RPM-Oriented Query Rewriting Framework for
E-commerce Keyword-Based Sponsored Search (Student Abstract)

Xiuying Chen,[1,2] **Daorui Xiao,**[3] **Shen Gao,**[2] **Guojun Liu,**[3] **Wei Lin,**[3] **Bo Zheng,**[3]
Dongyan Zhao,[1,2] **Rui Yan**[1,2*]

[1]Center for Data Science, Peking University, Beijing, China,
[2]Wangxuan Institute of Computer Technology, Peking University, China, [3]Alibaba Group, China
Email: xy-chen@pku.edu.cn

Abstract

Sponsored search optimizes revenue and relevance, which is estimated by Revenue Per Mille (RPM). Existing sponsored search models are all based on traditional statistical models, which have poor RPM performance when queries follow a heavy-tailed distribution. Here, we propose an RPM-oriented Query Rewriting Framework (RQRF) which outputs related bid keywords that can yield high RPM. RQRF embeds both queries and bid keywords to vectors in the same implicit space, converting the rewriting probability between each query and keyword to the distance between the two vectors. For label construction, we propose an RPM-oriented sample construction method, labeling keywords based on whether or not they can lead to high RPM. Extensive experiments are conducted to evaluate performance of RQRF. In a one month large-scale real-world traffic of e-commerce sponsored search system, the proposed model significantly outperforms traditional baseline.

Introduction

Web search seeks to optimize relevance, while sponsored search optimizes RPM, which takes both relevance and bid price together into consideration. Obviously, it is not possible for advertisers to identify and collect all relevant bid keywords to target their ads. A common mechanism to solve this problem is query rewriting, which outputs a list of bid keywords (rewrites) that are related to a given query. Query rewriting has been extensively studied in the context of traditional web search. Initial studies only focuses on the relevance between a query and bid keyword, without considering the traffic allocation mechanism on the bid keyword side to maximize revenue. More comprehensive works are based on graphs. For example, (Malekian et al. 2008) formulates a family of graph covering problems whose goals are to suggest a subset of ads with the maximum benefit by suggesting rewrites for a given query. However, these models also have shortcomings: (1) Each query is regarded as a non-splittable graph node, which hinders the full understanding of queries. (2) They are all essentially statistic-based memory models, and thus cannot handle heavy-tailed ads.

To overcome obstacles mentioned above, we come up with an RPM-oriented Query Rewrite Framework (RQRF). RQRF embeds both queries and bid keywords to vectors in the same implicit space. Specifically, we combine a modified version of a convolutional neural network (CNN) with a self-attention mechanism to learn representations of queries and bid keywords. For the label construction, since users click on ads instead of the bid keywords, we propose an RPM-oriented sampling method to bridge the gap between labeled ads and unlabeled keywords.
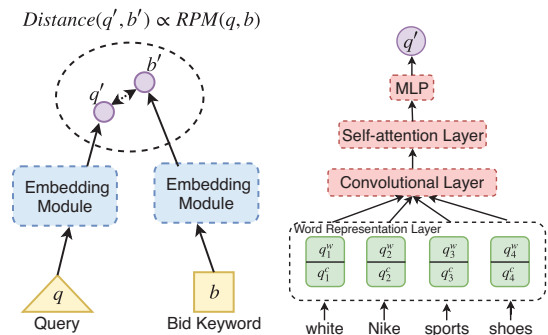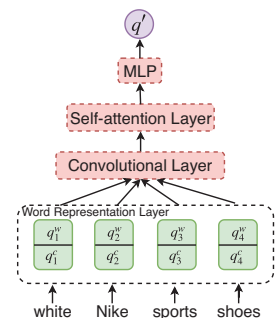
Model

Figure 1: Overview of RQRF

Figure 2: Illustration of the Embedding Module.

Embedding Module. As shown in Figure 1, our model maps each query-bidword pair to a low dimensional vector q' and b' respectively. We first introduce the embedding module using query as an example, as shown in Figure 2. Bid keywords are embedded in a similar way. We adopt standard techniques to obtain the embedding of each word q_t by concatenating its word embedding q_t^w and character embedding q_t^c. The output of a given word q_t from this layer is calculated as $m_t = [q_t^w, q_t^c]$, where "[,]" denotes the concatenation between vectors.

Next, a novel deep convolutional neural network architecture from (Chollet 2017) and the self-attention mechanism are adopted, producing α_t. In this way, the model learns the

interactions and exchanges of information between words by itself. Finally, a fully-connected layer is applied to obtain the final query embedding:

$$q' = FC([a_1, a_2, ..., a_T]). \tag{1}$$

Label Construction. Since there are no click labels on bid keywords, only click labels on ads. Herein, we propose an RPM-oriented sampling method to construct limited number of training samples, which greatly reduces the size of training data and mitigates the Matthew effect. Let the number of ads that bid on keyword b_i be n_{b_i}, and the price ad a bid on b_i be price$(b_i|a)$ for a sample $< q, a >$. The probability of keyword b_i being positively sampled, i.e. score$(b_i|q, a)$ is calculated as:

$$\text{score}(b_i|q, a) = \frac{\text{price}(b_i|a) \times \text{relevance}(b_i, q)}{\log(n_{b_i} + 1)},$$

where relevance$(b_i, q) = \text{cosine}(b_i, q)$. Note that the relevance is calculated by pre-trained word embeddings instead of embedding in our model. We add factor n_{b_i} to allow the model to choose keywords that are not that popular, in order to balance advertisement competition. As for the relevance(b_i, q) factor, it is used to enhance the importance of relevance between query and bid keyword. Then the scores of all keywords bought by advertisement a are normalized and we obtain the final sample probability $p(b_i|a, q)$. where $B(a)$ represents all keywords bought by a. This sampling method leads to a positive correlation between the probability that q matches b_i and the RPM value. The details of derivation are attached in Appendix.

After obtaining the value of p, we use p as a sampling probability to find positive keyword samples, and $1 - p$ as a sampling probability to obtain negative samples. The training objective is to maximize the cosine distance between a query embedding and positive keyword embeddings.

Experiments

Dataset. We use the search logs from one of the largest e-commerce websites in China to train RQRF. Overall, there are 9.8 billion training cases and 66 million evaluation and 66 million test cases. On average, there are 3.94 words and 8.66 characters in queries and 2.88 words and 6.34 characters in bid keywords.

Offline Evaluation Apart from NLL, we also employ commonly-used ranking metrics, including mean average precision (MAP), mean reciprocal rank (MRR) and normalized discounted cumulative gain (NDCG) to evaluate RQRF. We compare our query embedding model with various baselines, including **BM25**, **PM**(non-linear projection),**CLSM**(Ahsan and Agrawal 2018),**BIDAF**(Seo et al. 2016). The main results of our experiments are summarized in Table 1 where RQRF outperforms all baselines. Statistical significance of observed differences between the performance of two runs are tested using a two-tailed paired t-test and is denoted using ▲(or ▼) for strong significance for $\alpha = 0.01$.

Online Evaluation. RQRF is deployed on a real-world commercial search engine to serve ads as part of A/B

Table 1: Comparison with query embedding baselines.

Models	NLL	MAP	MRR	NDCG
BM25	0.870	0.149	0.009	0.157
PM	0.662	0.166	0.011	0.191
BIDAF	0.623	0.180	0.012	0.201
CLSM	0.631	0.191	0.014	0.226
RQRF	**0.617**▼	**0.207** ▲	**0.015**▲	**0.253**▲

testing. The traditional statistic-based memory model from (Malekian et al. 2008) is used as the control model and RQRF as the treatment. The flight tests are conducted for a period of one month. The online results are summarized in Table . Our model outperforms the baseline in the head traffic by 8.79% in RPM value. For tail queries, our model achieves consistently better performance, outperforming the baseline by 16.44%. This demonstrates that our model is especially effective for tail queries. The ads retrieved by RQRM account for 24.7% of the total queries in the online test, and bring about 3.07% RPM increase in total traffic.

Table 2: RPM lift rate result on head and tail of query distribution.

Traffic type	RPM Lift Rate
Head	8.79%
Tail	16.44%
Head&Tail	12.75%

Conclusion

In this paper, we propose an RQRM query rewriting model that outputs related bid keywords for a given query. We are the first to apply a deep neural network model in a keyword-based sponsored search system. Extensive experiments are conducted to evaluate the performance of RQRF.

Acknowledgments

References

Ahsan, H., and Agrawal, R. 2018. Approximating categorical similarity in sponsored search relevance. *arXiv preprint arXiv:1812.00158.*

Chollet, F. 2017. Xception: Deep learning with depthwise separable convolutions. *arXiv preprint* 1610–02357.

Malekian, A.; Chang, C.-C.; Kumar, R.; and Wang, G. 2008. Optimizing query rewrites for keyword-based advertising. In *ACM EC*, 10–19. ACM.

Seo, M.; Kembhavi, A.; Farhadi, A.; and Hajishirzi, H. 2016. Bidirectional attention flow for machine comprehension. *arXiv preprint arXiv:1611.01603.*