

Beyond Keyword Targeting: An End-to-End Ad Retrieval Framework for Sponsored Search

Xiao Yang*, Zhi Guo, Shouke Qin, Zongyao Ding

Baidu Inc.

{yangxiao04, guozhi, dingzongyao, qinshouke}@baidu.com

ABSTRACT

As the main revenue source for search engines, sponsored search system retrieves and allocates ads to display on the search result pages. Keyword targeting is widely adopted by most sponsored search systems as the basic model for expressing the advertiser's business and retrieving related ads. In this targeting model, the advertiser should cautiously select lots of keywords relevant to their business to optimize their campaigns, and the sponsored search system retrieves ads based on the relevance between queries and keywords. However, since there is a huge inventory of possible queries and the new queries grow dramatically, it is a great challenge for advertisers to identify and collect lots of relevant bid keywords for their ads, and it also takes great effort to select and maintain high-quality keywords and set corresponding match types for them. In the meantime, the keyword targeting leads to a multi-stage retrieval architecture as it contains the matching between query and keywords and the matching between keywords and ads. The retrieval funnel based on keyword targeting cannot achieve straightforward and optimal matching between search queries and ads. Consequently, traditional keyword targeting method gradually becomes the bottleneck of optimizing advertisers' campaigns and improving the monetization of search ads.

In this paper, we present an end-to-end ad retrieval framework for sponsored search. This framework can break the limits of keyword targeting and can achieve direct matching from query to related ads. The framework has been deployed at Baidu's sponsored search system and the experimental result validates its effectiveness. We will describe the design and architecture of this framework, and hope that this framework can give some inspiration for the sponsored search industry.

ACM Reference Format:

Xiao Yang*, Zhi Guo, Shouke Qin, Zongyao Ding. 2019. Beyond Keyword Targeting: An End-to-End Ad Retrieval Framework for Sponsored Search. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19)*, July 21–25, 2019, Paris, France. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3331184.3331429>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGIR '19, July 21–25, 2019, Paris, France

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-6172-9/19/07...\$15.00

<https://doi.org/10.1145/3331184.3331429>

1 INTRODUCTION

Keyword targeting is the key to the business model of sponsored search and has more than 20 years of history. Nowadays, the major search engines have billions of search queries each day and the new queries grow dramatically. As the search engine has massive and various queries and the search volume changes dynamically, it poses huge challenges for advertisers in the selection and maintenance of relative keywords.

In the keyword targeting model, the keyword is used as the intermediary between query and ad, and modern sponsored search platform often adopts a multi-stage retrieval procedure which is shown in Figure 1. The multi-stage retrieval procedure is a cascaded funnel. As each stage has its respective optimization target and has inevitable errors, these errors will be accumulated and enlarged by stage-wise processing which will lead to losses of efficiency and accuracy in retrieving related ads for target query. Moreover, as the keywords are short texts and are the brief descriptions of advertisers' business, it can't provide sufficient information to diminish the semantic gap between queries and advertisers' business.

Are the keywords really necessary for sponsored search? To liberate advertisers from the difficult keywords selection task and increase the efficiency of search ads retrieval, in this paper, we will present an end-to-end ad retrieval framework for sponsored search in which the retrieval procedure is more straightforward and the keywords are only used as supplementary information.

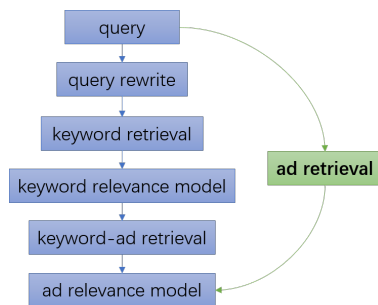


Figure 1: The multi-stage procedure of ad retrieval. The blue part is the traditional keyword-based retrieval procedure, and the green part is the end-to-end ad retrieval procedure.

Based on this framework, we can break the limits of keyword targeting, and more rich information can be used to represent the advertiser's business. The retrieval results of advertiser's keywords will be expanded and the ad retrieval task can be transformed into model-based semantic matching problem. We deployed the framework in Baidu's sponsored search system and verify the effectiveness of the framework.

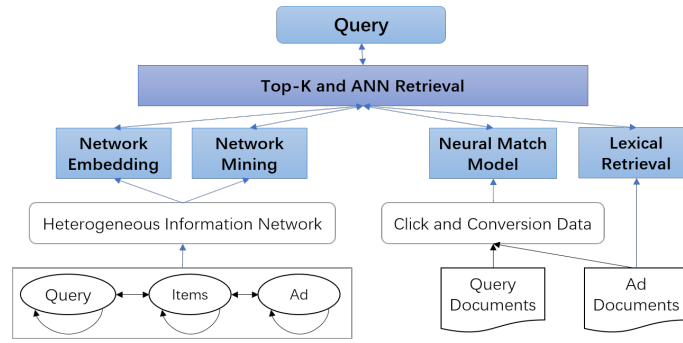


Figure 2: The end-to-end ad retrieval framework.

2 FRAMEWORK OVERVIEW

The overall architecture of the proposed framework is shown in Figure 2. In this framework, we provide two types of methods to expand the keyword-based ad retrieval and to break the limits of multi-stage retrieval procedure of keyword targeting. Both methods can retrieve related ads directly from the query.

2.1 Basic Requirements

To support the end-to-end ad retrieval framework, there are some basic requirements, which are: (1) Basic data: Several sources of data should be used to sufficiently represent the advertiser’s business and the relationship between queries and ads, such as ad click data, ad content data and search log data, and so on. (2) Bidding for the new retrieved ads: Casting off the dependence of keyword-level bid prices settled by advertiser is a necessary prerequisite for end-to-end ad retrieval. We adopt the conversion rate based bid adjustment[5] and the target-CPA bidding strategy[9] to generate the bid prices for the new retrieved ads to ensure the ROI of advertisers. In the meantime, if the bidding strategy can ensure the ROI, then the match types of traditional keyword targeting are no more necessary and can treat the advanced broad match as the default match type.

2.2 Ad Retrieval on Heterogeneous Network

Firstly, We can model the problem of ad retrieval as a network-based mining task. Inspired by[8], we construct a heterogeneous information network. As shown in Figure 2, we can use various items such as keywords, ads, urls, intent tags, clicked search results, etc, to connect the query nodes with the ad nodes. Based on this network, we can adopt the meta-path-based mining methods such as [7] and network embedding models such as[2] to find potential unconnected query-ad pairs. Meanwhile, to find the suitable retrieval results for long-tail nodes and incorporate the attributes of nodes and edges in the network, we also adopt the graph neural network models such as the GraphSAGE framework [3] to learn the node embeddings. By combining the network mining results and the learned node embeddings, we can retrieve the top-K ad nodes directly for the target query node.

2.3 Ad Retrieval as Text Matching

Secondly, we can treat the ad retrieval task as the text matching problem by representing the query and ad as text documents. In this

way, we can use rich content information of queries and ads and can retrieve ads for the target query just the same as the retrieval methods of organic search results. We expand the query text with its rewrites and clicked search results, and extract the content of the landing page and ad creation as the ad text and expand it with its related keywords and clicked creations. Based on these documents, we develop an inverted index of ad documents, and retrieve the ads for target query just as the traditional IR task. Meanwhile, we also adopt the neural matching models such as [6] and [4] to learn the representation and matching of queries and ads based on training on the click and conversion feedback data. The term-based representations from transformers[1] are also used by pre-training on the corpus of queries and ads.

ACKNOWLEDGMENTS

We would like to thank Ruiwei Zhu, Junwei Zhang, Penghe Liu, Kai Xu, Xinran Liu, Shaogui Ye, Yang Wang for their contributions.

REFERENCES

- [1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
- [2] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. 2017. metapath2vec: Scalable representation learning for heterogeneous networks. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 135–144.
- [3] Will Hamilton, Zhitaoying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*. 1024–1034.
- [4] Liang Pang, Yanyan Lan, Jiafeng Guo, Jun Xu, Shengxian Wan, and Xueqi Cheng. 2016. Text matching as image recognition. In *Thirtieth AAAI Conference on Artificial Intelligence*.
- [5] Benjamin Rey and Ashvin Kannan. 2010. Conversion rate based bid adjustment for sponsored search. In *Proceedings of the 19th international conference on World wide web*. ACM, 1173–1174.
- [6] Yelong Shen, Xiaodong He, Jianfeng Gao, Li Deng, and Grégoire Mesnil. 2014. Learning semantic representations using convolutional neural networks for web search. In *Proceedings of the 23rd International Conference on World Wide Web*. ACM, 373–374.
- [7] Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S Yu, and Tianyi Wu. 2011. PathsSim: Meta-path-based top-k similarity search in heterogeneous information networks. *Proceedings of the VLDB Endowment* 4, 11 (2011), 992–1003.
- [8] Xiao Yang, Tao Deng, Zhi Guo, and Zongyao Ding. 2017. Advertising keyword recommendation based on supervised link prediction in multi-relational network. In *Proceedings of the 26th International Conference on World Wide Web Companion*. ACM, 863–864.
- [9] Xiao Yang, Daren Sun, Ruiwei Zhu, Tao Deng, Zhi Guo, Zongyao Ding, Jiao Ding, and Shouke Qin. 2019. AiAds: automated and intelligent advertising system for sponsored search. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 1021–1030.