September 22, 2017

Competitor Identification Based On Organic Impressions

Yunting Sun

Aiyou Chen

Abinaya Selvam

COMPETITOR IDENTIFICATION BASED ON ORGANIC IMPRESSIONS

Competitor identification is one of the first steps in developing an overall marketing strategy to increase return on investment for a particular brand. For example, in advertising and marketing research, a competitor set is used to construct benchmark advertisement spend (e.g., average media spend) and cost per click data, which businesses rely on in their media planning (aka, comparable company analysis). Competitor sets are generally not exposed to the advertiser directly. In media mix modeling, the search query volume associated with top competitors may be used to correct search-selection bias. In the acquisition of a small business, knowledge of a business's competitors can help to surface additional leads. That is, if a website is shown to perform well as an online advertiser, then the ability to identify similar competitors (even those competitors advertising on a different advertisement platform than the website) may be helpful in generating future customers.

Problems with Existing Methodologies

Various conventional systems attempt to identify competitors of a brand by analyzing their advertising data, such as pay-per-click campaigns sponsored by a search engine. However, these systems generally suffer from significant drawbacks limiting their capability to accurately identify a company's true competitors. For example, these systems require access to large volumes of advertising data associated with each competitor in order to make their decisions. Consequently, these systems fail to identify competitors that produce a low volume of advertising data, such as competitors newly entering the market or competitors choosing not to significantly invest in pay-per-click

campaigns. Also, it cannot identify competitors for brands that produce a low volume of advertising data, limiting the usefulness of the algorithm when the brands are new customers. Furthermore, advertising data tends to fluctuate over time, which causes these systems to produce inconsistent and unstable results. Moreover, such systems tend to produce irrelevant results because paid search queries are notoriously noisy and many are not directly relevant to the business at all.

Using Organic Search Data to Rank Competitors

A competitor identification system may identify the top competitors of a brand based on organic search data, instead of advertising data. In particular, the system identifies competitors by assigning a weighting score to each candidate website appearing in the search results produced from a selection of search queries associated with the brand website. Candidate websites that appear more frequently in the search results are assigned a higher weighting score than those that appear less frequently. That is, the system narrows down the list of possible candidate websites by calculating the conditional probability that the candidate website would appear in the organic search results when searching for keywords relevant to the brand.

A competitor identification system that is not based on advertising data has significant advantages over conventional systems. First, the system produces less biased results because it may identify competitors despite the competitor not paying (or paying very little) for advertisement space on the search engine. As such, the system includes a wider range of possible candidates in its analysis than the conventional systems. It also makes it possible to identify competitors for brands not paying or paying very little for advertisement on the search engine. Second, the system produces more stable results

because its algorithm is immune to the fluctuations in the advertisement data, which are commonly caused by the competitors frequently changing their advertisement bid amount and/or bidding keyword. Lastly, the system ensures that if a candidate website is identified as a competitor to the brand website, that the reverse is also true – i.e., the brand website is a competitor to the candidate website.

In one implementation, referred to herein as a "User-Independent" approach, the system includes operations that can run either in series or in parallel. Any of these operations may also run as a background process. The following operations may be performed by a competitor identification system that is implemented in hardware, software, firmware, or any combination thereof.

a) Identify Search Queries

The system identifies a set of relevant search queries that produced organic impressions or search results that led to the brand website b. The set of relevant search queries is denoted as Ω_b. The search queries are identified from a database of historical search data, that may be filtered based on any parameter or combination of parameters, such as a fixed time period (e.g., all searches within the past 3 months), a geographical area (e.g., all searches made by a device located in San Diego, CA), search volume, and/or a device type (e.g., all searches made with a mobile device). For example, the system may identify "tent," "camper," and "lantern" as a set of search queries that consistently produce www.examplecamping.com (a fictitious brand website) in the search results requested by mobile phones in San Diego, CA. These search queries are input data to the competitor identification system.

b) Weight Assignment

The system assigns a weight (w_i) to each search query associated with the brand website by dividing the number of impressions that query-i drives to the brand website by the total number of impressions that all relevant queries (determined in Step 'a') drive to the brand website. The weight measures the relevance of each query to the brand website.

c) Calculate Organic Impression Count

For each candidate competitor (c), the system determines the number of organic impressions (N_{ic}) from the i-th relevant query to the candidate website (c), which is normalized by N_i; where,

- 'i' is a search query (e.g., tent) from the set of relevant search queries.
- 'N_{ic}' is the number of impressions associated with a particular search query (i) to candidate competitor website (c) pairing. That is, this is the total number of impressions for the candidate competitor website produced from the search query (i).
- 'N_i' equal to $N_{ib} + \sum_c N_{ic}$ is the total number of impressions for the search query (i).

In other words, N_{ic} / N_i represents the proportion of impressions associated with the candidate competitor website for a given search query (i), among all the impressions produced from the search query (i).

d) Calculate Score

The system calculates the score for each candidate competitor (c), as it relates to the brand website (b), using the following equation:

$$\text{score}(c \mid b) = \sum_{i \in \Omega_b} w_i N_{ic} / N_i$$

This score indicates the degree in which the candidate competitor (c) is a competitor to the brand website (b).

e) Identify Competitors based on Rank

The system then ranks/sorts each candidate brand based on its score (c | b), as calculated in Step 'd', to determine the top competitors for the brand website (b). Top competitors may be identified as those candidates that fall within an upper range (e.g., top 2%, top 5%, top 10%, etc.) of the score listing.

In some implementations, the system also computes the reverse competitor score (i.e., score (b | c)) in order to filter candidate competitors (c) that do not consider the brand website (b) as an equal competitor. That is, if a candidate competitor (c) also considers the brand website (e.g., www.examplecamping.com) as a competitor, then it is more likely that the candidate competitor (c) is indeed a competitor of the brand website.

Example Operation

Fig. 1 illustrates a competitor identification system implementing the "User-Independent" approach to identify competitors for a brand website (e.g., www.flowers-A.com, a fictitious brand website). The bottom portion of the diagram shows the two most relevant search queries (e.g., flower delivery" and "online flowers") associated with the brand website, which were selected from historical search data in Step 'a'. The upper portion of the diagram shows the brand website and the two top candidate websites (e.g., www.flowers-B.com and www.flowers-C.com, each fictitious competitor websites) that the system identified based on the "User-Independent" approach.

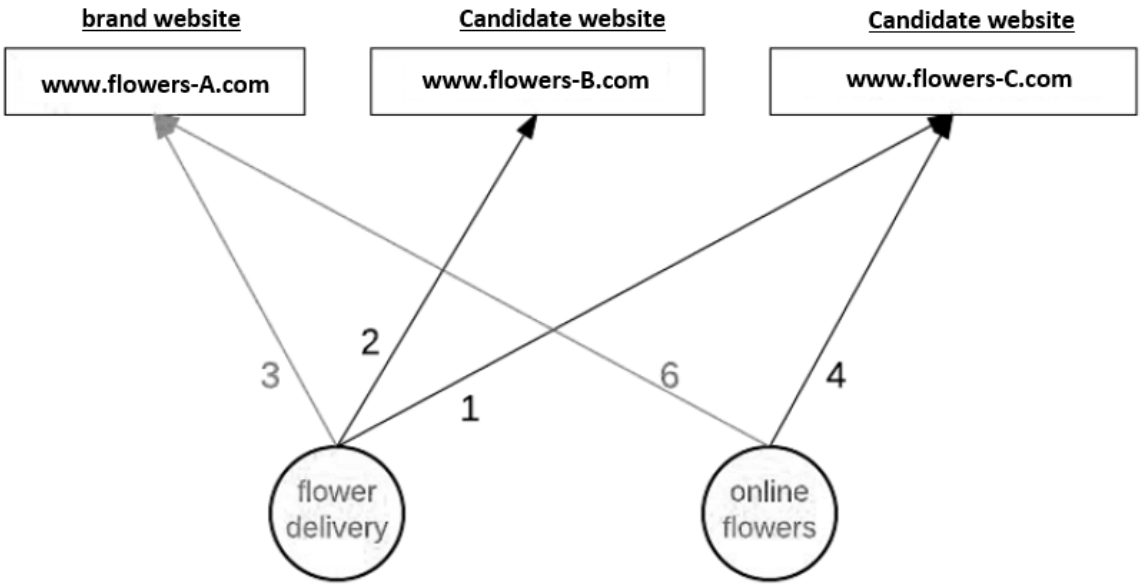

Fig. 1 - Diagram for www.flowers-A.com

For example, at Step 'b', the system assigns a weight to each relevant search query. If the brand website (www.flowers-A.com) receives 3 organic impressions from query "flower delivery" and 6 impressions from "online flowers," then the system would assign a weight of 3/(3+6)=1/3 to "flower delivery" and 2/3 to "online flowers." At Steps 'c'-'d', the system calculates the score for each search query (i). At Step 'e', the system ranks the scores to determine that the brand website's top competitors based on the "flower delivery" search query is (top to bottom): www.flowers-B.com and www.flowers-C.com. The system also determines that the brand website's top competitors based on the "online flowers" search query is (top to bottom): www.flowers-C.com. Therefore, based on an analysis of the rankings for both search queries, the system determines that the brand website's top competitors are www.flowers-C.com (#1) and www.flowers-B.com (#2).

Improving Accuracy with Organic Click Data

In an alternate implementation, referred to herein as a "User-Dependent" approach, the competitor identification system incorporates the organic click data into its impression-based algorithm to improve accuracy and confidence in the system. For example, the system may calculate a weighting score for each search query (i) by combining the impression data and the click data, where each click data point is worth a factor (e.g., 5 times, 10 times, etc.) more than each impression data point. This approach assumes that users who search for the brand website are likely to also click on impressions leading to the website for a competitor to the brand website. In another example, the system may separately rank the impression data and the click data for a given search query (i) before combining the two data sets.

Exemplary Uses of the Approach

a) Marketing Mix Modeling

The system may be used in research on Marketing Mix Modeling. This type of research is generally associated with marketing and optimization platforms designed to improve return on advertising expenditure.

b) Benchmark Metric Modeling

The system may be used to calculate a Benchmark Metric Model for a comparable analysis study, which companies generally rely on in their decisions of how much to spend on marketing in each media channel.

c) Lead Generation Modeling

The system may be used to construct a Lead Generation Model listing all the potential competitor companies associated with a particular brand website which is

shown to perform well as an online advertiser. The ability to identify similar non-client competitors may be helpful in generating future customers.

Abstract

This document describes a technique for identifying competitors of a brand based on organic impression data. The system identifies relevant search queries as ones that lead to organic impressions to the brand website based on historical data. The system assigns a weight to each search query based on the number of organic impressions it drives to the brand website. The system calculates a score for each candidate competitor as it relates to the brand website. The system ranks/sorts each candidate brand based on its calculated score to determine the top competitors for the brand website. The system may improve confidence in the result by computing the reverse competitor score in order to filter candidate competitors that do not consider the brand website as an equal competitor.