Towards Intent-Driven Bidterm Suggestion

William Chang[‡] [*], Patrick Pantel[†], Ana-Maria Popescu[†], Evgeniy Gabrilovich[†]
‡ USC Information Sciences Institute, 4676 Admiralty Way, Marina del Rey, CA 90292, USA
† Yahoo! Labs, 2821 Mission College Blvd, Santa Clara, CA 95054, USA
wchang@isi.edu | {ppantel | amp | gabr}@yahoo-inc.com

ABSTRACT

In online advertising, pervasive in commercial search engines, advertisers typically bid on few terms, and the scarcity of data makes ad matching difficult. Suggesting additional bidterms can significantly improve ad clickability and conversion rates. In this paper, we present a large-scale bidterm suggestion system that models an advertiser's intent and finds new bidterms consistent with that intent. Preliminary experiments show that our system significantly increases the coverage of a state of the art production system used at Yahoo while maintaining comparable precision.

Categories and Subject Descriptors: H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval
General Terms: Algorithms, Experimentation
Keywords: sponsored search

1. INTRODUCTION

Suggesting high quality bidterms to advertisers received significant attention in recent years, since relevant and profitable bidterms lead to improved ad clickability and increased conversion rates. Bidterm suggestion is similar to query expansion in mainstream IR [3] and in ad retrieval [9]. Existing approaches to bidterm suggestion rely on three main data sources: search engine results [1, 2, 6, 8], search engine logs [2] and advertiser bidding patterns [2, 5].

We describe a large-scale bidterm suggestion system that seeks to model the user's intent implicitly targeted by an ad, and finds new bidterms consistent with that intent. Ad intents are derived from a large collection of bidterm sets explicitly enumerated by advertisers (obtained from Yahoo's ad database), rendering our system close in spirit to [2] and [5]. In contrast to previous work that uses advertiser bidding patterns, our methods use second order co-bidding information and run on an industrial size database rather than on small test sets. We also demonstrate that the system increases the coverage of Yahoo's state of the art production system while maintaining the same precision.

2. INTENT DRIVEN MODEL

The intent (or information need) of a search engine user is expressed as short textual queries [7]. A typical *ad* consists of a set of bidterms (search terms purchased by the advertiser) whose underlying intents are likely to cause a user to be interested in the advertised product or service. Given a very large collection of such ads, we can begin to learn mappings between bidterms and the hidden intents.

We hypothesize that most ads try to capture a small set of intents (e.g., the purchase of a specific product or service), and therefore the set of bidterms in an ad is likely to be associated with the same hidden intent. Given a bidterm b, we model its set of hidden intents by the set of all other bidterms co-bidded with b (i.e., other bidterms occurring in the same ad as b). Some co-bidded terms are more discriminative than others so we weigh them by the strength of their association with b as follows.

Let $PMI(b) = (pmi_{b_1}, pmi_{b_2}, \ldots, pmi_{b_m})$ denote a pointwise mutual information feature vector, constructed for each bidterm b, where pmi_{b_f} is the pointwise mutual information between bidterm b and co-bidded term f:

$$pmi_{bf} = log \frac{\frac{c_{bf}}{N}}{\frac{\sum_{i=1}^{n} c_{if}}{N} \times \frac{\sum_{j=1}^{n} c_{bj}}{N}} \tag{1}$$

where c_{bf} is the number of times b and f are co-bidded, n is the number of unique bidterms, and N is the total bidterm occurrences.

By our hypothesis, two bidterms that capture the same intents will have more similar feature vectors than two bidterms that capture different intents. In this paper, we define the similarity between two bidterms b_i and b_j using the cosine similarity metric between their PMI feature vectors, $sim(b_i, b_j) = cosine(PMI(b_i), PMI(b_j))$.

Bidterm suggestion algorithm (IDBS)

Given an ad consisting of k bidterms, $\{b_1, b_2, \ldots, b_k\}$, we rank each bidterm b ever seen in our ad network by summing $sim(b, b_i)$ for $i = [1..k]$. We call this system *Intent-Driven Bidterm Suggestion* (IDBS).

The calculation of the similarity between all pairs of bidterms is computationally intensive. A brute force implementation is $O(n^2 f)$, where n is the number of bidterms and f is the size of the feature space ($f = n$ in our system). For a large real-life collection of bidterms, optimizations and parallelization are necessary.

Our optimization strategy follows a generalized sparse-matrix multiplication approach [10], which is based on the observation that a scalar product of two vectors depends only on the coordinates for which both vectors have non-zero values. Similarly, cosine similarity is determined solely by the features shared by both vectors. Since most of our feature vectors are very sparse (i.e., most bidterms never co-occur with any particular bidterm), the computation can be greatly sped up. Determining which vectors share a non-zero

*The research described herein was conducted while the first author was a summer intern at Yahoo!

WWW 2009, April 20–24, 2009, Madrid, Spain.
ACM 978-1-60558-487-4/09/04.

Table 1: Excerpt of IDBS output for a random ad.

Ad Bidterms	IDBS Suggestions
22 sail boat	sail boat sales
23 sail boat	old sail boat
24 sail boat	boat for sale by owner
25 sail boat	used boat for sale
capri	used yacht
catalina	sailing boat
catalina capri	used power boat for sale
catalina sail boat	sail boat for sale by owner

feature can easily be achieved by first building an inverted index for the features. The computational cost of IDBS is $\sum_i N_i^2$, where N_i is the number of vectors that have a non-zero i^{th} coordinate; this cost can be further reduced by thresholding low PMI values. On our datasets, we observed near linear average running time in the corpus size. Our MapReduce implementation is an extension of the approach of Elsayed et al. [4].

3. EXPERIMENTAL RESULTS

3.1 Setup

We randomly sampled 200 ads from Yahoo's sponsored search ad database, such that each ad had fewer than 50 bidterms. We scraped the Yahoo production bidterm suggestion system, a variant of [2], which generates up to 11 suggestions for each ad. We also generated up to 11 suggestions using *IDBS* and a baseline system, which was a simplification of *IDBS* where a bidterm is represented by a vector of ids of the ads to which it belongs (first order co-occurrence), similar to [2] and [5]. We extracted statistics to build our baseline and *IDBS* using Yahoo's ad network. We experimentally set the cosine thresholds for the baseline to 0.1 and for *IDBS* to 0.4. Table 1 shows sample IDBS output.

Each bidterm suggestion from each of the three systems was manually judged by two editors. The editorial guidelines asked the judges to mark a suggestion as correct *if any reasonable intent that would generate the suggestion as a search term matches any reasonable intent captured by the ad bidterms*[1]. Our intent-based judgments yielded high inter-annotator agreement: the kappa score is 0.86 on the 2,045 judged suggestions.

3.2 System Performance

Using the editorial judgments, we assess system performance using macro-averaged precision and coverage statistics. Coverage is defined as the ratio between the number of suggestions produced by a system and the maximum number of allowed suggestions (11 per ad in our setup, to match the Yahoo production system). Our hypothesis is that the Yahoo production system would generate much higher coverage than *IDBS* since it has access to many more features such as session logs and click data [2]. We show below, however, that *IDBS* adds value by making accurate suggestions when Yahoo's production system fails to do so.

Yahoo's overall precision was 0.87 ± 0.02 (95% confidence) with a coverage of 43%. *IDBS* achieved half the coverage for the same precision, however Table 2 shows that our system adds 13% coverage to the Yahoo system with little loss in precision. Table 2 also lists the added value of our baseline and that of *Combined*, a fourth system that combines the baseline system with our intent-driven system by interlacing the suggestions. *Combined* achieves the same precision as Yahoo's production system and increases its coverage by 13%. Figure 1 illustrates the tradeoff between precision and added coverage relative to the Yahoo production system.

[1]The guidelines only consider bidterm relevance and do not account for the cost of a bidterm.

Table 2: Added value vs. Yahoo's production system

	Baseline	IDBS	Combined
Precision	100%	83%	87%
95-Confidence	±0%	±5%	±5%
Added Coverage	6%	13%	13%

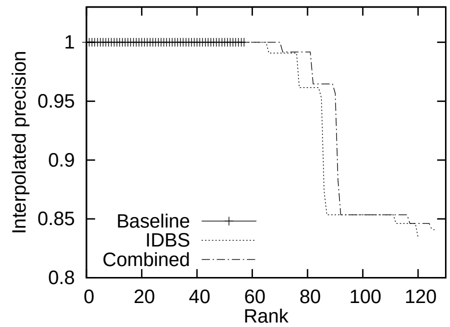

Figure 1: Interpolated precision vs. added coverage

4. CONCLUSION

This paper presents a large-scale bidterm suggestion system that models an advertiser's intent and finds new bidterms consistent with that intent. Preliminary experiments show that our system increases the coverage of Yahoo's production system by 13% while maintaining comparable precision.

5. REFERENCES

[1] V. Abhishek and K. Hosanagar. Keyword generation for search engine advertising using semantic similarity between terms. In *ICEC*, 2007.

[2] K. Bartz, V. Murthi, and S. Sebastian. Logistic regression and collaborative filtering for sponsored search term recommendation. In *Second Workshop on Sponsored Search Auctions*, 2006.

[3] C. Buckley, G. Salton, and A. Singhal. Automatic query expansion using SMART: TREC-3. In *Overview of TREC-3*, pages 69–80. DIANE Publishing, 1995.

[4] T. Elsayed, J. Lin, and D. Oard. Pairwise document similarity in large collections with MapReduce. In *ACL*, 2008.

[5] D. Gleich and L. Zhukov. SVD based term suggestion and ranking system. In *ICDM*, 2004.

[6] A. Joshi and R. Motwani. Keyword generation for search engine advertising. In *ICDM*, 2006.

[7] X. Li, Y.-Y. Wang, and A. Acero. Learning query intent from regularized click graphs. In *SIGIR*, 2008.

[8] V. M. Peter Anick and S. Sebastian. Similar term discovery using web search. In *LREC'08*, May 2008.

[9] F. Radlinski, A. Broder, P. Ciccolo, E. Gabrilovich, V. Josifovski, and L. Riedel. Optimizing relevance and revenue in ad search: a query substitution approach. In *SIGIR*, 2008.

[10] S. Sarawagi and A. Kirpal. Efficient set joins on similarity predicates. In *SIGMOD*, 2004.