Inferring Clickthrough Rates on Ads from Click Behavior on Search Results

Sreenivas Gollapudi
Microsoft Research
Mountain View, CA, USA
sreenig@microsoft.com

Rina Panigrahy
Microsoft Research
Mountain View, CA, USA
rina@microsoft.com

Moises Goldszmidt
Microsoft Research
Mountain View, CA, USA
moises@microsoft.com

ABSTRACT

Click logs provide valuable information that can be used to infer several parameters related to the relevance of search results and ads to queries. However often there are much fewer clicks on ads as compared to search results. Thus while the click logs can easily be used to study parameters related to search results the sparseness of the number of clicks for ads renders it infeasible to infer parameters related to ads. We show that the click behavior on ads is correlated with the click behavior on search results. The first correlation we show is that the drop in the CTR values as we move down the search results is correlated to the drop in CTR as we move down the ads. Second, we study the effect of the number of mainline ads on the CTR of the top ad. Finally, as an application of such correlations, we show how one can use the click data on search results to infer approximately the optimal number of ads for each query to maximize revenue from the top ad on the page.

1. INTRODUCTION

Click logs provide valuable information that can be used to infer several parameters related to the relevance of search results and ads to queries. However often there are much fewer clicks on ads as compared to search results. Thus while the click logs can easily be used to study click parameters for search results, the sparseness of the number of clicks for ads renders makes it difficult to study the same parameters for ads. We will show how some click behavior for ads can be extrapolated from that of the search results

The revenue of a search engine is directly related to the CTRs of the ad. The CTR depends not only on the quality of an ad, but also its position and other conditions such as the total number of ads in the page. Consider a specific scenario in which the search results page consists of up to three ads followed by ten search results. Is it better to show fewer ads or more ads? Is it even clear that a fixed number of ads works best for all queries? One intuition would suggest that more ads probably implies more clicks which might result in higher revenue. On the other hand, more ads may dilute the CTR of the top ad and since typically the revenues are dominated by those from the top ad, any gain in revenue from the additional ads may be less than the loss in revenue in the top ad, thus favoring few ads. Experimentally computing the optimal number of ads by varying the number for each query separately may not be feasible for several reasons; there are too many queries, the query set is dynamic, and a large number of trials may need to be wasted in estimating these CTRs (since CTR of ads is usually lower than that of search results.) It is therefore useful to to be able to deduce this optimal number in other ways without explicit experimentation for each query.

WSDM'11, February 9–12, 2011, Hong Kong, China.
ACM 978-1-4503-0493-1/11/02.

In this paper we show that the click behavior on ads is correlated with the click behavior on search results. The first correlation we show is the drop in the CTR values as we move down positions in the search results is correlated to the drop in CTR as we move down the ads. Next we study the effect of the number of mainline ads (ads on top of the page) on the CTR of the top ad and we show that this is also correlated to several click-based features in the search results. We show how these correlations can be used to infer the optimal number of ads for each query. We will demonstrate that the optimal number of ads is very much dependent on the query and can be estimated to some extent by observing other properties of the query such as CTR of the search result positions for the query, its distribution, frequency and type of the query – for example, we found that if the CTR of the top search result is low then it helps to show few ads.

For privacy reasons, we will study the problem of maximizing the number of ad clicks instead of the revenue so as not to reveal the bid values of the ads. All our techniques can be applied to the problem of maximizing revenue. For the purposes of our study we make use of a data set of click logs consisting of $62,921$ queries over a period of 3 months on a commercial search engine. These are randomly chosen head queries for which results have been shown with 3 ads as well as with 1 ad. Thus they can be used to study the effect of changing number of ads on CTR values. We measure the CTR of the search result positions and the ads positions. We will refer to the 3 ads on top of the page as main line ads and use $ML1 \dots ML3$ to refer to the positions. The algorithm result positions will be referred to as $S1 \dots S10$. CTR_{Si} and CTR_{MLi} will denote the CTRs of these positions for some given query.

2. CORRELATION BETWEEN DROP IN CTR VALUES IN ADS AND SEARCH RESULTS

It is well known that the CTR values of different search positions tends to decrease as we go down and this is often referred to as position bias. In this section, we will highlight the correlation between drop in CTRs as we go down the positions for both ads and search results. A simple mea-

query-bucket	median entropy search positions	median entropy ads
1	0.4260	0.3450
2	0.6140	0.4180
3	0.7780	0.4190
4	0.9580	0.4280
5	1.2260	0.4345

Table 1: Median values of the entropy of CTR distributions in search result and ad positions across the five buckets. Observe that the values are in the same order for both search results and ads thus indicating a correlation

surement in our data set shows a correlation between the drop in CTRs in the top five search result positions and the three ads. To quantify this drop, we compute the entropy of the CTRs for the five search result positions as a distribution (i.e., we scale the CTRs so that they sum to 1 and then compute the entropy); let S_q denote this entropy for the distribution for a query q. Similarly, let A_q denote the corresponding entropy for the three ad positions. Note that sharp drop's in CTRs means the entropy will be low and a near equal values of CTRs across positions will give a high entropy.

The correlation between these two query features measured using the Kendall-Tau coefficient is 0.042; this value is statistically significant as the value for two random streams is normally distributed with mean 0 and variance $2(2n+5)/(9n(n-1)) = 4/(9 \cdot 62921) = 0.0000076$ [12]. To further illustrate the correlation, we sort the queries by S_q and divide into five equal sized buckets. Table 1 shows the median value of S_q and A_q in these five buckets. Observe that not only is the median value of S_q in increasing order (as expected), but also is the median value of A_q. Note that if S_q and A_q were two random features, the probability of these two sets of five median values being in the same order would be $1/5! = 1/120$. This indicates a significant correlation between S_q and A_q.

Next, we also normalize the CTR of all search result positions by the CTR of the top search result position and that of all ad positions by the CTR of the top ad position; that is we compute CTR_{Si}/CTR_{S1} for the search positions and CTR_{MLi}/CTR_{ML1} for the ads. We then compute the median value of the normalized CTR for all positions in each bucket. Figures 1 and 2 shows the plots median values in the five buckets for the search and ad positions respectively. Again, note that the curves for the ad positions are in in strictly increasing order thus demonstrating correlation.

3. EFFECT OF NUMBER OF ADS ON THE CTR OF TOP AD

We will now see how to decide the optimal number of Ads for a query so as to maximize the CTR of the top mainline position. We note that the revenues from the top mainline position typically dominates the total revenues of a search engine as the CTR and the bids tend to be highest at the top position. We will restrict ourselves to the options of showing only one ad or three ads. We will assume the ads are shown from a distribution. For a query q, let x_q denote the expected CTR of the top ad position $ML1$ assuming one ad. Similarly, let y_q denote the CTR of $ML1$ assuming three ads. Let $\Delta_q = x_q - y_q$ and let b_q be an indicator variable that is 1 if $\Delta_q > 0$ and zero otherwise. Ideally, if the value of b_q is

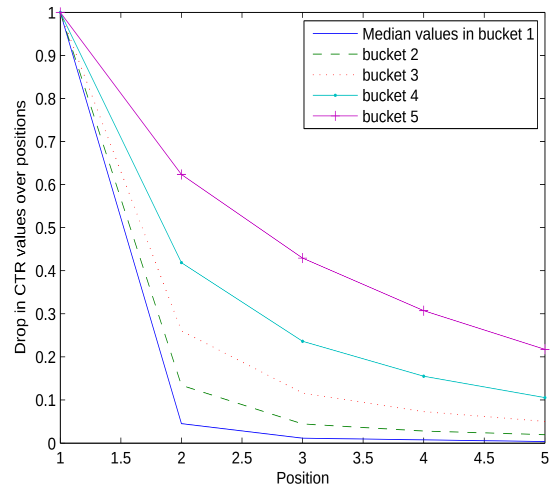

Figure 1: Median values of normalized CTR values in the five buckets for the search result positions

known separately for each query, we would use this to show the optimal number of ads for q. Unfortunately as stated before, computing b_q explicitly by trying out different values for every query is impractical. So we will try to estimate b_q from other known properties of the query. We make use of the query features listed in table 2. One of the features is commercial-intent that is a bit indicating if the query is likely to be a commercial query. This bit is computed using logistic regression on automatically generated trained data from a commercial toolbar logs [10].

We measure the CTRs of the different ad and search result positions for the 62,921 queries in our click log. Note that all these queries were shown with both a single ad and with three ads. These CTRs give us estimates for x_q and y_q for each query. Figure 3 shows the distribution of $\Delta_q = x_q - y_q$. Observe that it takes both positive and negative values. Note that if we show three ads for every query, the total number of clicks is $\sum f_q y_q = 26689496$, where f_q is the frequency of the query q. Whereas, when we show one ad, the total number of clicks drops slightly to $\sum f_q x_q = 25697231$. However, if we show the optimal number of ads based on b_q, the number of clicks is $\sum f_q \max(x_q, y_q) = 31450241$. Thus, by selectively showing one ad or three ads depending on the query, we can increase the number of clicks by $5256877 = 20\%$. The important task is to find the queries for which it takes positive values and show only one ad for these queries. We will compute an estimate $\tilde{b}_q$ for b_q and show one or three ads depending on whether $\tilde{b}_q$ is 1 or 0. The number of clicks in the top ad position based on this rule is

$$\sum f_q(x_q\tilde{b}_q + (1-\tilde{b}_q)y_q) = f_q(y_q + \tilde{b}_q\Delta_q) = f_q y_q + f_q\tilde{b}_q\Delta_q.$$

. Thus $\tilde{b}_q f_q \Delta_q$ denotes increase in number of clicks obtained over the default rule of always showing three ads. Thus we need an estimate $\tilde{b}$ so as to maximize $\tilde{b}_q f_q \Delta_q$.

In order to empirically verify the correlation between the search features and the gain in clicks we performed exploratory data analysis using various statistical techniques including, logistic and linear regression, Bayesian network classifiers,

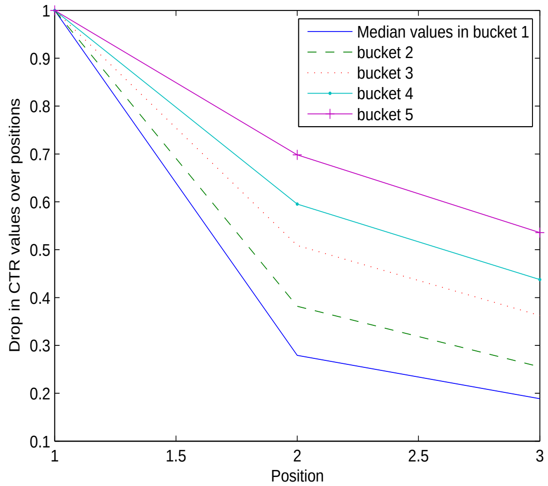

Figure 2: Median values of the normalized CTR values in the five buckets for the ad positions. Note that the five curves for ads and search results over the buckets are in the same order thus implying a correlation

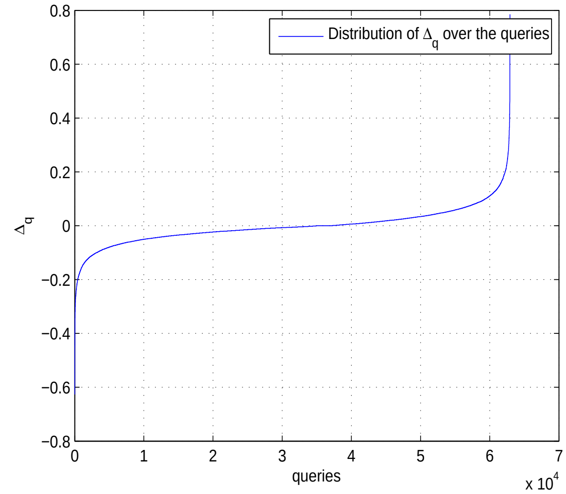

Figure 3: Distribution of Δ_q over queries

QueryFeature	Kendall Tau coeff.	Gain in clicks as % of best possible	Coeff.in linear model
CTR S1	-0.0261	10.5384	-105.2796
CTR S2	0.0171	4.6164	0
CTR S3	0.0302	6.5452	0
CTR S4	0.0354	5.8445	-172.3322
CTR S5	0.041	5.5136	-642.908
Frequency	0.0039	0.3601	580.0018
Commercial Intent	0.0017	0.0159	170.0388
Entropy of CTR distr.	0.0153	5.535	-0.0019

Table 2: List of features

and a classifier based on boosting (using decision tree stomps). We will report on the results for the linear regression as the rest of the models achieved similar gains and accuracy.

We used ridge linear regression to shrink some of the regressors — more sophisticated feature selection methods were not needed as we used at most 10 regressors [12]. We also used 5-fold cross-validation in order to estimate the mean of the gains and the variance, and to ensure that indeed whatever correlation we found was statistically significant. The linear model obtained is statistically significant as all the coefficients for the regressors exhibited a confidence interval away from zero with a small p-value, and none of the coefficients were inflated. Table 2shows the Kendall Tau correlations between Δ_q and the different features; It also shows the gain in the number of clicks (as a percentage of the best possible gain of 5256877) from each feature separately, and the coefficients of the features when used cumulatively in a linear model. We note that Δ_q is negatively correlated to the CTR of the top search result position $S1$. Figure 4 further illustrates this correlation. In this figure, we sort all the queries in our data set by CTR of $S1$ and then plot the cumulative sums $\sum_{i=1}^{x} \Delta_{q_i}$ where the sum is over the first x queries in this sorted order. Note that this sum seems to increase steadily (with small local drops) and then seems to drop almost steadily. This indicates that Δ_q is mostly positive till a certain point and then mostly negative. Thus if CTR of the top algorithmic result position is low then showing only one ad tends to increase the CTR of the top Ad. This observation may seem counter-intuitive but perhaps one rationale is that if the top algorithmic result is high quality and if there is only one ad above it then the ad needs to directly compete with the algorithmic result just below it. In such cases having a buffer of two more ads increases the distance between the top ad and the top algorithmic result and thus helps increase the CTR of the top mainline ad.

Using our model we get estimates $\tilde{b}_q$ for the bit b_q indicating the sign of Δ_q. The model is the sign of the hyperplane $-105c_1 - 172.3c_4 - 642.9c_5 + 580c_6 + 170c_7 - 0.0019c_8 + 78.9 > 0$, where $c_1 \ldots c_8$ are the eight features used. Using these predicted values, we are able to increase the total number of clicks by about 677055 which is about 13% of the best possible gain. The table 3 shows the gain as a percentage of the optimal gain in each of the five folds of the cross-validation. Note that the gains are consistently close to 13% in all the five validations. More sophisticated generalized linear models should be applied to close the gap between what a linear model can achieve and the maximal gain. Yet, that is the subject of future research.

4. RELATED WORK

Earlier work largely focused on studying Clickthrough rates on ads and search results separately [5, 9, 1, 16, 15, 14, 3, 17, 4, 11, 8]. The related work can be classified into click prediction in contextual advertisement and sponsored search. While research in contextual advertising has studied the effect of page content and the user on the clickthrough rate on an ad, most models on predicting clickthrough rates in sponsored search have not considered the effect of the user query and other links (specifically search results) on the page. In fact, the basic assumptions that underscored several of these works are that users inspect documents from top to bottom and the clickthrough rates on a result largely depended on a) the probability of a inspecting the document in a certain position that is *mostly* dependent on the quality of the re-

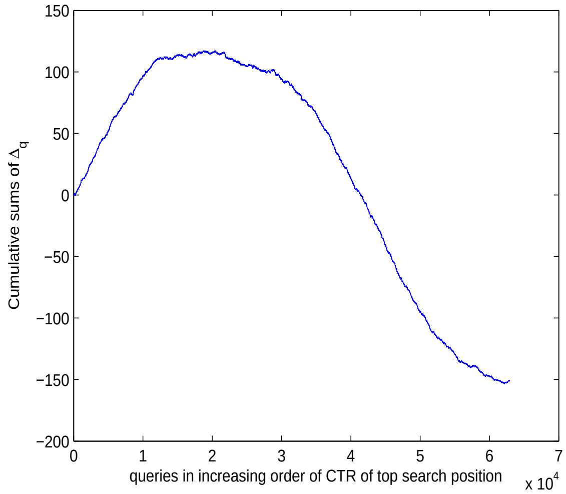

Figure 4: Correlation between CTR of top search result and effect of number of ads on ad clicks

Gain as percentage of optimal gain
12.13
14.78
14.62
11.89
11.43

Table 3: Gains in five fold cross validation as percentage of optimal gain

sults above it and b) the quality of the result itself. Such a behavior would not give rise to correlations in drops in CTRs in different regions of the results page that our study finds. Further our results also show that a CTR of a position also depends on the quality of the document below it. This corroborates with models that suggest the user always see document(s) immediately after the clicked document (see, e.g., [9, 6] and references within). Other related work on click prediction specifically deals with estimating the position bias in the user's browsing and clicking behavior on search results [5, 13]. Some more recent related work includes [2, 7]

5. CONCLUSIONS

In this paper we showed how the click data on ads is correlated to the click data on search results. Thus even in the absence of rich click data on ads, one can infer several important parameters from the click data on search results. In particular we showed how to predict the optimal number of mainline ads for each query separately to maximize the revenue from the top mainline ad. Indeed it should be possible to get much more accurate predictions by using additional features; we leave this task as an interesting open problem for future work.

6. REFERENCES

[1] Eugene Agichtein, Eric Brill, Susan T. Dumais, and Robert Ragno. Learning user interaction models for predicting web search result preferences. In *SIGIR*, pages 3–10, 2006.

[2] Andrei Broder, Massimiliano Ciaramita, Marcus Fontoura, Evgeniy Gabrilovich, Vanja Josifovski, Donald Metzler, Vanessa Murdock, and Vassilis Plachouras. To swing or not to swing: learning when (not) to advertise. In *Proceeding of the 17th ACM conference on Information and knowledge management*, CIKM '08, pages 1003–1012, New York, NY, USA, 2008. ACM.

[3] Andrei Z. Broder, Massimiliano Ciaramita, Marcus Fontoura, Evgeniy Gabrilovich, Vanja Josifovski, Donald Metzler, Vanessa Murdock, and Vassilis Plachouras. To swing or not to swing: learning when (not) to advertise. In *CIKM*, pages 1003–1012, 2008.

[4] Deepayan Chakrabarti, Deepak Agarwal, and Vanja Josifovski. Contextual advertising by combining relevance with click feedback. In *WWW*, pages 417–426, 2008.

[5] Nick Craswell, Onno Zoeter, Michael J. Taylor, and Bill Ramsey. An experimental comparison of click position-bias models. In *WSDM*, pages 87–94, 2008.

[6] Edward Cutrell and Zhiwei Guan. What are you looking for?: an eye-tracking study of information usage in web search. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 407–416, 2007.

[7] Cristian Danescu-Niculescu-Mizil, Andrei Z. Broder, Evgeniy Gabrilovich, Vanja Josifovski, and Bo Pang. Competing for users' attention: on the interplay between organic and sponsored search results. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 291–300, New York, NY, USA, 2010. ACM.

[8] Georges Dupret and Ciya Liao. A model to estimate intrinsic document relevance from the clickthrough logs of a web search engine. In *WSDM*, pages 181–190, 2010.

[9] Georges Dupret and Benjamin Piwowarski. A user browsing model to predict search engine click data from past observations. In *SIGIR*, pages 331–338, 2008.

[10] Ariel Fuxman, Anitha Kannan, Andrew B. Goldberg, Rakesh Agrawal, Panayiotis Tsaparas, and John C. Shafer. Improving classification accuracy using automatically extracted training data. In *KDD*, pages 1145–1154, 2009.

[11] Fan Guo, Chao Liu 0001, and Yi Min Wang. Efficient multiple-click models in web search. In *WSDM*, pages 124–131, 2009.

[12] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics, second edition edition, 2009.

[13] Thorsten Joachims, Laura A. Granka, Bing Pan, Helene Hembrooke, Filip Radlinski, and Geri Gay. Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. *ACM Trans. Inf. Syst.*, 25(2), 2007.

[14] M. Regelson and D. Fain. Predicting click-through rate using keyword clusters. In *Proceedings of the Second Workshop on Sponsored Search Auctions*, 2006.

[15] Matthew Richardson, Ewa Dominowska, and Robert Ragno. Predicting clicks: estimating the click-through

rate for new ads. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 521–530, 2007.

[16] Benyah Shaparenko, Özgür Çetin, and Rukmini Iyer. Data-driven text features for sponsored search click prediction. In *ADKDD '09: Proceedings of the Third International Workshop on Data Mining and Audience Intelligence for Advertising*, pages 46–54, 2009.

[17] Xuerui Wang, Andrei Z. Broder, Marcus Fontoura, and Vanja Josifovski. A note on search based forecasting of ad volume in contextual advertising. In *CIKM*, pages 1343–1344, 2008.