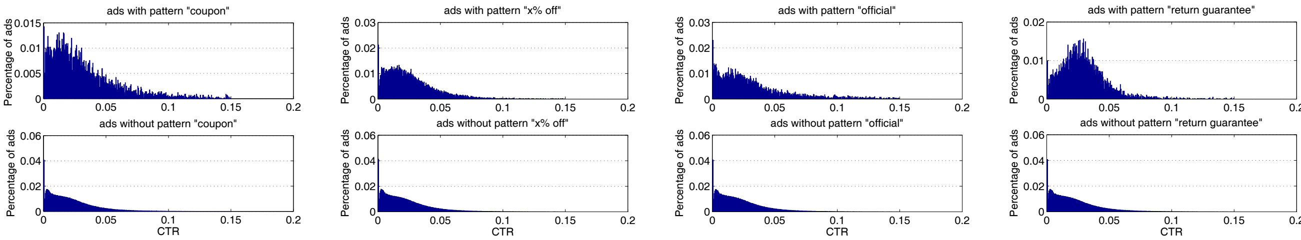

(a) "*coupon*" (b) "*x% off*" (c) "*official*" (d) "*return guarantee*"

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

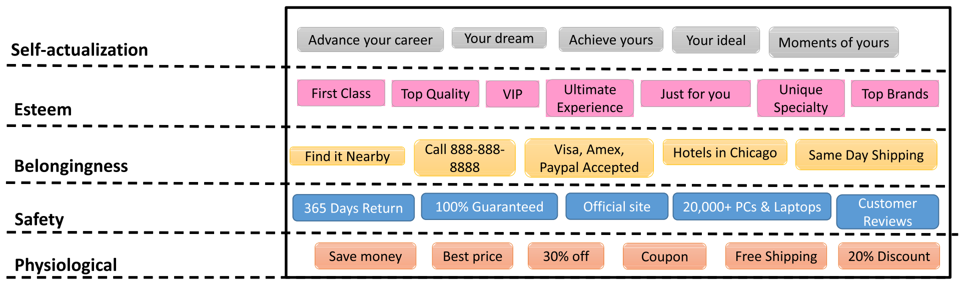

Figure 4: A Hierarchy of User Psychological Desires in Sponsored Search.

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

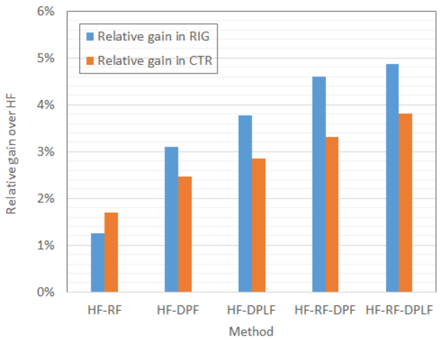

Figure 5: Relative gain of different methods over HF in terms of CTR and RIG.

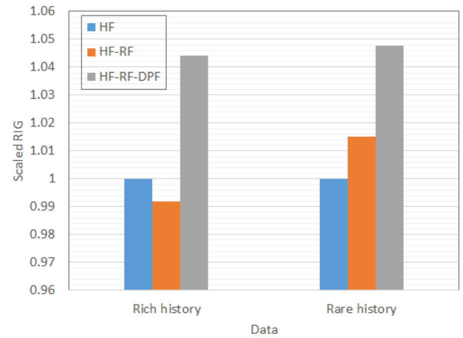

Figure 6: Relative gain of different methods over HF in terms of RIG, conducted on ads with rich or rare history.

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.