Learning to Rank Audience for Behavioral Targeting

Ning Liu[1]   Jun Yan[1]   Dou Shen[2]   Depin Chen[3]   Zheng Chen[1]   Ying Li[2]

[1]Microsoft Research Asia
Sigma Center, No.49,
Zhichun Road
Beijing, 100190, China

{ningl, junyan, zhengc}@microsoft.com

[2]Audience Intelligence
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052, USA

{doushen, yingli}@microsoft.com

[3]Department of Computer Science
University of Science and Technology
of China
Hefei, P.R. China

depin.chen@gmail.com

ABSTRACT

Behavioral Targeting (BT) is a recent trend of online advertising market. However, some classical BT solutions, which predefine the user segments for BT ads delivery, are sometimes too large to numerous long-tail advertisers, who cannot afford to buy any large user segments due to budget consideration. In this extend abstract, we propose to rank users according to their probability of interest in an advertisement in a learning to rank framework. We propose to extract three types of features between user behaviors such as search queries, ad click history etc and the ad content provided by advertisers. Through this way, a long-tail advertiser can select a certain number of top ranked users as needed from the user segments for ads delivery. In the experiments, we use a 30-days' ad click-through log from a commercial search engine. The results show that using our proposed features under a learning to rank framework, we can well rank users who potentially interest in an advertisement.

Categories and Subject Descriptors: H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *retrieval models*; H.4.m [**Information Systems Applications**]: Miscellaneous.

General Terms: Economics, Experimentation

Keywords
Online Advertising, Behavioral Targeting, Learning to Rank

1. INTRODUCTION

In recent years, user-targeted advertising is booming in industrial circles by collecting and analyzing user data in many aspects such as online behavior, demographic and geographic information. Among various sources of user data, the richest user behavioral data exhibit the biggest potential for further exploitation. Many major search engines are providing Behavioral Targeting (BT) [1] solutions to advertisers through acquiring BT companies or bringing behavioral targeting technique into their existing products. Generally speaking, the output of many BT solutions is hundreds of user segments, within each of which users are believed to have similar interests. Then, each segment is given a snippet or name like "sports enthusiasts" for advertisers to judge which segments they should buy for ads delivery. However, in this way, the relatedness between the user segment and the advertiser is assessed by human experience rather than from data mining directly. Moreover, since a single user segment can have millions of users, long-tail advertisers may not have enough budgets to buy a complete segment. They are supposed to know which part of users in a segment is the best audience desired by them, which results in a user ranking problem.

In this paper, we propose to rank users according to their potential interests on a given advertisement. Similar to the learning to rank problem in the field of information retrieval for document ranking, we consider our problem in the same framework by treating "ads" as "queries" and "users" as "documents". We propose three types of features based on the user behavioral data and the ad content. The behavioral data include users' query history and ad click history. The first type of features reflects the direct content matching between users' behavioral data and the ad content. For the second type of features, we map both user behaviors and ads contents into high-level topics and then compute their topic similarity. Finally, the third type of features gives some priori user behavioral statistics, e.g. global ad click-through rate (CTR), a high value of which indicates that the user may have habit to view ads and perform clicks. Using these user behavior oriented features, we use a subset of ads together with users who ever viewed or clicked these ads as training data to learn a user ranking model. The experimental results show that using our proposed features under a learning to rank framework, we can well rank users who potentially interest in an advertisement.

2. LEARNING TO RANK AUDIENCE

Suppose an advertiser from a certain industry wants to run his ad campaigns online and comes to an online ad publisher (e.g. a search engine or an ad network). He knows much about his products which are going to be advertised, but has no idea about the online audience. There are billions of people on the Internet, and hence it is impossible and unaffordable to deliver his ads to all of them. Although he may buy some predefined user segments with descriptions from the BT companies, the number of user could be large for many long-tail advertisers. The cost of buying user segments will exceed his budget. To make efficient use of his budget, he wants to find a smaller part of best audience within a user segment, and only displays ads to them during online activities. To address this problem, we propose a user ranking mechanism according to the advertiser's ad campaigns.

Analogous to learning to rank documents in document retrieval [2], in the user ranking problem, there should be queries, document repository, relevance label definition and feature set. In terms of learning to rank audience, the "query" is the advertisement a. An online user u is treated as a "document", with the label l representing whether this user has viewed or clicked the given ad, where $l=-1$ stands for the user u has never viewed a, $l=0$ stands for u has viewed a but did not click it, and $l=1$ stands for u has viewed a and clicked it. In this work, we make use of two types of behavioral data: user search query history and user ad click history. Features are extracted from the ad and these behavioral data simultaneously. We list the description of all three types of features we propose to use in Table 1. Using this set of

features, which are represented by a feature vector *x*, for each user-ad pair with label *l*, i.e. *<u, a, l>*, we can train a ranking model under the learning to rank framework. And then, for any new ad and all candidate users, we can rank the audience according to their probability to click the given ad.

Table 1. Feature description

No.	feature name	description
Relevance features		
#1	'query history to ad title' similarity	BM25 score computed between a user's query history and the title of the given ad.
#2	'query history to ad description similarity	BM25 score computed between a user's query history and the description of the given ad.
#3	'query history to ad page similarity	BM25 score computed between a user's query history and the landing page content of the given ad.
#4	'clicked query history to ad page similarity	BM25 score computed between a user's query history (queries with ad click) and the landing page content of the given ad.
#5	'non-click query history to ad page similarity	BM25 score computed between a user's query history (queries without ad click) and the landing page of the ad.
#6	'clicked ad history to ad page similarity	BM25 score computed between a user's ad click history and the whole content of the given ad.
Topic-based relevance features		
#7	'query history to ad page topic similarity	Cosine similarity of topics between query history and the ad landing page content.
#8	'clicked ad history to ad page topic similarity	Cosine similarity of topics between ad click history and the ad landing page content.
Priori behavioral statistics features		
#9	user global CTR	User's global ad click CTR in the user behavioral history log.
#10	topic-specific CTR	User's ad click CTR when inputting queries which have the same topic with the given ad.
#11	topic-specific %query	The percentage of queries in user query history that have the same topic with the given ad.
#12	topic-specific %ad-click	The percentage of user clicked ads having the same topic with the given ad.
#13	# user query topic	The number of topics covered by the user submitted queries in history.
#14	# user ad-click topic	The number of topics covered by the ads that have been clicked by a user.

After surveying existing learning to rank methods, we choose Ranking SVM [2] as our main method due to its proven efficacy in prior research. Running Ranking SVM is equivalent to solving the following optimization problem.

$$\min_{\omega, \xi_{ij}} \frac{1}{2}\|\omega\|^2 + C\sum \xi_{ij}$$

$$\text{s.t. } \langle \omega, x_i - x_j \rangle > 1 - \xi_{ij}, \forall x_i \succ x_j, \xi_{ij} > 0$$

where the subscripts *i* and *j* are used to distinguish different feature vectors extracted from behaviors of different users. Suppose the solution of this optimization problem is ω^* and a series of ξ_{ij}, the obtained ranking model is,

$$f(x) = \langle \omega^*, x \rangle.$$

3. VALIDATION

For experiments, we use a dataset comes from a commonly used commercial search engine which lasts for a full month. It contains the fields such as user ID, query text of user, the displayed ad to user, the clicked ad by user etc. We select 100 frequently clicked ads from the ad repository for demonstration. We only consider the users who have viewed and/or clicked these 100 ads for ranking. To label users for experiments, if a user clicked an ad, he is labeled as relevant to the ad; the users who have been displayed the ad but did not click are labeled as irrelevant. For all users correlate to the 100 ads, we create the profile for each user, including the user query history and ad click history. Then, for each ad, we extract the 14 features defined in Table 1to train the ranking model. We randomly divide the whole user dataset into 5 folds and we run the experiments for 5 times. In each round, 1 fold is used as testing dataset and others are used for training.

Figure 1 shows the normalized distributions of relevant users in the ranking list. The x-axis indicates positions in the ranked list. The y-axis indicates the proportion of relevant users in a specific position over the 100 ads. From the figure we can see that through using our audience ranking solution, more relevant users are ranked to higher positions.
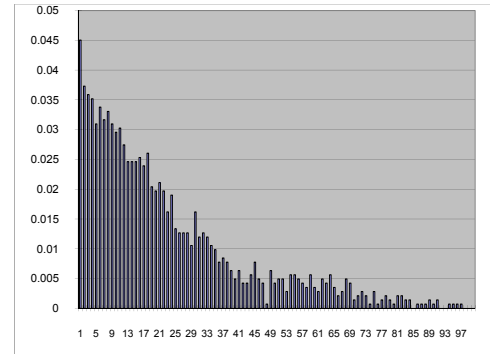

Figure 1. The distribution of relevant users

Over the 100 ads and all related users in our dataset, the averaged ad CTR is about 2.02%. After ranking, at the first position of each ad, CTR of ad is as high as 16.7% and 11.2% at the second position. This means we can well rank users who potentially interest in an advertisement.

4. CONCLUSION

In this paper, we propose a novel user ranking problem to answer "whom to deliver ads" for online advertisers. We utilize the user behavioral data including user search query history and ad click history to create the user profile. 14 features are defined and extracted between user profile and the content of a given ad. We then embed the user ranking problem into the learning to rank framework and employ Ranking SVM to obtain the ranking model. Experimental results show that our approach can effectively rank the relevant users on top.

5. REFERENCES

[1] J. Yan, N. Liu, G. Wang, W. Zhang, Y. Jiang, and Z. Chen. How much behavioral targeting help online advertising? In *Proceedings of the 18th International World Wide Web Conference (WWW)*, 2009.

[2] R. Herbrich, T. Graepel, and K. Obermayer. Large margin rank boundaries for ordinal regression. In *Advances in Large Margin Classifiers*, MA, MIT Press, 2000.