0.6204	0.6221	0.6141	0.6097	0.6064	0.6016
TCM	0.6167	0.6317	0.6294	0.6251	0.6386	0.6378
FRM	0.6359	0.6418	0.6400	0.6339	0.6404	0.6443
Impr. to baseline	3.39%	4.05%	4.79%	5.86%	7.50%	7.47%
Impr. to CCM	2.49%	3.16%	4.21%	3.96%	5.60%	7.09%

it would perform better. This result demonstrates that the FRM is effective in modeling both kinds of influences. The improvement of our proposed model is significant. Compared with the competitive model CCM, the improvements in different configurations are from 2.5% to 7.5%. Especially, from Table 2, we could see that the proposed model is very effective in improving the performance for the lower positions.

## 5. CONCLUSION

In this paper, we investigate how to improve click models by utilizing the influence among ads. Through real-world data analysis, we identify that the competing influence only takes effect when the qualities of co-occurred ads are above the average, and when the qualities of those are poor, its opposite, collaborating influence would take more effect. We propose a novel click model to incorporate both collaborating and competing influences. Through experimental verification, the proposed model has a significant improvement compared with previous work.

## 6. ACKNOWLEDGMENTS

The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CUHK 415311 and No. CUHK 415210), a Google Focused Grant Project under “Mobile 2014” and a grant from Beijing Key Laboratory of Intelligent Information Technology in BIT.

## 7. REFERENCES

- [1] O. Chapelle and Y. Zhang. A dynamic bayesian network click model for web search ranking. In *Proceedings of the 18th international conference on World wide web, WWW '09*, pages 1–10, New York, NY, USA, 2009. ACM.
- [2] N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey. An experimental comparison of click position-bias models. In *Proceedings of the international conference on Web search and web data mining, WSDM '08*, pages 87–94, New York, NY, USA, 2008. ACM.
- [3] R. Cristi. *Hegel on Freedom and Authority*. 2005.
- [4] G. E. Dupret and B. Piwowarski. A user browsing model to predict search engine click data from past observations. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '08*, pages 331–338, New York, NY, USA, 2008. ACM.
- [5] F. Guo, C. Liu, A. Kannan, T. Minka, M. Taylor, Y.-M. Wang, and C. Faloutsos. Click chain model in web search. In *Proceedings of the 18th international conference on World wide web, WWW '09*, pages 11–20, New York, NY, USA, 2009. ACM.
- [6] F. Guo, C. Liu, and Y. M. Wang. Efficient multiple-click models in web search. In *Proceedings of the second international conference on Web search and data mining, WSDM '09*, pages 124–131, New York, NY, USA, 2009. ACM.
- [7] T. Joachims. Optimizing search engines using clickthrough data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 133–142. ACM, 2002.
- [8] C. Liu, F. Guo, and C. Faloutsos. Bbm: bayesian browsing model from petabyte-scale data. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '09*, pages 537–546, New York, NY, USA, 2009. ACM.
- [9] R. Srikant, S. Basu, N. Wang, and D. Pregibon. User browsing models: relevance versus examination. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '10*, pages 223–232, New York, NY, USA, 2010. ACM.
- [10] C. Xiong, T. Wang, W. Ding, Y. Shen, and T. Liu. Relational click prediction for sponsored search. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 493–502. ACM, 2012.
- [11] W. Xu, E. Manavoglu, and E. Cantu-Paz. Temporal click model for sponsored search. In *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval, SIGIR '10*, pages 106–113, New York, NY, USA, 2010. ACM.
- [12] F. Zhong, D. Wang, G. Wang, W. Chen, Y. Zhang, Z. Chen, and H. Wang. Incorporating post-click behaviors into a click model. In *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval, SIGIR '10*, pages 355–362, New York, NY, USA, 2010. ACM.
- [13] Z. A. Zhu, W. Chen, T. Minka, C. Zhu, and Z. Chen. A novel click model and its applications to online advertising. In *Proceedings of the third ACM international conference on Web search and data mining, WSDM '10*, pages 321–330, New York, NY, USA, 2010. ACM.