Query Augmentation based Intent Matching in Retail Vertical Ads

Huasha Zhao
University of California
Berkeley CA 94720
hzhao@cs.berkeley.edu

Vivian Zhang
Microsoft Corporation
Sunnyvale CA 94085
wzha@microsoft.com

Ye Chen
Microsoft Corporation
Sunnyvale CA 94085
yec@microsoft.com

John Canny
University of California
Berkeley CA 94720
jfc@cs.berkeley.edu

Tak Yan
Microsoft Corporation
Sunnyvale CA 94085
takyan@microsoft.com

ABSTRACT

Search advertising shows trends of vertical extension. Vertical ads typically offer better Return of Investment (ROI) to advertisers as a result of better user engagement. However, campaign and bids in vertical ads are not set at the keyword level. As a result, the matching between user query and ads suffers low recall rate and the match quality is heavily impacted by tail queries. In this paper, we propose a retail ads retrieval framework based on query rewrite using personal history data to improve ads recall rate. To insure ads quality, we also present a relevance model for matching rewritten queries with user search intent, with a particular focus on rare queries. Extensive experiments are carried out on large-scale logs collected from the Bing search engine, and results show our system achieves significant gains in ads retrieval rate without compromising ads quality. To our knowledge, this work is the first attempt to leverage user behavioral data in ad matching and apply it to the vertical ads domain.

Categories and Subject Descriptors

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

Keywords

ads retrieval, personalization, tail queries

1. INTRODUCTION

Online advertising is a multi-billion dollar business, and the market sees continuing double digit growth since its inception. Recent years have seen rapid growth of online vertical ads and an increasing number of advertisers start to adapt to this emerging campaign paradigm. Comparing with text ads product, vertical ads have shown significant better user engagement and advertiser satisfaction. Search users can see the exact products before they even reach advertiser's sites, which leads to more clicks/CTR and higher ROI for advertisers.

Different from the traditional text ads, vertical ads requires no keyword match. Whenever a user enters a search query with commercial intent, search engine retrieves relevant items from ads inventory. This makes it very easy for advertiser to promote their entire product inventory. On the other hand, search engine is having much greater responsibility on ads relevance and recall: it is a search task on its own in the product inventory. A key challenge for vertical ads is to match high quality ads with user intent (e.g. product in mind to buy), while maximizing ads recall. The current pressing issue is low recall rate and hence low auction density due to the limited number of listings. Moreover, we have observed that the ads for specific commercial queries outperforms general queries often for retail vertical ads. These ads inventory are often sparse, and the associated queries are also less frequently searched due to their specificity. How to effectively retrieve relevant ads for these tail queries is another challenge.

In this paper, we proposed an ads retrieval framework to address the problems above. The idea is to retrieve ads using both the original query and multiple of its personalized rewrites to increase recall. This is quite unlike most of other personalization works which instead focus on improving precision using personal data. The key challenge of personal rewrite is to find relevant queries from history. We developed a hybrid relevance model that captures query-level similarity at first level, and does word-level matching using a topic model at second level to address the issue of rare queries.

2. ADS RETRIEVAL FRAMEWORK

Figure 1 shows a diagram of the ads retrieval system architecture, where the new components for query rewriting are drawn within the dashed box. In the current production system, the original query are directly sent to the vertical ad index, and it is up to the ads engine to (1) retrieve ads with keywords matching the query; (2) estimate relevance scores for the retrieved ads; (3) return ads with top relevance scores.

Our ad retrieval system feed both the original query and multiple of its rewrites to the ad index which ultimately returns product ads for all the queries to display. Rewrite can-

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2577317.

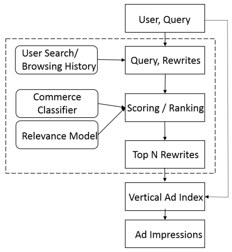


Figure 1: Ads Retrieval System Framework

didates are either directly chosen from user search/browsing histories (we use a Microsoft internal service to translate browsing url to query) or a combination of current query and history queries. The raw rewritten query can be noisy: they may not be relavant with user search intent (thus the ads returned may increase user frustration) or does not exibit commerce intent. Therefore, in our system, all query candidates would have to pass through a commerce intent classifier, and the rewritten queries must meet the relevance requirement with regarding to the original query according to our relevance model. Then, selected candidates are ranked according to the commerce intent score and relevance measure score, and top queries are selected are sent to the ad index to retrieve ads together with the original query.

The top relevant rewrites are good complements of the original query. An example would be rewriting “watches” to “watches louis vutton” if “louis vutton” is in the user history. This rewrite enriches user intent, and triggers more ads by replacing an abstract search term with a concrete product. It may potentially improve relevance as well, since it reflects user preference.

This framework requires minimal modification to the current production system. At the same time, the scheme is arguably the most effective and intuitive one to follow since it tackles the problem from the beginning of the work flow.

3. RELEVANCE MODEL

The key challenge of this framework is to find related rewrites. We describe here two techniques to measure relevance between two queries.

3.1 Query-Level Matching

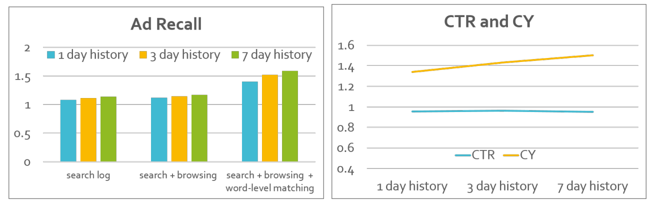
In query-level matching, query pairs are generated from successive search queries issued by a single user within a single search session [1]. We collapse repeated query pairs over users across all the sessions. So that the data consists of all query pairs and frequencies of appearance.

We then assign scores to each pair of queries according to a Bernoulli model using likelihood ratio [2], where higher likelihood ratio indicates higher relevance between two queries.

The formula to compute the log likelihood ratio of a query pair (q_1, q_2) is,

$$LLR = 2(\log L(\frac{k_1}{n_1}, k_1, n_1) + \log L(\frac{k_2}{n_2}, k_2, n_2) \quad (1)$$
$$- \log L(\frac{k_1+k_2}{n_1+n_2}, k_1, n_1) - \log L(\frac{k_1+k_2}{n_1+n_2}, k_2, n_2))$$

where where n_1 is the count of q_1 showing up in the first place, and k_1 is among those the count q_2 succeeding q_1. Similarly n_2 is the total number of query pairs whose first place is not q_1, and k_2 counts all the q_2 occurrences without q_1 showing up ahead of it. And $L(p, k, n) = p^k(1-p)^{n-k}$.



(a) Ad recall (b) CTR/CY

Figure 2: Performance

3.2 Word-Level Matching for Tail Queries

The query-level model above only covers less than 15% of all the new query pairs to be judged. All the other queries are scored zero because the system have never seen the pairs before and considered to be unrelated. So the recall is low. To solve this problem, we break down query into words and construct a similarity function based on a factor model.

We first build a product-word matrix D using click-through data from Bing shopping. We add one to $D_{(p,w)}$ when a query containing word w leads to a click on product p. We then factorize this matrix using LDA, $D = P \times W$, where D, P, W are $m \times n, k \times m, k \times n$ matrices respectively, and k is much smaller than n, m. We interpret k as the dimension of category space, and P and W are category representation of products and words respectively,

The category representation of the each query is then estimated using MLE of the query, which the average sum of the normalized words vector: $c_q = W \times v_q^T$, where v_q is the bag-of-word representation (uni-gram) of the query. Then the relevance score is obtained by computing the cosine similarity between two queries $\frac{c_{q_1} \cdot c_{q_2}}{||c_{q_1}|| \cdot ||c_{q_2}||}$.

4. EXPERIMENTS

We carry out extensive experiments on large-scale user log data collected from Bing. Figure 2a demonstrates the improvement in ads recall with different days of history used. We compare three configurations of the proposed system with the current production system. The first configuration uses search query log only, while the second one also incorporates browsing history. Both configurations filter queries using the query-level matching model only. The last configuration applies word-level query matching as well. The recall numbers of all configurations are relative to the baseline, of which all the numbers are normalized to 1. As illustrated in the figure, we show overall 60% improvement in ads recall. We also observe the effect of incorporating more personal data sources. With increased recall, we further compare the CTR and click yield (CY) in Figure 2b. The results show that the CTR remains largely unchanged, and the CY gain is proportional to the increase of recall.

5. CONCLUSIONS

This paper introduces an innovative framework for vertical ad retrieval, and demonstrates its performance gains. Work remains to tune the ranking model of the rewrites. We will also explore the benefit of incorporating other personal data.

6. REFERENCES

[1] P. N. Bennett, R. W. White, W. Chu, S. T. Dumais, P. Bailey, F. Borisyuk, and X. Cui. Modeling the impact of short- and long-term behavior on search personalization. In *SIGIR'12*.

[2] R. Jones, B. Rey, O. Madani, and W. Greiner. Generating query substitutions. In *WWW'06*.