Information & Management xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Information & Management

journal homepage: www.elsevier.com/locate/im

Finding competitive keywords from query logs to enhance search engine advertising

Dandan Qiao[a], Jin Zhang[b,*], Qiang Wei[a], Guoqing Chen[a]

[a] School of Economics and Management, Tsinghua University, Beijing, 100084, China
[b] School of Business, Renmin University of China, Beijing, 100872, China

ARTICLE INFO

Article history:
Received 30 April 2015
Received in revised form 21 September 2016
Accepted 21 November 2016
Available online xxx

Keywords:
Competitive advertising
Keywords suggestion
Topic modeling
Factor graph model
Search engine advertising
Query logs

ABSTRACT

This study has proposed a topic based competitive keywords suggestion method called TCK to enhance search engine advertising. On the basis of query logs, the method explores the indirect associations between keywords and extracts the hidden topic information to identify competitive keywords. It can help advertisers not only broaden the choices of keywords but also carry out a competitive strategy for search engine advertising. Extensive experiments have been conducted to demonstrate the effectiveness of the proposed method. Results prove that the proposed method performs better than existing keyword suggestion methods, contributing greatly to the keyword suggestion advertising market.

1. Introduction

Search engine advertising (SEA), which is a primary marketing channel in online advertising nowadays [1–4], has been one of the main venues to compete for customers, due to its high targeting, relatively low cost, high flexibility and high variability [5]. Generally, advertisers bid on specific keywords related to their business, and then corresponding advertisements are displayed alongside the search results when users query those bidden keywords [6]. For example, a Nestle milk dealer in Shanghai (i.e., an advertiser) would like to bid on the keyword “Nestle Shanghai”, and post the corresponding ad alongside the search results returned by querying with the keyword, since the search engine users querying “Nestle Shanghai” would be possibly interested in the ad and click it. To a great extent, SEA can help business merchants effectively target potential consumers. The advantages of SEA prompt the emergence of keyword suggestion methods to help achieve the business goal.

Generally, a typical keyword suggestion method is to generate a set of keywords that relate to the seed keywords predefined by advertisers. The generated keywords can be considered as an expanded description of the business. Taking the Nestle milk dealer as an example, relevant keywords like “Nestle milk price”, “Nestle milk quality” and “Nestle milk formulation” may reasonably be suggested/recommended for advertising. In recent years, numerous research efforts have been devoted to the problem of keyword suggestion. A common technique is to conduct co-occurrence analysis based on query logs [1,7,8,9–12]. Keywords frequently co-occurring with the seed keyword in query logs are selected to conduct search engine advertising. As query logs can timely reflect users’ intentions [4,13], this category of methods have been adopted as the mainstream technique in search engine advertising marketplace, e.g., Google AdWords[1], Baidu Tuiguang[2], etc. In most cases, however, the keywords suggested by the co-occurrence based methods are quite popular and expensive to bid on. They limit advertisers to only a handful of keywords though the total number of query keywords is estimated to exceed a billion [14]. Actually a long tail of other candidate keywords, which is also relevant and occupies a large fraction of the total traffic [15], is ignored in the existing methods. Moreover, this kind of co-occurrence based methods cannot suggest keywords from the competitive perspective, which is rather significant in today's fierce competition environment [16–19]. Currently, the competitive keyword advertising is emerging as a new type of advertising which attracts more and more attention from advertisers. This

* Corresponding author.
E-mail addresses: qiaodd.12@sem.tsinghua.edu.cn (D. Qiao), zhangjin@rbs.org.cn (J. Zhang), weiq@sem.tsinghua.edu.cn (Q. Wei), chengq@sem.tsinghua.edu.cn (G. Chen).

[1] https://adwords.google.com/.
[2] http://fengchao.baidu.com/.

http://dx.doi.org/10.1016/j.im.2016.11.003
0378-7206/

strategy is meant to display ads alongside the search results of competitors. Fig. 1 shows a typical example in Bing advertising marketplace. As a tablet provider, Google has bidden on a competitive keyword "kindle fire", thus prompting a "Nexus" ad to the users querying "kindle fire". This strategy of advertising is effective and valuable on two folds. First, it can help advertisers find and attract more potential consumers to watch their ads, which is the nature of search engine advertising. A user who queries certain products is very likely to be interested in products with similar features/functions provided by competitors. Therefore, as shown in Fig. 1, it increases the possibility for consumers, who want to buy a tablet, to switch from Kindle Fire to Nexus after scanning the ads. Second, the competitive advertising strategy can help advertisers seize potential market shares from competitors. This is critical and important for advertisers to sustain a competitive advantage in the market [19,20]. An advertiser hence may be induced to directly advertise on the competitor's keywords [17] as shown in Fig. 1. This competitive advertising strategy helps Nexus obtain more by free riding on the market created by Kindle Fire. Therefore, leveraging competitive keywords to achieve the strategy of competitive advertising on search engines is very promising and important for advertisers.

To effectively achieve the goal of competitive advertising on search engines, a large number of competitive keywords should be provided to advertisers. However, advertisers can normally perceive a very small fraction of competitive keywords in their minds due to bounded cognitions. In addition, the existing co-occurrence based methods can hardly accomplish the task since competitive keywords rarely co-occur with the seed keywords in users' queries. To address the problem, this study proposes a novel method called TCK to automatically suggest keywords for competitive advertising. The proposed method captures the underlying topics hidden in query logs to recommend competitive keywords. Firstly, given a seed keyword, the method TCK could mine a broad enough set of candidate competitive keywords based on their indirect connections with the seed keyword. Secondly, each candidate is mapped into a topic structure extracted from the query logs. The mined topic structure is further combined into a factor graph model to finally extract a set of effective competitive keywords. Data experiments were conducted across various business domains to demonstrate the effectiveness of TCK.

The remainder of this paper is organized as follows. First, state-of-the-art of keywords suggestion methods is reviewed. Then the problem definition is presented. Following this, a novel method of competitive keywords suggestion, namely TCK, is proposed. Then experimental results are shown to reveal the outperformance of the proposed method. In the final section, discussions and limitations are concluded.

2. Literature review

To effectively help advertisers expose ads to potential consumers who are more dependent on search engines to assist purchase decisions [21,22], a plethora of research efforts have been devoted to the problem of keywords suggestion. According to various types of data source, the methods for keywords suggestion can be categorized into three streams, i.e., query log based, proximity based and meta-tag crawlers based methods [8].

2.1. Query log based keywords suggestion

In the branch of query log based methods, keywords are mainly suggested by conducting association/co-occurrence analysis in search engine query logs [3,11]. For instance, "iPad" is found to usually occur jointly with the keyword "Apple" in query logs. So the keyword "iPad" is considered to be highly relevant to the keyword "Apple" and therefore can be suggested to advertisers. The technique of co-occurring analysis has been widely used in many commercial keywords suggestion tools like Google AdWords[1] and Baidu Tuiguang.[2] Query logs from mainstream search engines can timely capture users' search intentions [21], which are valuable for commercial advertising and query log based keywords suggestion methods. Leveraging query logs, Zhang et al. proposed a bidding mechanism to recommend a group of relevant yet less hot keywords [12]. Chuklin et al. employed all the query phrases that contain the seed keyword to conduct query expansions [23]. In the context-based framework of query auto-completion, Jiang et al. extracted those queries matching the leading characters from query logs to generate query candidates [24]. Ranked by the conditional probabilities, Massoudi et al. selected the terms co-occurring with the original query in posts to augment queries [25]. Fuxman et al. considered the strong associations existing between

Fig. 1. An example of competitive search advertising and the translation.

Please cite this article in press as: D. Qiao, et al., Finding competitive keywords from query logs to enhance search engine advertising, Inf. Manage. (2016), http://dx.doi.org/10.1016/j.im.2016.11.003