Psychological Advertising: Exploring User Psychology for Click Prediction in Sponsored Search

Taifeng Wang
Microsoft Research Asia
taifengw@microsoft.com

Jiang Bian
Microsoft Research Asia
jibian@microsoft.com

Shusen Liu*
South China Univ. of Tech.
liushusen.smart@gmail.com

Yuyu Zhang*
Inst. of Computing Technology
Chinese Academy of Sciences
i@zhangyuyu.com

Tie-Yan Liu
Microsoft Research Asia
tyliu@microsoft.com

ABSTRACT

Precise click prediction is one of the key components in the sponsored search system. Previous studies usually took advantage of two major kinds of information for click prediction, i.e., relevance information representing the similarity between ads and queries and historical click-through information representing users' previous preferences on the ads. These existing works mainly focused on interpreting ad clicks in terms of *what* users seek (i.e., relevance information) and *how* users choose to click (historically clicked-through information). However, few of them attempted to understand *why* users click the ads. In this paper, we aim at answering this "*why*" question. In our opinion, users click those ads that can convince them to take further actions, and the critical factor is if those ads can trigger users' desires in their hearts. Our data analysis on a commercial search engine reveals that specific text patterns, e.g., "official site", "x% off", and "guaranteed return in x days", are very effective in triggering users' desires, and therefore lead to significant differences in terms of click-through rate (CTR). These observations motivate us to systematically model user psychological desire in order for a precise prediction on ad clicks. To this end, we propose modeling user psychological desire in sponsored search according to Maslow's desire theory, which categorizes psychological desire into five levels and each one is represented by a set of textual patterns automatically mined from ad texts. We then construct novel features for both ads and users based on our definition on psychological desire and incorporate them into the learning framework of click prediction. Large scale evaluations on the click-through logs from a commercial search engine demonstrate that this approach can result in significant improvement in terms of click prediction accuracy, for both the ads with rich historical data and those with rare one. Further analysis reveals that specific pattern combinations are especially effective in driving click-through rates, which provides a good guideline for advertisers to improve their ad textual descriptions.

*This work was done when the third and fourth authors were interns in Microsoft Research Asia.

KDD'13, August 11–14, 2013, Chicago, Illinois, USA.

Categories and Subject Descriptors

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

General Terms

Economics, Algorithms, Experimentation

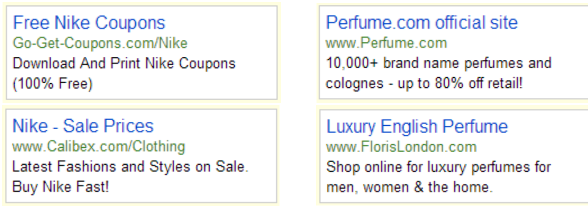
Keywords

Online Advertising; Sponsored Search; Click Prediction; User Psychological Desire

1. INTRODUCTION

As an online advertising system, sponsored search [10] [14] has been one of the most important business models for commercial Web search engines. It generates most of the revenue of search engines by presenting to users sponsored search results, i.e., advertisements (ads), along with organic search results. To deliver the most interesting ads to the users, a sponsored search system consists of a couple of technical components, including query-to-ad matching [1], click prediction for matched ads [8] [11], filtration of the ads according to thresholds for relevance and click probability, and auction to determine the ranking, placement, and pricing of the remaining ads [9]. In today's industry, generalized second price auction (GSP) [9] is the most widely-used auction mechanism, in which the price that an advertiser has to pay depends on the predicted click probability of his/her own ad as well as the bid price and predicted click probability of the ad ranked in the next position. As can be seen from the above descriptions, accurate prediction of click probability is an essential problem in sponsored search, since it is widely used in the filtration, ranking, placement, and pricing of the ads.

State-of-the-art sponsored search systems typically employ a machine learning model to predict the probability that a user clicks an ad. In previous studies [8] [2] [11], the historical click information for each ad is shown to be effective for predicting the future click probability of the ad. In practical sponsored search systems, however, there are many ads without adequate historical click-through data (even after aggregation at different levels, e.g., campaign, advertiser, and query levels). To tackle this data sparsity issue, it is necessary to explore other information than the clicks. Several types of features have been considered for this purpose, which we call *relevance features*. The relevance features are mostly based on the similarity between query and ad, and the quality of the ad [7] [23] [21] [20] [24]. By incorporating these relevance features into the process of click predictions, improved prediction accuracy have been obtained.



(a) "*nike*" (b) "*perfume*"

Figure 1: Example ads for two queries, "*nike*" and "*perfume*", and two ads under the same query yield similar relevance to the query.

Despite the usefulness of the relevance and historical click features, we would like to point out that they are constructed from the perspective of *what* users click and *how* users click. Specifically, relevance information can indicate what relevant content users seek to click. However, as it is well-known that users are not active to search for ads, the search engine, instead, has to recommend ads to users during their generic Web search. Therefore, the relevance between query and ad cannot perform as the key driver for click. Moreover, historical click information implies how users click, unfortunately, this information is not applicable for all of the users due to the data sparsity. In our opinion, in order for more accurate click prediction, we need to examine why users click. Some previous studies [21] [15] have proposed some features to describe the attractiveness of ads. They have made some attempts to touch the question of why users click ads. However, those features are simply arbitrary uni-grams extracted by some heuristics. It is still difficult to explain the reason of users' clicks merely based on those term based signals without systematic knowledge on users' desire behind ad clicks. To this end, we would like to answer the "*why*" question from the perspective of user desire by applying psychological theories.

Two kinds of literatures about psychological research shed light on our analysis of user desire in sponsored search. First, according to literatures on consumer behavior analyses [13] [22] [19], many factors will influence the decision making for consumption, including thought-based effects and feeling-based effects. Thought-based effects are basically win/loss analysis (e.g., trade-off between price and quality); while feeling-based effects are more subjective (e.g., brand loyalty and luxury seeking). Note users clicking the ad usually are with the intention to purchase something. In this situation, it is natural that the factors mentioned in consumer behavior analyses will influence their click behaviors. Moreover, as being widely leveraged by advertisers when designing their ad descriptions to meet different users' needs, psychological theory on human desire, especially Maslow's hierarchy of needs [16], can perform as an important theoretical foundation to model user desire behind ad clicks. Further data analysis, as shown in Section 2, will demonstrates how these factors take effects on users' clicks on ads in sponsored search.

Our data analysis reveals that when different ads yield similar relevance features, there can be big differences in terms of click-through rates. Here we would like to give some intuitive examples. Figure 1(a) shows two ads for the query "*nike*". We can observe that these two ads are both relevant to the query. However, the sponsored search logs show that the first ad gives rise to much more clicks than the second one. In particular, the first ad generates a click-through rate of 0.073 while the second one only results in a click-through rate of 0.005. For another example, Figure 1(b) shows two ads for the query "*perfume*". We can also find that these two ads are both relevant to the query. But the sponsored search logs demonstrate that the first one yields a much higher click-through rate (i.e., 0.167) than the second one (0.005). To understand such differences from the perspective of users, we conduct a user study and ask 100 common web users (diverse in their jobs and ages) to compare the corresponding two ads under each of the queries above, and to tell us why they prefer an ad than the other. Under the study of the query "*nike*", 81% users who attended the study chose the first ad and told us that this is because they would like to take petty advantage by using the "*coupon*" as mentioned in the ad. Under the study of the query "*perfume*", 85% users who attended the study chose the first ad and their major reason was that they feel more trustable on the "*official site*" as mentioned in the ad.

Note that, taking petty advantage and feeling trustable are both desires as mentioned in [13], on the other hand, we can naturally map them into "Physiological" needs and "Safety" needs as introduced in [16]. These findings give us a strong hint that certain textual contents in the ad descriptions can trigger specific psychological desires, and therefore affect the click behaviors of the users. We call such factors for search ad clicks as "*psychological desires factors*". If we can extract effective features representing such "*psychological desire factors*", we should be able to greatly enhance the click prediction accuracy, especially for those ads with inadequate historical click information.

In this paper, we propose to model user psychological desire for both ads and users based on textual patterns, which are mined from ad texts and representative for the psychological desires. In particular, we first take deeper data analysis to verify that the user psychological desire, especially in the form of special textual content, can affect the click-through rates in sponsored search. After that, we propose a data mining based method to automatically extract the textual contents (i.e., n-grams) representative for user desires from the ad descriptions. We then cluster different textual contents into clusters(i.e textual patterns) according to heuristics. To make further generalization on the user desires, we organize all desire patterns into a hierarchy of users psychological desires by following the principle of Maslow's hierarchy of needs [16].

Based on extracted textual patterns and generalized hierarchy of user desires, we extract new features describing ads, users, and their correlation in terms of psychological desire and incorporate these features into the learning framework of click prediction in sponsored search. We have conducted a large scale evaluation on the effectiveness of these new features using the click-through log from a commercial search engine. Experimental results demonstrate that our proposed new features can give rise to significant improvement in terms of accuracy of click prediction. Additional experiments show that our proposed features are especially useful for the ads without rich historical data. Moreover, further analysis illustrates that specific pattern combinations are quite effective for driving up click-through rates, which provides a good reference for advertisers to improve the textual descriptions of their ads..

To sum up, the contributions of our work include:

- A comprehensive data analysis illustrating that user psychological desires can play an important role in click prediction in sponsored search.
- An effective data mining based method to automatically extract textual patterns that can attract user psychological desires.
- Introduction of new features about user psychological desire into the learning framework of click prediction in sponsored search.

The remainder parts of this paper are organized as follows. In Section 2, we present our data analysis on how user psychological desires affect the CTR in sponsored search. Section 3 introduces our proposed approach for extracting textual patterns representing user desires. The click prediction modeling will be presented in

Section 4, in which we will propose to incorporate extracted patterns as new features into the whole modeling. Experimental setup and results are presented in Section 5. At last, we conclude the paper and discuss the future work in Section 6.

2. DATA ANALYSIS ON USER PSYCHOLOGICAL DESIRES

To gain more understanding on why user psychological desire is important to click prediction, in this section, we will take discussions on the user decision making process and its relationship with sponsored search. Based on some specific examples, we will demonstrate how user psychological desires are usually reflected in the ad content. After that, by conducting some data analyses, we try to understand how those user psychological desires affect click behaviors. As all of discussions in this section will be provided based on data analysis on real sponsored search data, we first introduce the settings for our data analysis.

2.1 Data Settings

All of our data used in this work are collected from a commercial search engine. In this section, we will use two datasets for the following purposes:

Dataset 1: In order to study whether the decision making only relies on relevance, we randomly sample a set of queries with all the ads displayed under them from the entire dataset of June, 2012. Then, we compute both CTR and the relevance score of each $\langle$query, ad$\rangle$ pair. To get a reliable CTR, we filter those $\langle$query, ad$\rangle$ pairs with fewer than 100 impressions. Finally, we collect $10,786$ queries and $44,079$ $\langle$query, ad$\rangle$ pairs.

Dataset 2: In order to explore the user desires in the ad, we sample about 20M ad impressions from those occurred in the log of June, 2012. We compute CTR for each ad in the set based on the click-through log. In Section 3, this dataset will be used again as source to mine out those patterns. Finally, we collect $5,296,935$ unique ads, generated by $171,495$ advertisers.

2.2 Consumer Decision Making Process

In sponsored search, once a user clicks the returned ads, the natural goal of him/her would be conversion. Thus, ad clicking becomes a critical step on the consuming behavior chain since it pre-selects what will be potentially consumed. To predict users' click behaviors on ads, it is necessary to investigate what people will consider during their consuming behaviors.

As a widely adopted course on consumer behavior, authors of the book [13] formulate the process of making consumption decisions into four stages: problem recognition, information search, decision making, and post-purchase evaluation. Through the first two stages, consumers will identify candidate consumption, while in the most important stage, i.e., decision making, they will take thorough consideration and decide whether to complete the consumption. According to [13], the decision making is affected by three effects: contextual, thought-based, and feeling-based effects. In particular, contextual effects correspond to environmental effects which are usually treated as the background of consumption; thought-based effects, such as pricing discount, deliver time limitation, etc., are more quantifiable; feeling-based effects, such as brand preference, trustworthiness, luxury seeking, etc., are more related to subjective. Among these three types of effects, thought-based and feeling-based effects play as the more critical factors in decision making and form up consumer psychological desires. Moreover, since consumer psychological desires are diverse, it is naturally to organize them into a hierarchy to model consumer behaviors in a more effective way. As being widely leveraged by advertisers when designing their ad descriptions to meet consumers' needs, psychological theory on human desire, especially Maslow's hierarchy of needs [16], can perform as an important theoretical foundation to model and organize user desires behind ad clicks.

In sponsored search, ad clicking can be viewed as selecting potential consumption. Thus, user psychological desires obviously play important roles in deciding the click behaviors. Unfortunately, most of state-of-the-art works rely on relevance and historical click features. Although historical click information implies *how* users click, unfortunately, this information is not applicable for all of the users due to the data sparsity. Moreover, relevance can indicate *what* content users seek, but it is just a preliminary condition to find candidate consumption and cannot answer *why* users click according to the decision making process. It's highly possible that some ads, though with high relevance to the query, could not fit on the user desires. To verify that user desires do affect ad clicks beyond relevance, we conducted a data analysis to investigate that whether ads with the same relevance level under one query will get quite different CTR values [1].

Here, the CTR is computed based on the click-through log in Dataset 1. To measure the relevance between the query and ad, we first build a relevance model via learning-to-rank approach. In particular, for each $\langle$query, ad$\rangle$ pair, we generate a vector of relevance features, such as BM25, language modeling, category matching etc, and each pair is associated with a human judged label representing the degrees of relevance of the ad with respect to the query. There are five levels of relevance: perfect, excellent, good, fair, and bad. We apply a widely-used learning-to-rank algorithm, i.e., RankSVM, to learn the relevance model. All parameters are tuned based on 5-fold cross-validation.

After obtaining the CTR for each ad and the relevance score for each $\langle$query, ad$\rangle$ pair, we conduct further data analysis on Dataset 1. Particularly, among the ads under one specific query, we find those ad pairs satisfying that the relevance scores of both ads yield small difference (in practice, we normalize the relevance scores into the scale of [0, 1], and we judge the normalized relevance scores differing less than 0.01 as small difference.) Then, we checked the CTR difference of ad pairs with small relevance difference. Figure 2 demonstrates the distribution of relative CTR difference among all the ad pairs with similar relevance. The relative CTR difference between two ads, a_i and a_j, is computed as $\frac{|\mathrm{CTR}_{a_i} - \mathrm{CTR}_{a_j}|}{\max(\mathrm{CTR}_{a_i}, \mathrm{CTR}_{a_j})}$. From this figure, we can find that about 75% of ad pairs yield more than 20% CTR difference between two ads. This analysis result indicates that relevance is not adequate to provide accurate click prediction. According to the aforementioned decision making process, it becomes necessary to explore the other complimentary factors, i.e., user psychology desire, to achieve better click prediction.

2.3 Effects of User Psychological Desire

In this subsection, we examine the effects of user psychology desires on click behaviors in sponsored search. Actually, advertisers are profit-seekers and really pioneers in employing such user desires. After browsing a sample of ads from the commercial search engine, we find that there are a number of textual contents that are often used by advertisers in the ad text. Table 1 illustrates some examples of the user desires under thought-based and feeling-based effects, which are extracted from ads by human experts in this field. From this table, it is obvious that advertisers do put specific textual contents in ad texts to target user psychological desires. In the

[1]A previous study [4] reveals that display positions will greatly influence the click probability of ads. In order for accurate comparison, all the CTR values used in our work are normalized based on ads' display positions. The normalization coefficients are obtained based on the analysis from a random online flight in the commercial search engine.

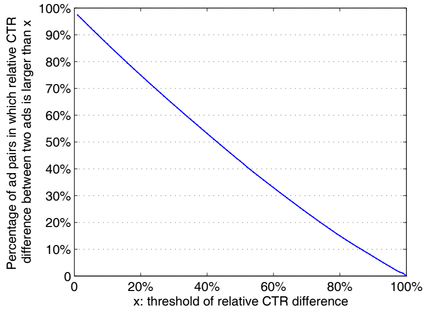


Figure 2: Distribution of CTR difference among all the ad pairs with similar relevance.

following of this section, we study some examples of these contents and investigate how they can affect the CTR of ads. This part of analysis is conducted based on Dataset 2 as mentioned in Section 2.1.

Table 1: Examples of the user desires under thought-based and feeling-based effects.

Effects	User Desires	Examples
Thought-based	Petty advantage	*big discount, good deal, coupon*
	Quantity/quality advantage	*popular brand, large selection space*
	Extra convenience	*quick delivering, flexible payment method*
Feeling-based	Trustworthy	*official seller, service with guarantee*
	Brand loyalty	*Ebay, Amazon*
	Luxury seeking	*first class brand*

One example of thought-based effects that affects user decision is petty advantage (in other words price temptation), which is a widely-observed textual content in the ads. In particular, to let more users pay attention to the specific products or services, some advertisers usually emphasize discounts or sales information in the text of their ads. Table 2 shows some examples of ads aiming at users psychological desire of taking petty advantages which have been found in a large number of ads. From this figure, we can find that "*coupon*" and "*x% off*" are two textual patterns which can be instantiated by various text content. Taking "*coupon*" as example, there exist at least the following variations: "*coupon code*", "*online coupon*", "*free coupons*", etc. To investigate if these textual patterns will affect the users' clicks, for each of these two patterns, we separately compute the CTR of those ads containing the pattern compared with the average CTR of all ads, the results of which are shown in Table 3. From this table, we can find that the ads matched with either the pattern "*coupon*" or "*x% off*" yield CTR which is significant larger than the average CTR over all ads. Besides of average CTR, Figure 3(a) compares the CTR distribution of ads matched with pattern "*coupon*" with those without this pattern. From the figure, we can observe that the pattern "*coupon*" can cause a big difference in the shape of CTR distribution as well. We can find the similar observation for the pattern "*x% off*", as shown in Figure 3(b).

Toward the feeling-based effects, we also find some textual content in ads which are often employed by advertisers to target user

Table 2: Examples of ads catching the user desires of taking petty advantages.

Text pattern	Ad's text
coupon	Printable **Free Coupons** - Get Free November 2012 Printable
	Verified **Coupon Codes** w/ 100% Success Rate Up To 30% Off + Hot Deal. Save Now!
	The World's #1 **Online Coupon** Site. Check Out Nov. 2012 Coupons!
x% off	Cheap Flights Tickets Sale. Book Now & Get an Instant **$15 Off**.
	Be Hurry or Be Sorry, Sale Ends Soon. **Up to 60% Off**. Book Flights Now!
	Get 50% off the new 2012 Titanium Security. Top rated antivirus.

Table 3: CTR difference between the ads matched with the certain desire pattern and overall ads.

Desire pattern	Percentage of matched ads	CTR change
"*coupon*"	2.2%	+47.5%
"*x% off*"	4.1%	+19.7%
"*official*"	2.6%	+25.0%
"*return guarantee*"	1.9%	+31.4%

psychological desire. Taking the trustworthy as example, "*official*" and "*return guarantees*" are often utilized to attract users' attentions in this category. Table 4 shows some related examples. There are also quite a few variations in textual content to express "*official*" and "*return guarantees*". Taking "*return guarantees*" as example, we can find "*365 Day Returns*", "*30 Day Right of Return*", "*No-Hassle Returns*", etc. And after conducting similar CTR analyses, "*official*" and "*return guarantees*" yield significantly larger average CTR compared to that of all ads, as shown in Table 3. Significant changes in CTR distribution of ads matched with pattern "*official*" or "*return guarantees*" can also be observed as shown in Figure 3(c) and 3(d).

Table 4: Examples of ads catching the user desires of trustworthy.

Text pattern	Ad's text
official	Shop At Lowe's **Official Site**!
	(**Official**) Online Number Bed Sale!. Save Up To $1900
	Official Alaska Marine Highway, Reservations & Specials
return guarantee	Semiconductor processing equipment Large stock. **30 Day Right of Return**.
	Free Shipping Both Ways on Puma. **365 Day Returns**.
	Up to 80% Off. Free Shipping. **No-Hassle Returns**!

Based on all these studies, we can find that user psychological desire has been well utilized by advertisers to lead users to click more on their ads. Therefore, it has become necessary to leverage user desire information to improve the accuracy of click prediction, since those textual patterns catching user psychological desire can drive up the CTR significantly. To take advantage of this important information, it is essential to find an effective way to extract those patterns from ads. In the next section, we will propose a data mining method to automatically extract those patterns from ads according to the heuristics get from data analysis.

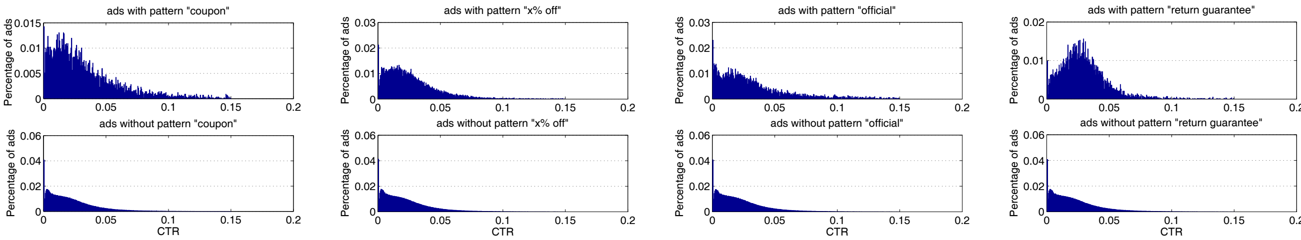
3. DISCOVERING USERS PSYCHOLOGICAL DESIRE FROM ADS

To extract content reflecting user desires from ad texts, we need to first understand the textual content of ads. Usually, there are two parts of content in one ad's text. One targets at indicating the relevance between the query and the ad; while the other part, according to definitions in Section 2.2, contains the content reflecting the user psychological desires. Therefore, our target is to mine out the textual content reflecting user psychological desires from the part not indicating relevance. To this end, our first step is to filter out the relevance part.

After extracting text content reflecting user psychological desires, we will summarize the corresponding desires of these text contents from the nature of their functionality. And, we will follow 3 principles:

• The text content should cover enough volume in real ad traffic, as quite a few experienced advertisers and even ad agents are very likely to add similar content to trigger user desires. In fact, we can find content like "*get a free coupon*" covers over 10K ads in our **Dataset 2**.

• Similar content can reflect the specific same desire. As ad is formed as free text content, advertisers have many choices to express the same meaning. For example, as to the coupon, there can be "*coupon code*", "*get coupon*", "*free coupons*" . These variants



(a) "*coupon*" (b) "*x% off*" (c) "*official*" (d) "*return guarantee*"

Figure 3: Comparison in terms of CTR distribution between ads with and without certain text patterns.

actually share the same text phrase. Therefore, many text contents reflecting a specific same desire can be organized in a cluster of text phrases. To easiness of reference, we call one cluster a **user desire pattern**.

• Content from experienced advertisers are more important. Experienced advertisers are more likely to put text content representing desires in their ads since they are more familiar with user behaviors than new advertisers. Moreover, they might even actively create novel text phrases to represent user psychological desires. Thus, text content from these advertisers will be highly possible to form useful desire patterns.

Accordingly, we propose a three-step approach to automatically extract user desire pattern from ad texts. In the following of this section, we will introduce it step by step, followed by a demonstration of the mining results with further human validation.

3.1 Mining User Desire Patterns

Step 1: Cleaning up content targeting for relevance
Based on the definition of user psychological desires, the content reflecting such desires will not directly contribute to the relevance between user's search query and ads. In some cases, e.g., the search query is "cheap car", some words in query, e.g., "cheap", might contribute to both relevance and user psychological desires. But, it is obvious that the noun word "car" should not be related to such desires. Hence, in our approach, we first filter out the noun words matched with search queries from ads, which can greatly reduce the content in ad texts.

Step 2: Finding n-grams with high frequency
Ads content contains two major parts, one targeting for relevance, while the other aiming at convincing users to consume. According to our definition, text patterns related to user psychological desire will be mined from the second part. After step 1, we have roughly removed contents targeting for relevance. To further discover effective patterns for user desires, the most straight forward way is to find out n-grams from ads corpus as the candidate contents to form patterns. Note that each useful pattern should not be too long, otherwise it can be split into several shorter patterns. Therefore, we limited each n-gram to be at most of 6-gram. Then, we scan all ads from our collected data and extract all the n-grams. According to the first principle discussed above, useful n-gram should cover enough volume. Hence, we only keep those patterns that can cover 1000 ads and 100 advertisers. These parameters are tuned manually via cross validation in a sampled data set.

Step 3: Pattern generalization via clustering
After first two steps, there are still a large number of extracted n-grams. And it is uneasy to judge whether these n-grams are really related to user desires. Considering the hints from the second and third principles, we leverage a dedicatedly designed clustering algorithm, as shown in Algorithm 1, to further process remained n-grams towards a better representation of user desires. The intuition of the clustering method lies: 1) Those n-grams sharing some words are usually variations of the same user desires; thus, textual similarity based clustering method will tend to group those variations into one cluster, named as a textual pattern for user desire. 2) Those advertisers with more experience will tend to incorporate user desire patterns in their ad texts; thus, we can put more weight on n-grams extracted from their ads when computing the clustering center. In this work, we describe the mature status of an advertiser as the number of clicks targeting at any of his/her ads in one month, and the weight of an n-gram is accordingly set as the maximum mature status among advertisers who ever used the n-gram in their ads.

In the clustering process, we need to determine the cluster number K, i.e., the number of textual patterns for user desires. Since we are aiming at enhancing click prediction, we finally set the $K = 300$ based on cross validate on the final click prediction performance.

Algorithm 1 Clustering Algorithm for User Desires Generalization

Input: $\{P_1, P_2, \cdots, P_N\}$: N text n-grams;
$\{\omega_1, \omega_2, \cdots, \omega_N\}$: weights of N n-grams obtained based on mature status of advertisers who used this n-gram;
K: the number of clusters;

Output: K clusters, each of which represent one generalized user desire;
$\{C_1, C_2, \cdots, C_N\}$: $C_i \in \{1, \cdots, K\}$ denotes which cluster P_i belongs to;

Algorithm:
It's basically a k-means framework:

1 Randomly select K n-gram as seeds: $S_1, \cdots, S_K$;
2 Cluster the n-grams based on similarity defined by distance metrics:
$C_i = \arg\min_k \text{Distance}\langle P_i, S_k\rangle, i \in 1, \cdots, N$
3 Update the cluster center according to center update method:
$S_i^{\text{new}} = \frac{\sum_{\forall C_j = i} \omega_j \cdot P_j}{\sum_{\forall C_j = i} \omega_j}$
4 go to 2 and loop 2,3 until the cluster centers converge

3.2 Hierarchy of User Psychological Desire

By using our clustering algorithm, we are able to extract a set of general user desires. To reduce the sparsity of user desires for each individual ad, we further organize extracted general desires into a hierarchy of user psychological desires according to Maslow's hierarchy of needs [16]. In particular, Maslow's hierarchy of needs define humans' needs as five levels: *Physiological, Safety, Belongingness, Esteem,* and *Self-Actualization*. We can directly map user psychological desires into these five levels, and each level is specified with a set of textual patterns, as shown in Figure 4. From this hierarchy, we can find that the lower level represents more basic

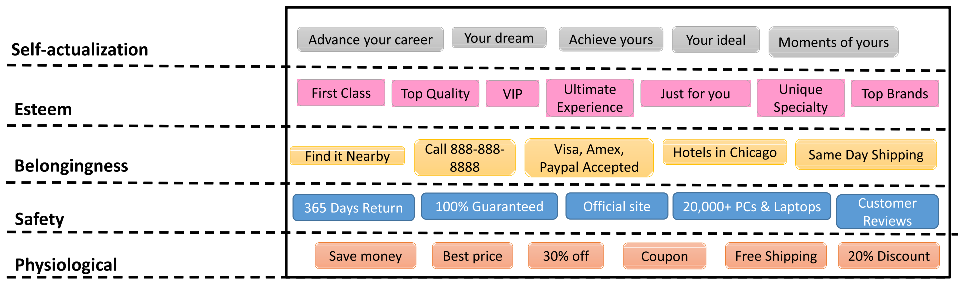


Figure 4: A Hierarchy of User Psychological Desires in Sponsored Search.

user psychological desires while the upper level represents higher level of user desires.

4. CLICK PREDICTION MODELING

As we have discussed in the data analysis part, textual patterns in ads which reflect user desire patterns will greatly drive the CTR. So after extracting textual patterns for user psychological desire, we are going to integrate such information into the click prediction modeling. In this section, we will first briefly outline our approach for click prediction. Then, we will discuss how to model the user psychological desire as new features for both ads and users into click prediction modeling.

4.1 Maximum-Entropy Modeling

We formulate click prediction in sponsored search as a supervised learning problem. In this paper, we will apply the maximum entropy model [3] for click prediction. We collected both click and non-click events from sponsored search logs as training samples, where each sample represents a ⟨query, ad, user, position⟩ tuple, representing that the ad was presented to the user at the certain position after she submits the query to search engine. Assume there is a set of N training samples,

$$\mathbb{D}_{\text{train}} = \{\langle \mathbf{f}(q_i, a_i, u_i, p_i), c_i \rangle\}$$

where $\mathbf{f}(q_i, a_i, u_i, p_i) \in \mathbf{R}^d$ represents the d-dimensional feature space for the i-th tuple and $c_i \in \{0, 1\}$ denotes the corresponding class label, i.e., 1 for click while 0 for non-click.

Given a query q, an ad a, a user u, and ad's displayed position p, the problem is to compute the probability of click $p(c|q, a, u, p)$. The maximum entropy model [3] is well suited for this task since its strength in combining diverse forms of contextual information, and formulates the click probability for a ⟨query, ad, user, position⟩ tuple as follows:

$$p(c|q, a, u, p) = \frac{1}{1 + \exp(\sum_{j=1}^{d} \omega_j f_j(q, a, u, p))}$$

where $f_j(q, a, u, p)$ is the j-th feature derived for ⟨query, ad, user, position⟩ tuple and $\omega_j \in \mathbf{w}$ is the associated weight. Given the training set $\mathbb{D}_{\text{train}}$, the maximum entropy model learns the weight vector $\mathbf{w}$ by maximizing the likelihood of exponential models as:

$$\mathbf{w} = \arg\max_{w}(\sum_{i=1}^{n} \log(p(c_i|q_i, a_i, u_i, p_i)) + \log(p(\mathbf{w})))$$

where the first part represents the likelihood function and the second part utilizes a Gaussian prior on the weight vector $\mathbf{w}$ to smooth the maximum entropy model. There are many approaches available in the literature [17] to solve this kind of optimization problems including iterative scaling and its variants, quasi-Newton algorithms, and conjugate gradient ascent. Given the large collection of samples and high dimensional feature space, we use a nonlinear conjugate gradient algorithm [18].

An accurate maximum entropy model relies greatly on the design of features. According to the state-of-the-art works in click prediction, there are two major kinds of features, which are relevance features and historical click features. In this work, we use some representative features according to previous work [5] [21]:

- For relevance features we employed edit distance of ad and query, edit distance of ad and bid keyword, cosine similarity between ad and query, the category matching between ad and query, etc.
- For historical features we employed history COEC (position normalized CTR) for ⟨query, ad⟩ pair, query, and ad, respectively, smoothed COEC according to query term, ad term, etc.

4.2 Integrating User Psychological Desires into Click Prediction

Beyond all the features described above, we aim at incorporating user desire information as new features into the click prediction modeling, since those user desires can result in more influence on users' click behaviors.

4.2.1 Modeling Psychological Desire as Ad Features

As we have mined a set of n-gram clusters as textual patterns, which are representative for user desires, we are able to match each ad against these textual patterns, so as to produce a series of binary features, each of which indicates the existence of a certain desire in the particular ad's text. Moreover, since we have generalized five levels of user desires to reduce the feature sparsity, it is possible to match each ad against these desire levels, so as to generate another five binary features, which imply the belongingness of an ad to the certain desire level. Specifically, these binary features are determined as follows:

• ***Ad desire pattern features***: For each ad a, we will check if a is matched with each of textual patterns by checking the existence of any n-gram belonged to this textual pattern P. If a is matched with one desire, i.e., it contains a specific n-gram belonged to the certain desire pattern P, the corresponding feature $\mathcal{D}_a(P)$ will be set as 1, otherwise, it will be set as 0.

• ***Ad desire level features***: For each ad a, we will check if a is matched with textual patterns belonging to each of desire levels. If a is matched with one desire level L, i.e., it contains a desire pattern that is included in the specific desire level, the corresponding feature value $\mathcal{D}_a(L)$ will be set as 1, otherwise, it will be set as 0.

These binary values will be directly used as binary feature in the maximum entropy model for click prediction. As describe above, we have generated 300 textual patterns and generalized them into 5 levels. Therefore, for each ad a, we will employ 300 desire pattern features as well as 5 desire level features to represent ad psychological desire in the click prediction modeling.

4.2.2 Modeling Psychological Desire as User Features

Our mined psychological desire patterns can also be leveraged to represent each user's interests from the perspective of psychological desire. Intuitively, if the user tends to click ads containing a certain psychological desire pattern frequently, it is very likely that this user has a strong demand on the corresponding psychological desire. Therefore, we could determine user features representative for the user's demand on each desire pattern or desire level as follows:

• ***User desire pattern features***: To describe a user's demand on a specific desire pattern, we take advantage of position normalized CTR of this user on all the ads containing the specific desire pattern. Particularly, for a user $\hat{u}$ and a desire pattern P, $\hat{u}$'s demand on P is computed as:

$$\mathcal{D}_{\hat{u}}(P) = \frac{\sum_{\langle q,a,u,p,c\rangle} \mathbf{I}(P \in a \wedge u = \hat{u} \wedge c = 1)\phi(p)}{\sum_{\langle q,a,u,p,c\rangle} \mathbf{I}(P \in a \wedge u = \hat{u})}$$

where $\mathbf{I}(\cdot)$ denotes an indicator function; and, $\phi(p)$ represents the position normalized coefficient, which gives larger weight to the click happened at lower position. [2]

• ***User desire level features***: To describe a user's demand on a specific desire level, we take advantage of the CTR of this user on all the ads containing any of desire patterns belonging to this specific desire level. Particularly, for a user $\hat{u}$ and a desire level L, $\hat{u}$'s demand on L is computed as:

$$\mathcal{D}_{\hat{u}}(L) = \frac{\sum_{\langle q,a,u,p,c\rangle} \mathbf{I}(L \cap a \neq \emptyset \wedge u = \hat{u} \wedge c = 1)\phi(p)}{\sum_{\langle q,a,u,p,c\rangle} \mathbf{I}(L \cap a \neq \emptyset \wedge u = \hat{u})}$$

As describe above, for each user, we can obtain 300 user desire pattern features corresponding to each desire pattern and 5 user desire level features for each desire level. Consequently, we will integrate these new features representative for user psychological desire into the click prediction modeling.

4.2.3 Modeling Desire Matching Between Users and Ads

After extracting psychological desire features for both ads and users, we are able to generate features to describe desire matching between users and ads.

• ***Desire pattern matching features***: After representing an ad a as a vector of ad desire pattern features, i.e., $\mathbf{D}_a^p = \langle \mathcal{D}_a(P_1), \mathcal{D}_a(P_2), \cdots, \mathcal{D}_a(P_{300})\rangle$, and representing a user u as a vector of user desire pattern features i.e., $\mathbf{D}_u^p = \langle \mathcal{D}_u(P_1), \mathcal{D}_u(P_2), \cdots, \mathcal{D}_u(P_{300})\rangle$, we could compute the desire pattern matching features between a and u based on the similarity between $\mathbf{D}_a^p$ and $\mathbf{D}_a^p$.

• ***Desire level matching features***: After representing an ad a as a vector of ad desire level features, i.e., $\mathbf{D}_a^l = \langle \mathcal{D}_a(L_1), \mathcal{D}_a(L_2), \cdots, \mathcal{D}_a(L_5)\rangle$, and representing a user u as a vector of user desire pattern features i.e., $\mathbf{D}_u^l = \langle \mathcal{D}_u(L_1), \mathcal{D}_u(L_2), \cdots, \mathcal{D}_u(L_5)\rangle$, we could compute the desire level matching features between a and u based on the similarity between $\mathbf{D}_a^l$ and $\mathbf{D}_a^l$. Note that, our desire pattern and desire level matching features are quite general in that we could apply any similarity function when compute these two kinds of features. In this paper, we will apply cosine similarity to compute the desire pattern and desire level matching features.

[2]Similar to [4], we obtain $\phi(p)$ based on the analysis from a random online flight in the commercial search engine.

5. EXPERIMENTS

In this section, we first describe the settings of our experiments and then report the experimental results.

5.1 Experimental Settings

5.1.1 Data set

To validate whether user psychological desire features we mined out can really help enhance the click prediction accuracy, we conduct experiments based on the click-through logs of a real world commercial search engine. In particular, we collect the whole click-through logs of a two-week period from this search engine as our experimental dataset. And, we randomly sample a set of query events from the original whole traffic. We finally collect about 20M ad impressions in each of these two weeks. After that, we divide this dataset into two parts, each containing the data of one week. Then, we use the first week's data to train the click prediction model, and use the second for testing. Detailed statistics of the dataset can be found in Table 5.

Table 5: Statistics of the datasets for training and testing the click prediction model.

	ad impressions	unique ad	unique query
Training	20, 835, 369	4, 251, 061	2, 569, 386
Testing	19, 812, 476	5, 311, 800	2, 533, 796

5.1.2 Compared Methods

As mentioned in Section 4, we employ maximum entropy modeling to train the click prediction model. In our experiments, we will compare the performance of different click prediction models trained with different feature sets. In order to show the effectiveness of those desire features, we employed the following feature settings: (details about the feature sets can be found in Section 4)

- **HF**: only uses historical click features.
- **HF-RF**: uses historical click features and relevance features.
- **HF-DPF**: uses historical click features and desire pattern features.
- **HF-DPLF**: uses historical click features and both desire pattern and desire level features.
- **HF-RF-DPF**: uses historical click features, relevance features, and desire pattern features.
- **HF-RF-DPLF**: uses historical click features, relevance features, and both desire pattern and desire level features.

We set HF and HF-RF as a baseline because previous studies [6] [12] have demonstrated that the historical click features and relevance features play the most important role in the click prediction task. Further experiments compare the performance of HF and HF-DPF/HF-DPLF to examine whether the proposed user psychological desire features can benefit click prediction beyond historical features. Comparison between HF-RF and HF-DPF/HF-DPLF will provide us with more understanding on the predicting power of relevance features and user desire features, respectively. Experiments on HF-RF-DPF/HF-RF-DPLF aim at recognizing the contributions of user desire features to click prediction beyond both historical click features and relevance features. Moreover, we compare the performance between HF-RF-DPF and HF-RF-DPLF to investigate if desire level features are good complement to desire pattern features.

5.1.3 Evaluation Metrics

In our work, the Maximum Entropy modeling is applied to predict click probability for every ad impression. We use recorded user actions, i.e., click or non-click, in the log data as labels. To evaluate the overall performance for the model, we employ average Relative

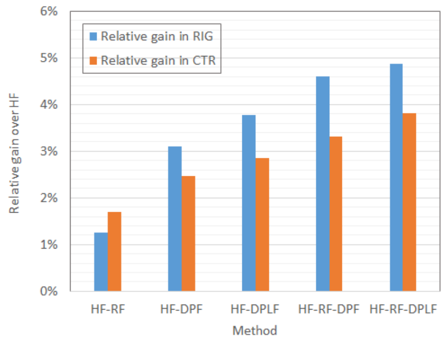


Figure 5: Relative gain of different methods over HF in terms of CTR and RIG.

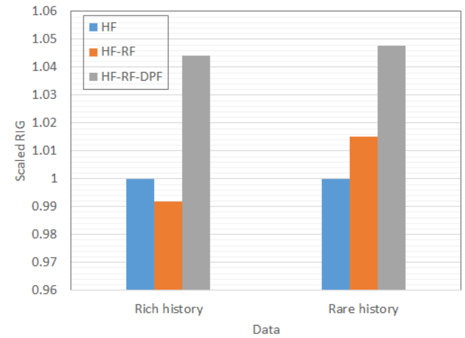


Figure 6: Relative gain of different methods over HF in terms of RIG, conducted on ads with rich or rare history.

Information Gain(RIG) [11] as the evaluation metric. Moreover, we employ CTR as another evaluation metric. Since it is a little difficult to run real online experiment, we apply a replay-based simulation method for evaluation. In particular, for each query event in the log data, we re-rank ads list according to our new click prediction model and use the real clicks as ground-truth to compute the CTR for the specific model.

5.2 Experimental Result

5.2.1 Overall Performance

Figure 5 demonstrates the relative gain of HF-RF, HF-DPF/HF-DPLF, and HF-RF-DPF/HF-RF-DPLF over HF in terms of CTR and RIG. From this figure, we can find that user desire patterns can lead to significant improvement on click prediction over the baseline method HF. In particular, in terms of RIG, there are about 3% relative improvement by using both desire features and historical click features over only using historical click features, while there are only about 1.2% relative improvement by using both relevance features and historical click features over historical click features only. And, there is also more than 3% improvement by using HF-RF-DPF or HF-RF-DPLF over HF-RF.

Figure 5 also reports the comparison between different models in terms of CTR. From the figure, we can found there are about 0.7% CTR improvement by HF-DPF over HF-RF while there are about 1.1% improvement by HF-DPLF over HF-RF. Moreover, HF-RF-DPF generates a relatively 1.6% CTR improvement by using both desire features and relevance features over HF-RF; meanwhile, HF-RF-DPLF, after adding desire level feature, gives rise to more CTR improvement, i.e., relatively 2.1%, over HF-RF. These results imply a big impact of desire features on click prediction accuracy improvement. Furthermore, the results showing that HF-RF-DPLF/HF-DPLF outperform HF-RF-DPF/HF-DPF, respectively, also indicates that desire level features are good complements to desire pattern features. We hypothesize the reason is that desire level features can reduce the sparsity of desire features for individual ads.

Actually, in real sponsor search system, increasing 1% on the click-through rate is already a big improvement. According to [9], 1% ctr improvement will drive additional hundreds of million revenue per month. In this sense, 2.1% relative improvement is really significant in click prediction.

5.2.2 Impacts on Ads with Rich v.s. Rare History

Click prediction task typically faces two kind of data: ads with rich history and ads with rare history. Usually, when an ad is with rich history, its click prediction can achieve good performance by referring to its historical CTR; while relevance features are often used to help the click prediction especially for those ads with rare historical information.

In this experiment, we would like to examine the impacts of new desire features with regarding to these two cases, respectively. In particular, we first separate all ⟨query, ad⟩ pairs into two subsets: *rich history set* includes all ⟨query, ad⟩ pairs with more than 70 historical impressions in one month; and, *rare history set* contains all the other ⟨query, ad⟩ pairs. Figure 6 reports the scaled RIG by three methods on these two sets. From this figure, we can find that, by adding desire features, HF-RF-DPF can increase the RIG performance over HF-RF by 5.2% for *rich history set* and 3.2% for *rare history set*.

For those ads in the *rare history set*, click prediction is mainly based on the understanding on users' click intents. Although relevance features can describe textual similarity between the query and the ad, it may not indicate if the user consider the ad as a potential fit for consuming the product associated with the ad. Nevertheless, user psychological desires can better reflect this kind of user desires. Therefore, our extracted textual desire patterns can effectively predict users' clicks on those ads with rare history.

For those ads in the *rich history set*, Figure 6 illustrates that HF-RF results in a decreasing RIG compared HF, which indicates that relevance features fail to help click prediction when historical information is rich. On the contrary, we can find user desire features can help to further increase the accuracy in this part by more than 4% with respect to HF. Basically, the ad's textual content is usually quite stable along the history. It is highly possible that the CTR already encoded the relevance features since intuitively users do not click irrelevant ads. Therefore, when historical data is rich, we can directly predict the CTR according to history while ignoring the relevance features. However, those user desire related patterns might change rapidly since advertisers will adjust ad text slightly for promoting specific features according to their knowledge. Even the ad was just created one month, some user desire related patterns might already change multiple time in the period. If we consider the effect of the current pattern, it is straightforward that the CTR of the ad will be predicted more accurately.

5.2.3 Effects of User Desires on Different Ads Categories

Intuitively, we hypothesize that the user psychological desires do influence the CTR. In particular, if the ad description can fit user's desire well, the corresponding CTR will be driven up. Thus, user desire features can lead to more accurate click prediction. It is natural that specific desires will work differently in different categories of ads. In this experiment, we employ the widely-used text categorization, i.e., ODP [3]. And, we apply a basic text classification model to automatically categorize ads into different concept categories. Then, we study several combinations between desire patterns and ads categories to verify the click prediction accuracy lift caused by new desire features. We listed some observations in Table 6, an interesting finding from which is: When Physiological

[3]http://www.dmoz.org/

level patterns matched in ads related to jewelry, the click prediction accuracy will be decreased slightly by 1.58%. While when such level patterns matches in ads related to travel and hotel, the click prediction accuracy will be increased largely (3.60%).

Table 6: RIG improvement of example combinations between Physiological level desire patterns and diverse ads categories.

Ads category	Patterns	RIG improvement
Music	Free, Official site	4.11%
Clothing	x% Off, Official site, save x	3.29%
Travel	Book Now, great deal	3.60%
Jewelry	x% off, Free shipping	−1.58%

5.2.4 *Effects of Combinations over Desire Patterns*

According to our discussion in Section 3, desire patterns can be further organized into 5-level hierarchy, including *Physiological, Safety, Belongingness, Esteem,* and *Self-Actualization*. Usually, successful advertisers will combine desire patterns from different levels together to achieve a higher CTR. In this section, we would like to check what kind of combinations between desire levels will provide better impacts on users click behavior.

In our experiment, we first select the top 500 effective pattern combinations according to their corresponding lift to prediction accuracy, then we match the detailed patterns in these combinations into desire levels. After that, we can get a set of combination of general desire levels which is effective in enhancing the click prediction accuracy. We listed the hottest five combinations of desire levels in Table 7.

Table 7: Hottest combination of general desires.

Self-Actualization + Physiological
Safety + Self-Actualization + Physiological
Belongingness + Self-Actualization
Safety + Physiological
Belongingness + Physiological

Generally, advertisers might take advantage of these combinations to enhance their ads copy so as to achieve better CTR. However, the second order effect does exist in the economic world, which indicates that if every one follows the same golden rule, the advantage will be vanished. Hence, it becomes necessary to address this potential issue in optimizing the ads copy with those patterns in identical way. Fortunately, on the other hand, it's not that hopeless since these combinations are actually conceptual and the detail patterns might be created by advertisers actively along the time. And that's the reason for us to use data mining algorithms to get those patterns automatically. As long as we can periodically get the latest patterns, we can help the click prediction effectively.

6. CONCLUSION AND FUTURE WORK

Advertising by natural focuses on commercial values, and it is indeed out of the scope of information retrieval. This paper takes an earlier attempt to connect click prediction in sponsored search with user behavior analysis. And, our research explores a new way for computational advertising to embrace the traditional psychological analysis to enhance the computational advertising through its real nature. In particular, we aim at answering "*why*" users click search ads by exploring user psychological desire according to consumer behavior analysis and Maslow's desire theory. We construct novel features for both ads and users based on our definition on psychological desire and incorporate them into the learning framework of click prediction. Large scale evaluations demonstrate that it can significantly increase the accuracy of click prediction by incorporating mined desire features into the learning framework of click prediction.

Leveraging psychology knowledge for improving online advertising, especially computational advertising, is still at early stage. But, it is indeed valuable to incorporate these cross discipline knowledge to push the boundary of the ads research. We will keep investigating on this direction. In details, 1) we will study if users' psychological desire is dependent with queries or other kinds of search context and study how to model context-aware users' desire. 2) As Maslow's theory mentioned, desires can be organized into a hierarchy. We will examine whether this hierarchical relationship can be leveraged when we are going to match users' and ads'desire. 3) As users' desire may change along the time, we plan to study how to model users' temporal psychological desire and detect their emerging interests in terms of desire at real-time.

7. REFERENCES

[1] V. Abhishek and K. Hosanagar. Keyword generation for search engine advertising using semantic similarity between terms. In *Proc. of EC*, 2007.

[2] J. Attenberg, S. Pandey, and T. Suel. Modeling and predicting user behavior in sponsored search. In *Proc. of KDD*, 2009.

[3] A. Berger and V. Pietra. A maximum entropy approach to natural language processing. In *Computational Linguistics*, 1996.

[4] Y. Chen and T. W. Yan. Position-normalized click prediction in search advertising. In *Proc. of KDD*, 2012.

[5] H. Cheng and E. Cantu-Paz. Personalized click prediction in sponsored search. In *Proc. of WSDM*, 2010.

[6] M. Ciaramita, V. Murdock, and V. Plachouras. Online learning from click data for sponsored search. In *Proc. of WWW*, 2008.

[7] C. Clarke, E. Agichtein, S. Dumais, and R. White. The influence of caption features oh clickthrough patterns in web search. In *Proc. of SIGIR*, 2007.

[8] K. Dembczynski, W. Kotlowski, and D. Weiss. Predicting ads click-through rate with decision rules. In *Workshop on Targeting and Ranking in Online Advertising*, 2008.

[9] B. Edelman, M. Ostrovsky, and M. Schwarz. Internet adverstising and the generalized second-price auction: selling billions of dollars worth of keywords. In *The American Economic Review*, 2007.

[10] D. Fain and J. Pedersen. Sponsored search: a brief history. In *Proc. of 2nd Workshop on Sponsored Search Auctions*, 2006.

[11] T. Graepel, J. Candela, T. Borchert, and R. Herbrich. Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft's bing search engine. In *Proc. of ICML*, 2010.

[12] D. Hillard, E. Manavoglu, H. Raghavan, C. Leggetter, E. Cantls-Paz, and R. Iyer. The sum of its parts: reducing sparsity in click estimation with query segments. In *Information Retrieval Journal*, 2011.

[13] W. D. Hoyer and D. J. Maclnnis. *Consumer Behavior*.

[14] B. Jansen and T. Mullen. Sponsored search: an overview of the concept, history, and technology. In *International Journal of Electric Business*, 2008.

[15] S. Kim, T. Qin, H. Yu, and T.-Y. Liu. Advertiser-centric approach to understand user click behavior in sponsored search. In *Proc. of CIKM*, 2011.

[16] A. H. Maslow. A theory of human motivation. In *Psychological Review*, 1943.

[17] T. P. Minka. A comparison of numerical optimizers for logistic regression. In *Technical report, Microsoft*, 2003.

[18] A. Mordecai. *Nonlinear Programming: Analysis and Methods*.

[19] G. N. Punj and D. W. Stewart. An interaction framework of consumer decision making. In *Journal of Consumer Research*, 1983.

[20] F. Radlinski, A. Broder, P. Ciccolo, E. Gabrilovich, V. Josifovski, and L. Riedel. Optimizing relevance and revenue in ad search: a query substitution approach. In *Proc. of SIGIR*, 2008.

[21] M. Richardson, E. Dominowska, and R. Ragno. Predicting clicks: estimating the click-through rate for new ads. In *Proc. of WWW*, 2007.

[22] R. L. Sandhusen. *Market, Business Review*.

[23] B. Shaparenko, O. Cetin, and R. Iyer. Data-driven text features for sponsored search click prediction. In *Proc. of ADKDD*, 2009.

[24] W. Zhang, X. He, B. Rey, and R. Jones. Query rewriting using active learning for sponsored search. In *Proc. of SIGIR*, 2007.