Get more Clicks!

Derek Hao Hu
derekhh@cse.ust.hk

Evan Wei Xiang
wxiang@cse.ust.hk

Qiang Yang
qyang@cse.ust.hk

Department of Computer Science and Engineering
Hong Kong University of Science and Technology

ABSTRACT

Sponsored search has become increasingly important due to the rapid development of Web search engines and pay per click (PPC) is amongst one of the most important advertising models search engines currently use. One of the key questions in sponsored search is that: Given a query or a substituted keyword, which ads should search engines display to the users in order to maximize their revenue? In other words, given a keyword, how can we choose ads out of a candidate list that will have higher click-through rates (CTR)? Previous works have attempted to estimate the CTR of ads via a query-independent perspective. In this paper, instead of predicting the CTR of ads, we will propose a new ranking-based approach to select ads that would have higher click-through rates via a query-dependent perspective. We first analyze some manually constructed heuristic rules that could be used to distinguish good ads from bad ones and then show how we could combine these rules into our ranking-based approach to reach our aim. Experiments on real-world datasets have confirmed the effectiveness of our proposed approach.

Categories and Subject Descriptors

I.2.6 [**Artificial Intelligence**]: Learning; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms

Algorithms, Experimentation, Economics

Keywords

click-through rate, sponsored search, paid search, pay per click, ranking

1. INTRODUCTION

Internet-based advertising (also known as paid search, sponsored search etc.) has undoubtedly become one of the most popular ways of textual advertising these days. A large portion of search engine companies' revenue come from paid search. The market for Internet-based advertising has risen to \$10 billion and will approximately reach \$24 billion by 2013[1].

There are many Internet advertising models and in this paper, we focus our attention on the commonly used cost per click (CPC) model (sometimes also known as pay per click (PPC) model). In CPC model, advertisers only pay when a user actually clicks on an advertisement and then visits the advertisers' website. Such a CPC model is currently widely being used by major search engines like Google, Yahoo and Microsoft Live Search. Other major advertising models include cost per impression (CPM), where advertisers pay money according to how many times there ads are shown, and cost per action (CPA), where advertisers only pay when the users actually complete a transaction.

Click-through rate (CTR) is a way of measuring the success of an online advertising campaign. A CTR is obtained by dividing the number of users who clicked on an ad on a web page by the number of times the ad was delivered (impressions). For example, if an ad was delivered 100 times (impressions delivered) and one person clicked on it (clicks recorded), then the resulting CTR would be 1 percent. In CPC model, it is obviously seen that the objective for search engines is to "persuade" people to take actions, in other words, click the shown advertisements, based solely on the advertisement descriptions - usually a few, at most around 100, well-chosen words.

The main process of CPC-based online advertising is as follows. Users submit a query and search engines substitute the query to a given keyword that may match some advertisers' bid keywords. Then, the search engines would choose some ads (usually 1 to 5) to be displayed according to some ad ranking criteria, amongst which one of the most important criteria is to rank by expected revenue, i.e. the product of the advertisements' bid amounts and the advertisements' estimated click-through rates. Since advertisements' bid amounts are known in advance, such a ranking procedure could be reduced to estimate the advertisements' click-through rates. Previous research work [5] have also attempted to estimate the click-through rate for new ads via a machine learning-based approach, or more precisely, a regression-based approach, where the click-through rates learned are irrelevant of the issued queries.

However, it is noteworthy to see that, when estimating the click-through rate for ads, it is more reasonable to take

[1] http://www.emarketer.com/Article.aspx?id=1006644

into the user issued queries into account. Since it would largely affect whether users' would actually click the shown ads. Thus, in this paper, we propose a new approach to estimate the *query-dependent* click-through rates for advertisements, instead of the *query-independent* approaches as shown in previous research works [5]. Furthermore, instead of the traditional regression-based approaches to estimate the click-through rates directly for advertisements, in this paper we propose a ranking-based approach to estimate the ranking of advertisements and then directly choose the top-ranked advertisements. Such a motivation is influenced by the EigenRank approach [2] proposed to solve the collaborative filtering problem and we would discuss our motivation for adopting the ranking-based approach later in detail.

The remainder of this paper is organized as follows. In Section 2, we would briefly discuss some previous research works related to sponsored search, in particular we would discuss some previous research works in predicting CTR, ad ranking and generating / substituting advertisement keywords. In Section 3, we would briefly review some important terminologies and background knowledge in CPC model and CTR prediction. In Section 4, we would discuss some manually constructed tips to distinguish good ads from ads and we would perform some statistical analysis of the usefulness and the coverage of such tips on a real-world commercial search engine data set. In Section 5, we would discuss our ranking approach in detail and in Section 6, we would show our algorithm performance on the real-world dataset and validate the effectiveness of our approach. Finally in Section 7, we would conclude this paper and discuss some possible directions in which we could carry on our future research work.

2. RELATED WORKS

There were many important research works related to sponsored search. Richardson et al. [5] tried to solve the problem of predicting the click-through rates for new ads or rarely clicked ads. As described in the previous section, their approach aims to estimate CTRs *explicitly* and is independent of issued queries. Craswell et al. evaluated different models of user search result browsing to analyze the positional bias of click-through rates and found that cascade model predicts the observed bias accurately.

Ad ranking has also been studied from a variety of perspectives. Some theoretical computer scientists try to model the ad ranking problem as a online bipartite matching problem and used revealing LP to derive an optimal algorithm with a competition ratio of $1 - 1/e$ of this problem [3]. Pandey and Olston [4] modeled the advertisement ranking problem as a multi-armed bandit problem and studied the tradeoff between exploration and exploitation.

Another interesting area worth studying is the query substitution problem for sponsored search, that is, user issued queries are replaced by other generated queries to match the keywords bid by advertisers. Jones et al. [1] built a model for selecting between candidate substitution queries by using a number of features relating the query-candidate pair. Later, in [6], an active learning algorithm is used to select the most informative samples to train the query rewriting relevance model. There are many important research works on the rather broad area of sponsored search. However, it is beyond our scope and page limit to describe them in detail.

3. BACKGROUND KNOWLEDGE

In this paper, our objective is to let search engines automatically choose the "good" ads to display, given a specific user query. So what is a "good" advertisement? In previous sections, we had already given the definition of click-through rates. It is natural and a conventional manner to use CTR as a means of "scoring criteria" for each ad. However, in real-world situations, an advertisement is usually binded to several keywords, and even the same advertisement might have different CTR when the keywords are different. Therefore, it is not reasonable to forget the impact of keyword on CTR.

Thus, our problem can be formally defined as follows: Given a keyword k_n, we have a set of candidate ads $A_n = \{ad_{ni}\}$, each of which is a potential match to the keyword k_n. Thus, we want to select the top-ranked advertisements such that their click-through rates (given the keyword is k_n) are maximum in the set A_n.

4. TIPS TO WRITE GOOD ADS

With the development of World Wide Web and search engines, web sites that provide news and information about search engines and search engine marketing, such as Search Engine Watch[2], are emerging everyday. Articles on such web sites are usually published by domain experts or speculators, who could offer a good analysis for search engines and service development. In this paper, We first gather some expert tips from such authoritative sources, and then investigate how to use such tips to boost our CTR performance by rearranging the ads for their corresponding keywords.

4.1 Tips from Human Experts

Ad Group Keywords: It is required that the ad group's keywords should appear at least once in the ad, and it is much better if it can appear in the headline such that we could draw the user's attention.

Speak Directly to Your Audience: Ad readers would tend to feel better when they believe the ads are specially written for them. Therefore, using words like "you" and "yours" would make the readers feel you are directly offering service to them.

Call Them to Action: It is much better to use imperative verbs in your ad descriptions like "Get", "Shop", "See", "Find", "Buy" rather than phrases like "visit our site" or "click to see". It's best to call your readers to action and tell them what you want them to do.

Create a Sense of Urgency: The ad descriptions should try to create a sense of emergency in texts, e.g. let the readers believe they would suffer or fail to benefit if they don't act right away as told in the ads.

Free is Good: Trying to use free offers and explicitly mention them in your ad description would boost clicks and conversions a lot. Using free offers is quite beneficial, especially when your product or service is high-priced or rather complex, or the sales cycle is long.

Flaunt: If your product or service has many competitors, it would be better to underscore your advantages, e.g. using claims like "top ratings", "best-quality products", "maximize profits", etc. They could make your ads more attractive.

Capitalize Every Word: Often an ad with the first letter of each world capitalized often has a higher CTR than

[2] http://searchenginewatch.com/

a version of the same ad with lower-case letters. Capitalize individual words in the display URL also often boosts CTR.

4.2 Statistics from Historical Logs

We first categorize these expert tips into seven groups. For each tip, we filer the keyword-ad pairs S_t which satisfy the Tip_t:

$$S_t = \{\langle k_n, ad_{ni}\rangle | ad_{ni} \in \text{Tip}_t\}$$

We define the coverage for Tip_t as

$$\text{Coverage}(\text{Tip}_t) = \frac{|\{\langle k_n, ad_{ni}\rangle | \langle k_n, ad_{ni}\rangle \in S_t\}|}{|\{\langle k_n, ad_{ni}\rangle\}|}$$

Then we calculate the tip coverage and the average CTR over the satisfied ads for these keyword k_n:

$$\text{CTR}_{\text{Satisfied}}(k_n) = \text{average}(\{\text{CTR}(ad_{ni}) | \langle k_n, ad_{ni}\rangle \in S_t\})$$

We also calculate the overall average CTR for all the ads belonging to the keyword k_n:

$$\text{CTR}_{\text{Overall}}(k_n) = \text{average}(\{\text{CTR}(ad_{ni}) | \langle k_n, *\rangle\})$$

The effectiveness and coverage of some human expert tips are shown in Table 1. We can find that the CTRs of the ads which satisfy the tips are better than the average overall CTRs of the ads which belong to the same set of keywords. Some tips are able to boost the CTRs of the ads a lot, for example, Tip Set1, Set4, Set5 and Set6. However, we can also observe that the coverage of some powerful tips are quite low. For example, the first tip of Set4 and Set6 can only cover 7.5% of all the advertisements.

We try to use some thesaurus-based methods to improve the coverage of these tips. For example, we generalize the first and third tips of Set3 with an action verb list, and we also try to extend the first tip of Set6 with either a descriptive or a motivating adjective word list. We can find that, while the coverage of the extended tip is increasing, the improvement of the CTR drops. In the next section, we propose to use a ranking model to ensemble these effective but low covered tips together.

5. OUR RANKING MODEL

Previous works have attempted to estimate the CTR of ads via a query-independent approach. They propose to use a uniform regression model to predict the CTR of each ad independently. However, such a CTR prediction task is not essentially the objective of ad ranking for search engines. Search engines only need to pick out several top ranked advertisements for specific query keywords. Therefore, essentially we should focus on building a ranking model instead of a regression model. Such a motivation is influenced by [2], where the authors proposed a ranking model to predict the top ranked items for the collaborative filtering problem. Therefore, we plan to learn a ranking function to output the rank of the ads for each keyword instead of inferring their CTRs independently.

Assume that there exists an input space $X \in R^m$, where m denotes the number of features. For our ad ranking problem, the features can be either bag of words (BOW) representation of the ads or the expert tips. The output space of ranks is represented by a set of ordered labels $r^* = \{r_1, r_2, ...r_q\}$, in which $r_1 \succ r_2 \succ ... \succ r_q$. Here $r_{ni} \succ r_{nj}$ denotes a preference relationship, and in our ad ranking task it means ad_{ni} obtains better CTR than ad_{nj} for keyword k_n. Suppose given a set of ranked ads belong to N keywords $S = \{(\vec{x}_n, r^*_n)\}_{n=1}^N$, our objective is to learn a set of weights $\vec{w}$ via solving a constraint optimization problem:

$$\begin{aligned}
&\text{Minimize} : V(\vec{w}, \vec{\xi}) = \frac{1}{2}\vec{w}\cdot\vec{w} + C\sum \xi_{i,j,n} \\
&\text{Subject To :} \\
&\forall(ad_{1i}, ad_{1j}) \in r^*_1 : \vec{w}\cdot\Phi(k_1, ad_{1i}) \geq \vec{w}\cdot\Phi(k_1, ad_{1j}) + 1 - \xi_{i,j,1} \\
&... \\
&\forall(ad_{Ni}, ad_{Nj}) \in r^*_N : \vec{w}\cdot\Phi(k_N, ad_{Ni}) \geq \vec{w}\cdot\Phi(k_N, ad_{Nj}) + 1 - \xi_{i,j,N} \\
&\forall i \forall j \forall n : \xi_{i,j,n} \geq 0
\end{aligned} \tag{1}$$

where Φ is a linear function of the feature vector $\vec{x}$:

$$\Phi(k_n, ad_{ni}) = \langle \vec{w}, \vec{x}_{ni}\rangle$$

Comparing with the regression model, which aims to optimize the prediction loss on ads' CTR individually, our ranking model minimizes the misordered ad pairs with respect to their keywords.

6. EXPERIMENTS

6.1 Data Sets

In this section, we evaluate the effectiveness of our ranking model with historical ad clickthrough logs collected by a commercial search engine. For the historical logs, we collect two-week log data from a commercial search engine over 1.7 million ads for 0.3 million keywords. The log records the query the user issued, the displayed ads id with position information, and the clicked ads id. We filtered out the ads of which the impression number is lower than 200. After that there left 35,000 ads for about 6,000 keywords. We sampled 90% keywords for training and used the remaining 10% for testing, and we carried out the experiments for 10 times. For both the ranking model and regression model, we used the implementation in the SVM^{light} package[3].

6.2 Evaluation Metric

Kendall tau distance is a metric that counts the number of pairwise disagreements between two lists. The larger the distance, the more dissimilar the two lists are. We use the *normalized Kendall tau distance*[4] between the ground truth and the advertisement rankings predicted by our ranking model.

Discounted Cumulative Gain(DCG) is an increasingly popular metric for evaluating ranked results in information retrieval. Using a graded relevance scale of items in a search engine result set, DCG measures the usefulness, or gain, of a item based on its position in the result list. We use the *normalized Discounted Cumulative Gain*[5] as a measure of average performance of our ranking model for the ad ranking of different keywords.

6.3 Results

The first experiment demonstrates the effectiveness of the ensemble over the expert tips. We carried out three groups of experiments which consider 1) No Tip information, 2) Single

[3]http://svmlight.joachims.org/

[4]http://en.wikipedia.org/wiki/Kendall_tau_distance

[5]http://en.wikipedia.org/wiki/Discounted_cumulative_gain

Table 1: **Effectiveness and Coverage of Expert Tips**

Tip Sets	Tip	Satisfied CTR	Overall CTR	Coverage
Set1	Keyword in abstract	0.083	0.067	36%
	Keyword in URL	0.119	0.072	27.5%
Set2	Contain {you,your}	0.071	0.066	14.3%
Set3	Contain {get,shop,see,find,buy,take}	0.072	0.066	26.8%
	Contain words in action word list	0.070	0.067	49.5%
	First word is {get,shop,see,find,buy,take}	0.073	0.065	14.4%
	First word is in action word list	0.068	0.066	23.1%
Set4	Contain {now,before}	0.078	0.065	7.4%
Set5	Contain {free,zero}	0.079	0.065	15.2%
	Contain price symbol $	0.075	0.062	3.6%
	Contain digital letters	0.070	0.065	21.5%
Set6	Contain {top,best,max,most,latest,newst}	0.079	0.065	7.5%
	Contain words in descriptive word list	0.078	0.066	18.9%
	Contain words in motivating word list	0.078	0.067	27.9%
Set7	Capitalize every word	0.072	0.067	38.6%
	All words are short	0.072	0.067	21.5%

Table 2: Ensemble of Expert Tips

Ensemble	Kendall Tau	NDCG			
		Full	@1	@3	@5
BOW	0.637	0.930	0.725	0.867	0.897
BOW+Single Tip	0.640	0.931	0.727	0.869	0.898
BOW+All Tips	0.665	0.940	0.754	0.883	0.910

Table 3: Different Ensemble Models

Ensemble	Kendall Tau	NDCG			
		Full	@1	@3	@5
Regression Model - SVR					
BOW	0.630	0.928	0.703	0.862	0.892
BOW+All Tips	0.637	0.933	0.720	0.871	0.899
Ranking Model - Ranking SVM					
BOW	0.637	0.930	0.725	0.867	0.897
BOW+All Tips	0.665	0.940	0.754	0.883	0.910

Tip information, 3) All Tips information. The results are shown in Table 2. NDCG@1, NDCD@3 and NDCG@5 are also reported since we are more interested in the quality of the top ranked ads. We use the bag of words (BOW) of the ad content as the baseline feature representation which does not consider the expert tip information. "BOW + Single Tip" refers to the average result that adds only one expert tip to the BOW feature vector. "BOW + All Tips" ensembles all the expert tips with the original BOW features. We can find that "Single Tip" only boosts "BOW" a little, while our ranking based ensemble method outperforms the other two baselines sharply.

The second experiment shows that the ranking model can better capture the ads preference with respect to their corresponding keywords. We compare our ranking model with the regression model. We find that the performance of Support Vector Regression (SVR) is much worse than Ranking SVM. Moreover, if we incorporate all the expert tips using the regression model, the performance improvement is not significant over the BOW baseline. In contract, the ranking model can benefit more from the tips contributed by human experts.

7. CONCLUSION

In this paper, we investigate the problem of query-dependent ranking, i.e. given a keyword, how to choose ads out of a candidate list that will have higher clickthrough rates (CTR). Previous works have attempted to estimate the CTR of ads via a query-independent approach. In this paper, we will propose to select ads that would have higher clickthrough rates via a query-dependent ranking model. We first analyze some manually constructed heuristic rules that could be used to distinguish good ads from bad ones, and we find that the tips provided by human expert are quite effective by low value in coverage. Then we propose to combine these rules into our ranking-based ensemble model to reach our aim. Experiments on real-world data sets have confirmed the effectiveness of our proposed approach. In the future, we will investigate how to design a statistical model to generalize the expert tips to a wider range with the help of historical log data.

8. REFERENCES

[1] R. Jones, B. Rey, O. Madani, and W. Greiner. Generating query substitutions. In *WWW*, pages 387–396, 2006.

[2] N. N. Liu and Q. Yang. Eigenrank: a ranking-oriented approach to collaborative filtering. In *SIGIR*, pages 83–90, 2008.

[3] A. Mehta, A. Saberi, U. V. Vazirani, and V. V. Vazirani. Adwords and generalized online matching. *J. ACM*, 54(5), 2007.

[4] S. Pandey and C. Olston. Handling advertisements of unknown quality in search advertising. In *NIPS*, pages 1065–1072, 2006.

[5] M. Richardson, E. Dominowska, and R. Ragno. Predicting clicks: estimating the click-through rate for new ads. In *WWW*, pages 521–530, 2007.

[6] W. V. Zhang, X. He, B. Rey, and R. Jones. Query rewriting using active learning for sponsored search. In *SIGIR*, pages 853–854, 2007.