UC San Diego
UC San Diego Electronic Theses and Dissertations

Title
An Advertiser Centered Approach to Improve Sponsored Search Effectiveness

Permalink
https://escholarship.org/uc/item/7n56q5x3

Author
Vattikonda, Bhanu Chandra

Publication Date
2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

An Advertiser Centered Approach to Improve Sponsored Search Effectiveness

A dissertation submitted in partial satisfaction of the
requirements for the degree of Doctor of Philosophy

in

Computer Science

by

Bhanu Chandra Vattikonda

Committee in charge:

>   Alex C. Snoeren, Chair
>   Kirill Levchenko
>   George Papen
>   Stefan Savage
>   Geoffrey M. Voelker

2015

The Dissertation of Bhanu Chandra Vattikonda is approved and is accept-

able in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2015

iii

DEDICATION

To my mother.

EPIGRAPH

There are three classes of people: those who see,
those who see when they are shown,
those who do not see.

Leonardo da Vinci

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

As I finish my dissertation, I cannot help looking back on my graduate school life. I am confident that I have emerged wiser at the end of these six years. I am also grateful to have shared this time with a set of smart, inspiring and amiable people. I would like to take this opportunity to thank them.

First and foremost, I would like to thank my advisor Professor Alex Snoeren for the extraordinary support and guidance he gave me during my time at UCSD. His ability to crystallize ideas and provide sharp focus is amazing. His support and steady hand during the tough times have been invaluable in getting me over the line.

Saikat Guha provided me with exciting opportunities at Microsoft and has spared his time at all times of the day to brainstorm ideas. He was always willing to ask the tough questions and gave the belief that answers could be found.

Several current and past SysNet members have contributed to my progress. George Porter was a strong force behind Practical TDMA for Datacenter Ethernet. His support was critical to the success of the project both in terms of ideas and in setting up experiments. Vacha Dave has been an ever present sounding board for the past three years. Her constructive feedback strengthened my work on sponsored search. During the initial years of my PhD program I was fortunate to have worked with Kevin Webb from whom I learnt good programming practices. I would also like to thank Amin for setting me up with a chance to work on datacenter networks. Andreas Pitsillidis, David Wang shared their experiences in evaluating career prospects and have been part of many a friendly banter. Sambit Das, Rishi Kapoor, Harshit Chitalia made the long nights spent on course projects enjoyable.

I would also like to thank fellow systems and networking group members Tristan Halvorson, Neha Chachra, Chris Grier, Danny Huang, Lonnie Liu, Patrick Verkaik, Alex Rasmussen, Nathan Farrington, Gjergji Zyba, and office mates Frank Uyeda, Malveeka

Tewari, Sivasankar Radhakrishnan, Feng Lu, Ryan Huang and Robert Liu for providing a welcoming working environment in the department.

During my PhD, and for years before that, my family has been a continuous source of support. I dedicate this dissertation to my mother, who made incessant efforts to ensure my progress and well-being. I thank my father, sister and brother-in-law for their love, support and encouragement. My wife, Shruti, has been my constant companion for the last six years.

This dissertation would not have been complete without the support from my committee members: Kirill Levchenko, George Papen, Stefan Savage, Geoff Voelker. I thank each one of them for devoting their time to review my work.

Lastly, I thank my friends in San Diego, Atlanta and Bay Area for providing an outlet outside of work.

Chapter 3, in part, is a reprint of the material as it would appear in Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD15. Bhanu C. Vattikonda, Santhosh Kodipaka, Hongyan Zhou, Vacha Dave, Saikat Guha, Alex C. Snoeren. The dissertation author was the primary investigator and author of this paper.

Chapter 4 includes material that is submitted for publication as "Empirical Analysis of Search Advertising Strategies". Bhanu C. Vattikonda, Vacha Dave, Saikat Guha, Alex C. Snoeren. The dissertation author was the primary investigator and author of the paper.

VITA

2007	Bachelor of Technology, Indian Institute of Technology Kanpur
2007–2009	McKinsey and Company
2009–2015	Research Assistant, University of California, San Diego
2011	Master of Science, University of California, San Diego
2015	Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

Bhanu C. Vattikonda, Vacha Dave, Saikat Guha, Alex C. Snoeren. Empirical Analysis of Search Advertising Strategies. Under submission.

Bhanu C. Vattikonda, Santhosh Kodipaka, Hongyan Zhou, Vacha Dave, Saikat Guha, Alex C. Snoeren. Interpreting Advertiser Intent in Sponsored Search. To appear in Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD15.

Kevin C. Webb, Bhanu C. Vattikonda, Kenneth Yocum, and Alex C. Snoeren. Scal- able Coordination of a Tightly-coupled Service in the Wide Area. In Proceedings of the First ACM SIGOPS Conference on Timely Results in Operating Systems, TRIOS 13, pages 8:18:15, Farmington, Pennsylvania, USA, 2013. ACM.

Bhanu Chandra Vattikonda, George Porter, Amin Vahdat, and Alex C. Snoeren. Practical TDMA for Datacenter Ethernet. In Proceedings of the 7th ACM European Conference on Computer Systems, EuroSys 12, pages 225238, Bern, Switzerland, 2012. ACM.

Bhanu C. Vattikonda, Sambit Das, and Hovav Shacham. Eliminating Fine Grained Timers in Xen. In Proceedings of the 3rd ACM Workshop on Cloud Computing Security Workshop, CCSW 11, pages 4146, Chicago, Illinois, USA, 2011. ACM.

ABSTRACT OF THE DISSERTATION

An Advertiser Centered Approach to Improve Sponsored Search Effectiveness

by

Bhanu Chandra Vattikonda

Doctor of Philosophy in Computer Science

University of California, San Diego, 2015

Professor Alex C. Snoeren, Chair

Sponsored search is a form of advertising where advertisers pay a search engine to show their ads on the search engine results page. The ads, also known as sponsored results, are chosen and presented to the user in response to a user query alongside organic search results. Sponsored search holds the promise of allowing advertisers to precisely target their ads to the large number of users of a search engine. The rise in use of search engines and the opportunity they provide to target ads using fine-grained criteria has led to a 20% annual growth in sponsored search revenues over the last decade.

The targeting criteria chosen by an advertiser for their ads allow a search engine

to deliver the ads to the right users. At the same time, it also puts the onus on the advertiser to identify the right ad targeting criteria. In this dissertation, we take a two-pronged approach to improve the effectiveness of sponsored search in delivering value to advertisers and improve the quality of results shown to users.

First, we improve the ability of a search engine to interpret the targeting criteria specified by the advertiser. As part of the targeting criteria advertisers submit ad keywords which specify the user queries for which they would like to advertise. We leverage the search engine itself to interpret an ad keyword by submitting the ad keyword as an independent query. Using the search results of the ad keyword associated with an ad we determine if the ad is suitable for the original user query.

We then analyze the effectiveness of different targeting strategies followed by advertisers. We develop a simple metric called net acquisition benefit (NAB) that admits comparisons between the efficacy of different ad targeting strategies. Using this metric, we conduct the first large-scale measurement of different targeting strategies used by advertisers—measured in terms of incremental conversion gains. Considering data from a month in early 2015, we employ NAB to identify cases where these targeting strategies are justified.

Chapter 1

Introduction

The World Wide Web has grown at a rapid pace over the last fifteen years. Web search engines like Google, Bing, and Yahoo! [2, 8, 24] have come to play a critical role in how users find content and navigate around the web. Search engines help users find content by presenting them with a list of web results in response to a query they submit. These results, which are chosen based on their relevance to the user query, are also referred to as organic results. A 2012 Pew Internet survey finds that 91% of the adults online use search engines to find information on the web [16]. The net result has been an explosion in the number of searches that users make, with Google alone serving more than a trillion search requests in 2014 [9].

Search engines have discovered that, as gatekeepers of the web, they can also connect advertisers with users. Sponsored search, where ads (also referred to as sponsored results) are shown along with the organic results, allows advertisers to advertise their products and services to users of a search engine. The increasing role of search engines in enabling users to find content on the web and the ability of advertisers to run highly targeted ads has led to a tremendous growth in sponsored search revenues. Sponsored search attracted more than 50% of the $49.5 billion [58] spent on online ads by advertisers in 2014, growing at an annual rate of 20% over the last decade [58, 59].

Sponsored search generally involves advertisers expressing the user queries for

which they would like their ads to be shown when they list their ads with the search engine. The words or phrases used by advertisers to specify the queries for which they would like their ads to be shown are referred to as "ad keywords". Advertisers also express other targeting parameters and the price they are willing to pay for their ads to be shown when listing their ads with the search engine. Search engines then choose the ads to be shown to the users based on the relevance of an ad to the query issued by the user and the price advertiser is willing to pay. Sponsored search is seemingly advantageous to advertisers and users alike. From the perspective of the advertiser, they can show their ads to users who have already expressed their intent through the query they submit. From the user perspective, the ads are less intrusive than traditional ads because they are relevant to the query the user submitted to the search engine. For example, a user searching for "shoes" may be presented with an ad by Nike making it attractive to both Nike and the user.

It is important, then, for the search engine to deliver ads to the right users by taking the targeting criteria specified by the advertiser and user satisfaction into account. Search engines balance these goals by evaluating the relevance of an ad to the user query before deciding whether or not to show the ad to the user. I none of the ads listed with the search engine are relevant to the user query then it may not show any ads to the user [38]. In general, search engines use supervised machine learning algorithms [56, 85] to score an ad on it's relevance to a user query. These supervised machine learning algorithms are trained on large human-labeled datasets.

It is important for the advertisers to specify the right targeting criteria for their ads to maximize the impact of their ad campaigns. In this process advertisers are aided by the targeting options offered by the search engines to craft their ad campaigns and the fine-grained data that the advertisers themselves can collect on users visiting their websites. These capabilities allow the advertisers to run highly targeted campaigns and

then accurately track the performance of their campaigns to measure their effectiveness.

In this dissertation we first address the challenges faced by a search engine in identifying the right ads to be shown to users by taking advertiser goals into consideration. We identify features which capture the advertiser intent expressed by their choice of ad keywords that can be used by the supervised machine learning rankers to improve relevance ranking. Then, we evaluate the effectiveness of different targeting strategies chosen by advertisers.

1.1 Approach

In this dissertation, we explore ways in which we can take advantage of the role of advertisers in sponsored search to improve the accuracy of ads being shown to users and improve the performance of ad campaigns being run by advertisers. One of the strengths of sponsored search is the ability of advertisers to run targeted campaigns using fine-grained criteria. Advertisers can specify the ad keywords which determine the user queries for which their ads are shown. They can also specify other parameters like the device type, user demographics, time of day, region based on which their ads are shown to the users.

We first explore ways to improve the accuracy of the search engine in interpreting the targeting criteria specified by the advertisers. One of the main targeting parameters in sponsored search is the ad keyword. The ad keyword is used to target ads towards particular user queries. It also provides the advertiser with a unique advantage—it is completely hidden away from the user allowing the advertisers to express their intent freely. Accurately interpreting the ad keyword associated with an ad allows the search engine to measure the relevance of the ad to a user query. The challenge in interpreting the ad keyword however, is that it is often very short—just a few words long [56]. We overcome this challenge by using the capabilities of the search engine itself. Using years

of prior research search engines have built capabilities to interpret a query and identify web results which are relevant to the query. These search results, therefore, capture the meaning behind the query that has been submitted to the search engine. We use the search results associated with the ad keyword to identify features which capture the advertiser intent expressed through ad keywords. We then use these features to improve the relevance ranker used by a large search engine to eventually determine the ads to be shown in response to their user queries.

We then explore the successfulness of the advertisers in taking advantage of the targeting criteria offered by sponsored search. Using logs from the same search engine we examine the effectiveness of different targeting strategies used by the advertisers. The logs, which represent the traffic served by the search engine, can be aggregated to identify different traffic slices, each of which represents a particular advertising strategy. We then define a simple metric—net acquisition benefit (NAB)—that allows us capture the profits an advertiser receives from a particular slice of traffic. We analyze the performance of different advertising strategies by identifying different traffic slices that these strategies represent and then comparing the profitability over those slices of traffic.

In particular, we analyze three different targeting strategies—cannibalization, poaching and mobile ad extensions. Cannibalization is an advertising strategy where the advertiser chooses to advertise on queries where they are present in organic results as well, thereby creating the possibility that they could be cannibalizing clicks on their organic results through advertisements. Poaching is a strategy in which an an advertiser tries to poach users clearly looking for a competitor. For example, when the retailer J.C. Penny advertises on the ad keyword "macys" trying to attract users clearly searching for Macy's, then we say J.C. Penny is trying to poach users. Finally, we evaluate the benefit of targeting mobile devices by analyzing the profitability of advertising on mobile devices and comparing it with traditional advertising on computers.

1.2   Contributions

In this dissertation we take an advertiser-centered approach to improving the effectiveness of sponsored search. We propose new techniques to enable a search engine to interpret the targeting criteria specified by the advertisers and provide tools that help advertisers improve their targeting criteria.

We identify features that can be used by the relevance ranker of a search engine to score the ads accurately on their relevance to an incoming user query. The relevance scores then allow the search engine to avoid showing irrelevant ads to the users which would lead to a poor user experience and wasted ad expenditure by advertisers. Moreover, accurate delivery improves the effectiveness of the ad campaigns.

Finally, we present a simple metric—net acquisition benefit (NAB)—which can be used by ad agencies and other entities managing the ad campaigns for advertisers to estimate the profitability of a particular ad campaign without access to sensitive financial information from advertisers which they may be reticent to share. Using the metric we study three common campaign strategies used by advertisers on a large search engine: cannibalization, poaching, and ad extensions. Considering data from a month in early 2015, we employ NAB to identify cases where these campaign strategies are justified. Advertisers and ad agencies can replicate our methodology to apply it to other strategies of interest.

1.3   Organization

The remainder of this dissertation is organized as follows. Chapter 2 provides an overview of the sponsored search ecosystem and the background necessary for this dissertation.

Chapter 3 describes our technique of using existing capabilities of a search engine

to identify features which can capture advertiser intent behind their targeting criteria. We describe the challenges associated with measuring the relevance of ads to user queries and the challenge of interpreting advertiser intent expressed through the ad keywords which are only a few words long. We evaluate the benefits of using features capturing advertiser intent using data from a large search engine.

In Chapter 4, we introduce a new metric that can be used to compare the profitability of different advertising strategies without access to advertisers' private financial information. Using month long data from a large search engine we evaluate the effectiveness of different advertising strategies that are used by advertisers.

Finally, Chapter 5 summarizes the dissertation with a discussion of the main results and opportunities for future work.

Chapter 2

Sponsored Search

In this chapter we present an overview of sponsored search and related concepts. As we walk-through different aspects of the ecosystem, we also discuss the vast body of related work in sponsored search. Sponsored search is an interaction of three players: users, advertisers, and the search engine [32, 48, 61]. We examine the role of each player in the following sections.

2.1   Users

Users visit a web search engine to find content on the web related to a particular query that they have [36, 37, 72]. In response to the query, users expect the search engine to show them content that is strongly related to their query and satisfy their informational need. Search engines generally show two types of results to the users, sponsored and organic, ordered using proprietary search engine algorithms. In general, a search engine chooses organic results based purely on their relevance to the user query. On the other hand, advertisers pay the search engine to have their sponsored results listed. The sponsored results thus, are chosen based on their relevance to the user query and the revenue opportunity they provide to the search engine. A typical results page returned by the search engine to users is shown in Figure 2.1. It has been observed that users have a preference for organic results over sponsored results expecting certain editorial integrity

Figure 2.1. Results page of a typical search engine has organic and sponsored search results.

in the organic results [48, 61]. However, typical web search engines derive their revenue from sponsored results [51, 58].

The organic results are displayed with a title and are accompanied by small snippet of text, chosen by the search engine, which describes the content of the destination web page. Each sponsored result is designed by the advertiser (or a third party working on their behalf) paying for the sponsored result. As shown in Figure 2.1 a sponsored result

typically includes title and text describing the products and services being advertised. For certain types of queries, like knowledge queries [83], the information they are looking for is provided on the results page itself. For example, if a user submits the query "president of usa" to a typical web search engine like Google, Bing or Yahoo! [2, 8, 24] the search engine presents the name of the current President of the United States in the results page. In most other cases, the users choose to click on a particular result—sponsored or organic—and navigate to the website hosting content of interest to the user.

Once the user navigates to the destination website, through organic or sponsored result, they could take a range of actions. For example, on informational websites like Wikipedia [22] they could be seeking information on a certain topic. On e-commerce websites they could choose to purchase a product. Advertisers using sponsored search to advertise themselves are often interested in getting the users to finally purchase their products or services.

2.2 Advertisers

Advertisers interested in marketing their products and services run ad campaigns with the search engine. In ad campaigns, advertisers specify the creatives (i.e., content of the ad) they would like to show to the users, various targeting criteria and the price they are willing to pay if the user clicks on their ad.

2.2.1 Ad Creative

The content of an ad that is shown to the user is referred to as ad creative. The creative is designed by the advertiser or an agency working on behalf of the advertiser. It includes at a minimum a title, the advertiser's domain name, and two short lines of descriptive text, typically rendered in blue, green, and black, respectively as illustrated in Figure 2.1. The ad creative also contains one or more Uniform Resource Locators (URLs)

— called destination URLs — the user should be directed to if the ad is clicked. Search engines increasingly support ad extensions that allow advertisers to include additional information or actions in the rendered ad [7, 20]. A call extension, for instance, allows the advertiser to provide a phone number; when the ad is shown on a mobile phone, the extension is rendered as a button that invokes the phone dialer when clicked by the user. Other common extensions used by advertisers are location, sitelink and product extensions [7, 20].

Advertisers can list static or dynamic ad creatives with the search engine. In the case of dynamic ads, advertisers specify the template of the creative and have the search engine generate the final creative based on the different targeting parameters specified along with the ad [5, 18]. The dynamic creative generation is used to customize the ad being shown in response to user query. The dynamic ad creative generation also allows the advertiser to insert the targeting parameters into the destination URL. When the user clicks on the ad and navigates to the advertiser's website, the advertiser can identify the ad which led the user to their website and associate the ad with other user activity data collected on their own website. The user activity data can then be used for comprehensive campaign analytics. The ad creatives themselves are created by advertisers using semi-automated techniques and different ad creatives are tested for their effectiveness [1]. Over time, the promising ads are chosen and refined.

In this dissertation, we do not examine the strategies advertisers use to design the creatives and the impact these strategies have on the performance of the ad campaigns. When we examine the impact of different targeting choices advertisers make, we assume that the advertisers have already chosen the optimal creatives for a particular advertising strategy that they are executing. It is known that top advertisers use millions of different creatives in their ad campaigns [34] which are tested iteratively for their effectiveness [1].

2.2.2 Targeting

Advertisers express their intent to advertise for certain user queries by selecting individual words or phrases that must be present in the (normalized) user search query. These words or phrases are also known as "ad keywords". The ad keywords are the basic mechanism through which advertisers can target their ads. Often ads are shown to users when the ad keyword matches the user query — resulting in what is an exact match. But, it is challenging for an advertiser to enumerate all the queries for which they would like to advertise [40, 56]. Hence, a typical web search engine provides the option of matching the ad keywords with a broader range of relevant user queries, e.g., alternative spellings, synonyms, etc., potentially resulting in what we term a broad match [40, 56]. Broad match requires the search engine to identify semantic similarities between user query and the ad keywords when choosing the ads to be shown to users. A mismatch between the query and the ad keyword can result in an unsuitable ad being shown to the user potentially leading to wasteful spend by advertiser while at the same time hurting user experience. For example, showing an ad by "Virgin Mobile" when the user query is "virgin river utah" is detrimental to advertiser goals and user experience. So, it is important for the search engine to accurately capture intent expressed by advertiser through their choice of ad keywords. We explore this topic further in Chapter 3.

The ad keywords are frequently identified using keyword suggestion tools that the search engines themselves provide where the search engines make keyword recommendation using query log mining [67]. Early keyword suggestion tools started with seed keywords provided by the advertisers. Then, using frequent queries containing the seed terms, advertisers were provided with ad keyword suggestions. Advanced keyword suggestion mechanisms have become more prevalent where semantic similarities of the words are taken into account [67]. In one approach, search engines build a directed graph

between related terms using search engine results and user click behavior to identify words and phrases with semantic similarity [47, 67]. Using this graph, the search engines then recommend keywords to the advertisers based on the seed terms provided. The choice of ad keywords to advertise for, has direct impact on the performance of an advertising campaign. The final choice of ad keywords is made by the advertisers and they can choose the types of keywords to advertise for by choosing the seed terms when using the keyword suggestion tools. The techniques used to find ad keywords themselves are not a topic of study in this dissertation. We assume that advertisers do intend to advertise for the ad keywords they specify in their targeting criteria. However, in Chapter 4, we provide tools for advertisers and ad agencies to evaluate the benefit they would derive by advertising for different types of ad keywords. We evaluate a few different keyword selection strategies using data from a large search engine making the assumption that advertisers intentionally bid on the ad keyword we evaluate.

Advertisers may further target their ad by device type, geographic region, time of day and user demographics. Targeting is quite fine-grained in practice, with top advertisers managing tens of millions of ad keywords [34]. Search engines recommend advertisers to tailor their campaigns based on these targeting parameters [13]. Studies by advertising agencies have found that the effectiveness of ad campaigns can depend on the time of day [13], location [66] and age of the users [69]. These agencies report that targeting ads based on fine-grained parameters can improve the performance of ad campaigns. In conducting these studies, the ad agencies often use various performance metrics like click-through rate, which is the number of clicks on an ad as a fraction of the number of times the ad is shown and cost per acquisition, which is the total advertiser spend divided by the number of resulting user acquisitions—users who pay for advertisers' products or perform some other desirable action. We discuss the shortcomings of the metrics currently used by advertisers to evaluate the performance of ad campaigns in

Chapter 4. We then evaluate the effectiveness of different targeting strategies used by advertisers.

2.2.3 Bid Value

When listing ads with the search engine, advertisers are required to specify the price they are willing to pay if their ad is clicked. This price is called the bid value. Whether the ad is shown in response to the user query and if shown, the position of the ad in the results page, depends on the bid value specified by the advertiser. The problem of deciding the bid value to place with the search engine is a complex optimization problem for each advertiser [52, 53, 79]. Advertisers are known to use a variety of techniques to optimize the bids that they place with the search engine. The authors of [39] show that sophisticated advertisers' bids are based on a host of complex factors including clicks that they obtain from advertising, conversions, profit margins, number of impressions etc. [35, 79]. Bidding strategies also rely on historical prices to adaptively identify the keywords to bid on and the corresponding bid values [79]. Yet, very often advertisers use different heuristics and place constant bids on a group of keywords. This reliance on heuristics in identifying bid values is likely due to a lack of enough data to fully optimize the bids [77].

The final price an advertiser pays for a click is determined through an auction mechanism that is designed to encourage advertisers to bid the maximum amount they are willing to pay for a click [50]. The amount an advertiser is charged if an ad is clicked is based on the next lower-ranked bid in a form of generalized second-price auction [50]. The truth revealing nature of the bids that advertisers place [27] means that a rational advertiser would expect returns equal to (or more than) the bids they place with the search engine. We use the assumption that bid values are a reflection of advertiser profit margins to measure the profitability of different advertising strategies in Chapter 4.

2.2.4 Campaign Analysis

The purpose of ad campaigns is obviously to drive revenue to the advertiser. Sponsored search and online advertising in general holds the promise of making available large-scale, fine-grained user-activity data to advertisers allowing them to accurately measure the performance of their ad campaigns.

As discussed earlier, advertisers use custom destination URLs to determine the ads through which users were obtained. Advertisers then log user activity on their own website to keep track of user actions and identify ways in which ad campaigns and customer experience can be improved in the long run [55, 73]. Sophisticated search engines assist advertisers in monitoring the effectiveness of their campaigns by providing support for analytics. The analytics tools allow advertisers to track desirable user actions and identify the returns on their advertising campaigns. In particular, advertisers use analytics to measure the success of ad campaigns in leading to user acquisition—also referred to as conversions.

The advertisers can inform the search engine when a conversion occurs by embedding JavaScript code provided by the search engine on the page on which the conversion happens. The JavaScript code directs the browser to contact the search engine's server with a unique user identifier which can then be used to link the conversion event to previous user actions [60] on the search engine results page. The advertisers can then track campaign performance along different types of conversions by passing an opaque tag to the conversion JavaScript.

2.3 Search Engine

The search engine is the central player in the sponsored search ecosystem connecting users searching for content to advertisers advertising their products and services

to the users. In this section, we present an overview of how search engines determine the organic and sponsored results to be presented to the users. Every time a user makes a query, a "search impression" is created which contains the organic and sponsored results chosen by the search engine.

2.3.1 Organic Results

The organic results presented by search engines are determined through proprietary heuristic algorithms like PageRank [72]. The listed websites do not pay the search engine for placement or user clicks they receive from being listed by the search engine. The organic results are ranked based purely on the search engine's determination of their relevance to the user query. The search engines use a variety of information sources to determine the most relevant results for a particular query, for example, past click behaviors on results, user demographic information and location of the user issuing the search query [28, 29, 64, 65].

2.3.2 Sponsored Results

When the user makes a query, a search engine has to decide which sponsored results (if any) should be displayed along with the organic results. These sponsored results are the main source of revenue for a typical web search engine [58, 60] and sustain the infrastructure needed to serve user queries. However, an aggressive pursuit to maximize revenue every time a user makes a query (i.e., every search impression) could hurt user experience and, in the long run, the search engine's popularity and profitability as sponsored results are generally perceived to degrade user experience [62]. Thus, in some cases, it may be desirable to show few or even no ads if they do not meet a certain relevance threshold [38]. For example, if the query is "weather", an ad for "cold weather jackets" might be a match and occasionally generate revenue but it is not relevant to the
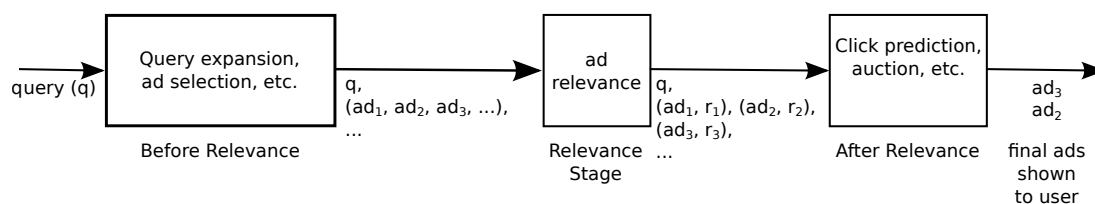
Figure 2.2. Sponsored results are chosen by search engine using a multi-stage pipeline.

user query if the user is simply seeking the current temperature—and especially if it is currently warm. In this case not showing the ad would be the prudent choice. Another canonical example is a so-called navigational query like "macys", where ads other than those from the retail chain Macy's could elicit a negative user response.

One of the key challenges for a search engine is thus, to balance user experience and revenue goals. A typical search engine uses a multi-stage pipeline to identify ads relevant to a user query. However, in this dissertation we abstract it into three stages as shown in Figure 2.2. Before the relevance stage, the query (q) is expanded to create expanded queries which are related to the original query depending on the match type indicated by the advertiser [40]. A lot of work has been done to interpret and expand the user query [40, 57]. For example, user click behavior [46, 57, 81] and electronic dictionaries [84] have been used to enhance the query and expand it. These techniques rely on the fact that relevance of words to a particular query is correlated to the user

Query expansion techniques have also used categorization of the query and topical information to achieve improvements [41, 68, 81] in ranking documents. These approaches use human-judged datasets to classify web pages into hierarchical categories. These categories are then used to find ads that may not be textually similar but belong to same category as the query. For example, the query "sneakers" would be categorized as belonging to "Apparel/Footwear" category allowing the search engine to match it to an ad for "shoes" whose destination URL is categorized similarly. Bennett et al. [33] use documents classified under the Open Directory Project (ODP) [4] to train a classifier and

show that categorizing web pages can improve ranking of relevant documents. Broder et al. [40] use search engine results (and the category classification of the results) to create an augmented query and then select ads using the augmented query. Each expanded query is then used to select candidate ads (ad_1, ad_2, ...) to be further evaluated. The goal at this stage is to identify all ads that are potentially related to the set of expanded queries.

In the relevance stage each ad picked thus far in the pipeline is evaluated for relevance to the original query. While the previous stages choose ads that are related to the expanded queries, this stage performs deeper inspection of the relevance of each ad (ad_1, ad_2, ...) to query. The output of this stage is a score for each ad (ad_1, ad_2,...) on how relevant it is to query (i.e., generate r_1, r_2,...). For example, an ad for "nike shirts" would be scored higher for the query "shirts" than would an ad for "jackets". To do so, the relevance stage trains a learning ranker [85] using a labeled training set of (query, ad) pairs and features computed for each (query, ad) pair. Basic features used to measure the relevance of an ad to the query capture the textual overlap between the creative and query [76]. Along with these basic features, search engines use past historical click-through rates, translation models [56] and category overlap between the ad and the query [40]. The trained ranker is then deployed to measure the relevance of an ad to the query.

After the relevance stage, the ads pipeline involves estimating the probability (click-through rate) that an ad would be clicked and conducting the second-price auction. The relevance of ad to query is a factor when estimating the click probability [**?**]. The probability estimate allows the search engine to calculate the expected revenue to be derived by showing a particular ad. These predictions are made using information about the ad from previous impressions of the ad [78]. For rare or new ads where such information is not available, information from semantically similar ads is used [49]. In

both the cases, the relevance of ad to query can be used as one of the features to predict the click-through-rates. So, while click-through-rate estimates do filter out ads, accurate relevance scores complement these efforts. Subsequently, most search engines rank ads by the product of click probability and advertiser bid value in an attempt to maximize expected revenue for the second-price auction [40, 56].

2.4 Summary

Sponsored search allows advertisers to target users who are searching for content related to a specific query. From the ads listed by the advertisers, search engines generally choose ads to be shown to users by considering the relevance of ads to the user query and the price advertisers are willing to pay to show their ads [80].

This dissertation presents techniques to improve the accuracy of the relevance ranker used by a search engine. In Chapter 3, we describe our methodology of using existing capabilities of a search engine to identify features which allow it to best capture advertiser intent. An improved relevance ranker would allow the search engine to satisfy advertiser goals and improve user satisfaction.

We then evaluate the successfulness of the advertisers in exploiting the targeting parameters offered by the search engine in Chapter 4. We introduce a simple metric—NAB—which allows us to measure the profitability observed by an advertiser from a particular slice of traffic they receive from the search engine. We use the conversion signals that advertisers send to the search engine to track their campaigns to measure the success of a campaign. Using NAB we then compare the profitability of three different advertising strategies: cannibalization, poaching and ad extensions.

Chapter 3

Advertiser Intent in Sponsored Search

As we discussed in the previous chapter, search engines balance user experience and revenue goals when choosing the ads to be shown to users. A key component in achieving this balance is to determine the relevance of a potential ad to the incoming user query. In this chapter, we discuss the ways in which search engines identify ads relevant to the incoming user query and our contribution in improving the relevance scoring mechanism. The ability to accurately determine relevance of ads to user query allows the search engine to enhance user experience and also better serve advertisers.

The search engine we studied, considers an ad to be relevant to a user query if the following four components are aligned [15]: *i*) query; i.e., what the user is looking for, *ii*) ad creative; what is being promised to the user, *iii*) ad landing page; the web page actually delivered to the user if the ad is clicked, and *iv*) ad keyword; which indicates the type of traffic the advertiser seeks to attract.

Interpreting a user's query is hard mainly because it is very short: 2.5-words long on average [83]. Over time search engines have incorporated large amounts of associated metadata such as the user's search pattern within a session [28] and click-through data [57], as well as employed various other query augmentation techniques [31, 70, 71, 75, 87] in attempts to accurately interpret user queries. Modern web search engines return highly accurate results which is reflected in the high user satisfaction [16].

JustFab™ Women's Shoes - JustFab.com
Ad www.justfab.com/ ▾
4.4 ★★★★⯪ rating for justfab.com
Get Your **Shoes** Handpicked for You! 2 Pairs for $39.95 + Free Shipping.
Deals On Heels: $39.95 - Take Our Fashion Quiz - Check Out Our Top Shoes

Figure 3.1. A typical ad create has a title, description, display URL, and one or more destination URLs to which the user is directed upon clicking the ad.

Similarly, various techniques have been used to understand the creative and landing page associated with the ads to improve ad selection [45]. Such approaches are effective because the creative is often a good reflection of what is being advertised and the landing page offers a rich set of features. However, the creative itself offers very little information (a typical creative is a few tens of characters, see Figure 3.1) and landing pages are known to be noisy [45]. These approaches are even more challenging to apply in the case of broad match. Broad match, as mentioned in the previous chapter, is a targeting criteria in which an advertiser allows a search engine to match the ad keywords to user queries which are semantically similar to the ad keywords. In the case of broad match the query and advertisement may not be textually similar. For example, an ad bidding on "sneakers" might be quite relevant for the query "shoes" but there may be little textual similarity between them.

In this chapter, we complement prior approaches by interpreting the ad keyword as well. Unlike the creative and landing page, both of which are to be displayed to the end user, the ad keyword represents an unconstrained opportunity for the advertiser to be direct about their desires without concern of offending or dissuading the user. Hence, we argue that it represents a very strong signal that should be mined to the fullest extent. We build on the fact that—as discussed above—search engines are good at interpreting a query. In particular, we determine advertiser intent by submitting the ad keyword to the search engine and use organic results that the search engine returns to provide additional context with which to interpret the ad keyword. Specifically, given an advertisement to

be scored for relevance against a particular query—we denote this as a (query, ad) pair in the remainder of the chapter—-we send the ad keyword associated with the ad to the search engine and use the top organic results returned to get additional information about the ad keyword. We then use features extracted from both these results and the organic results for query itself to measure the similarity between the user intent behind the query submitted and the advertiser's intent behind targeting the ad keyword.

We consider introducing two complimentary sets of features: 54 features that can be generated using just the query issued by the user and another 21 that require information capturing user intent, which we get by using organic results generated for the user query. We evaluate the benefits of adding each of these feature sets by comparing the performance of the resulting ranker to the best previously published baseline [76] and the production system at a large search engine. We achieve a 43.2% improvement in precision-recall area under the curve (AUC) over the baseline and 2.7% improvement over the highly engineered production system.

3.1 Motivation

The goal of the relevance stage is to compute the relevance of the ad to query in each (query, ad) pair that has been selected by the previous stages. While the stages prior to the relevance stage focus on casting a wide net to rapidly identify as many related ads as possible, the relevance stage uses a broader range of features to measure the relevance of ad to the query.

3.1.1 Capturing advertiser intent

The learning ranker used to compute a score measuring the relevance of ad to query in a (query, ad) pair is trained using a set of features computed for each pair. For the task of feature computation, the key fields available in the ad are: *i*) creative (Figure 3.1),

ii) ad keyword and *iii*) landing page.

Features that are currently used by the production system include text similarity features between query and these fields along with other external sources of information—including the click-through rate of the ad from past impressions [56].

In this chapter we explore the benefit of expanding the ad keyword and using the resulting features to measure relevance of ad to query. Our key insight here is that the ad keyword captures advertiser intent more accurately than the creative itself. The ad keyword is the only field in the entire ads pipeline through which an advertiser can explicitly express the type of traffic that they would like to attract. Other attributes of the ad, like the creative and landing page, are seen by the user which could prevent the advertiser from freely expressing their intent. Hillard et al. [56] observe that, for example, an ad for "limo rentals" would be quite relevant to a user query for "prom dresses". An advertiser might, thus, list an advertisement for "limo rentals" and bid for the keyword "prom dresses". In the absence of an understanding of the ad keyword, such an ad would be considered completely irrelevant to the user query "prom dresses". If one had a way to identify that that prom dresses and limos are frequently used together, however, better relevance scores could be computed improving the quality of ads delivered.

We choose to solve the problem of evaluating the relevance of an ad to query at the relevance stage because it is expensive to perform a deep evaluation of all possible variations of the query computed by the query expansion algorithms employed in the stages prior to the relevance stage. However, ad keywords associated with the ads provide us with a defined set of keywords on which deeper analysis can be performed.

3.1.2   Broad match opportunity

The importance of understanding the ad keyword is highlighted in the case of broad match when the query is expanded before being matched against the ad keyword,
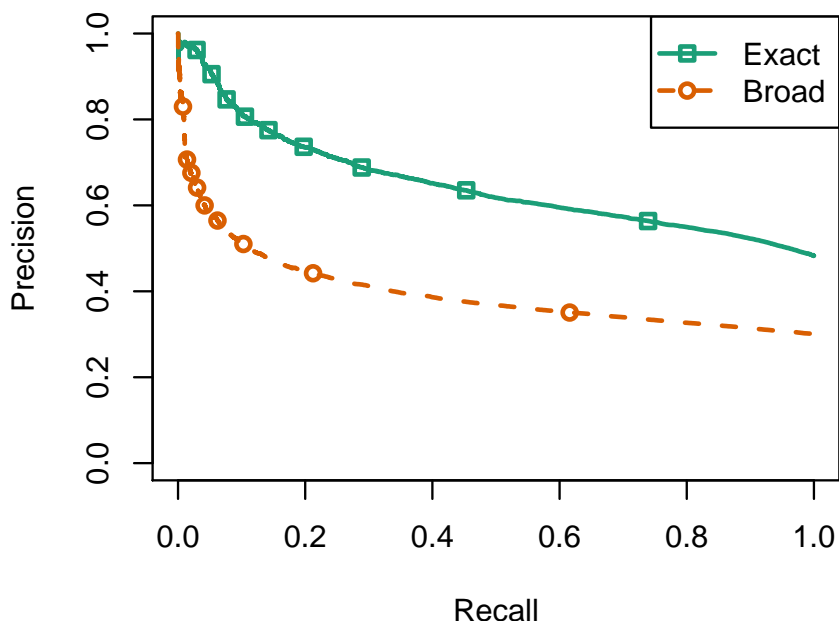
Figure 3.2. Ads chosen through exact match can be scored more accurately with previously published baseline features than those chosen through broad match.

increasing the likelihood that the user's intent behind the query may not align with advertiser's intent of targeting the ad keyword. To illustrate this, we compare the performance of a previously published baseline relevance ranker on (query, ad) pairs matched using exact match to those matched using broad match. The performance of the relevance ranker is evaluated using precision/recall values over a hold out validation set. Using precision/recall values allows for evaluation of the ranker independent of other factors like click probability and bid values which play a significant role in the final decision to show an ad to the user.

Figure 3.2 plots the precision/recall curve obtained by scoring (query, ad) pairs in the validation set using a relevance ranker trained on baseline features [76] (detailed in Section 3.2.3). High precision values indicate that a large fraction of the ads being selected are relevant—enhancing the user experience. Whereas, high recall values indicate that a greater number of relevant ads are being selected—improving revenue

opportunity for the search engine. Unsurprisingly, the baseline features that measure the textual similarity between only the query and the ad to determine relevance do not work as well in the case of broad match because broad match includes (query, ad) pairs which may be semantically similar but not textually similar.

3.2   Methodology

In this section we present our methodology for using the capabilities of the search engine to create features that represent the advertiser intent. Using the search engine to interpret a query, which is usually short, allows us to leverage years of research that has been done to interpret a short piece of text and identify web results. Using the organic web search results corresponding to the ad keyword provides us with detailed information about advertiser intent. We use these features to improve the performance of a learning ranker used to measure the relevance of an ad to a query. For comparison purposes we build upon a baseline ranker using the 19 features described by Raghavan et al. [76].

This section starts with a description of the datasets, learning ranker, and baseline features that we use. We then introduce our additional features, starting with features that can be extracted using just the user query along with organic results for the ad keyword. We then explore features that can be computed using organic results for the query which act as a proxy for user intent along with organic results for the ad keyword allowing us to capture the overlap between advertiser and user intent.

3.2.1   Data overview

The key datasets in our system are the training and validation sets comprising of (query, ad) pairs. Each ad listed with the search engine has an associated ad keyword indicating the type of queries from which the advertiser would like to attract traffic. An ad also has a creative and a landing page associated with it. Our dataset contains the title,
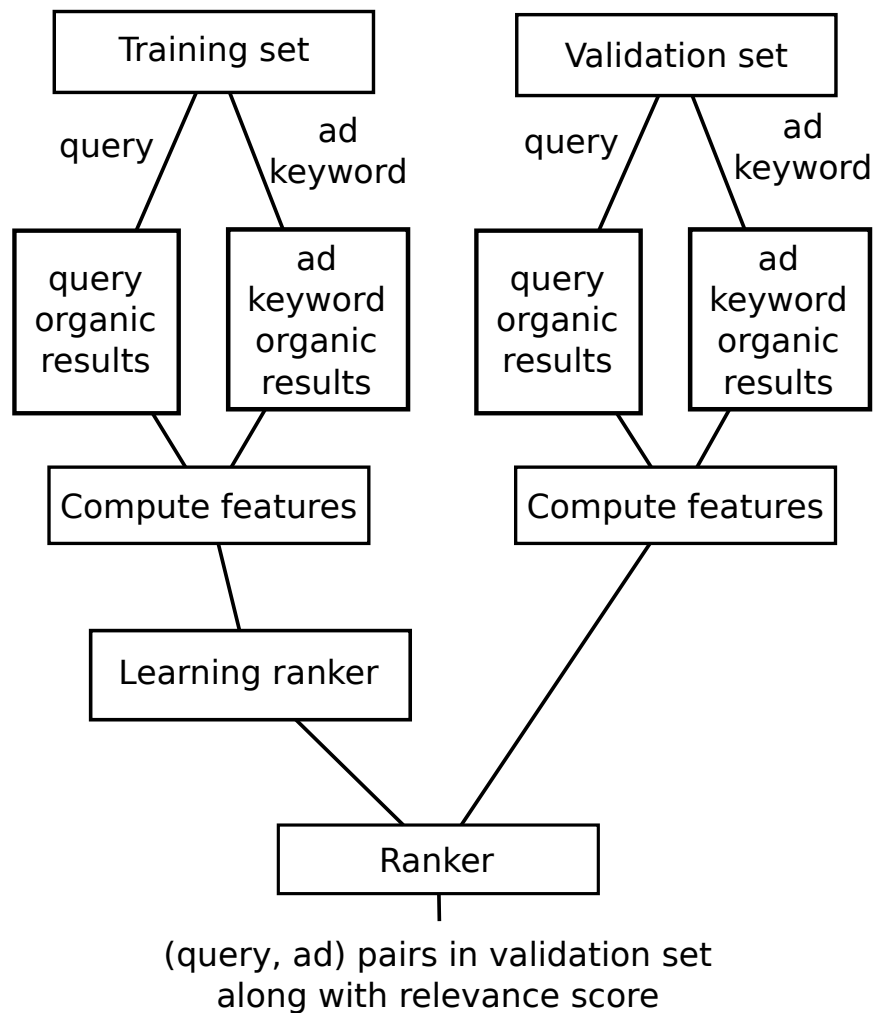
Figure 3.3. The learning ranker used for relevance ranking is trained on features computed over a training set. The accuracy of the resulting trained ranker is measured using a holdout validation set.

text, display URL associated with the creative and the title and a short snippet which represents the landing page. As shown in Figure 3.3, in the first phase, we obtain the top 40 organic results associated with the query and the ad keyword in an approach similar to the one taken by Broder et al. in [40].

Each of the 40 organic results returned for a query submitted to the search engine has the following fields associated with it: *i*) title of the web page as shown in the

web results, *ii*) snippet, a small piece of text displayed on the results page providing a summary of the result page itself, *iii*) description, a small piece of text from the web page which most accurately describes the web page, *iv*) ODP category, the ODP [4] category to which the result belongs, and *v*) URL of the result.

To clean up these fields, we remove stop words [6] and stem the title, snippet, and description of each result using the Porter stemmer [74]. We then concatenate all the titles, snippets, descriptions of results associated with each query to create a bag-of-words representation.

3.2.2 Ranker

For each (query, ad) pair in the training and validation sets, we compute features as described in the following sections. We train the LambdaMART learning ranker [43] on the features obtained over the training dataset. LambdaMART has been shown to be very effective in solving real-world ranking problems [33, 42]. LambdaMART is known to be robust to features that take a range of values and produces a tree-based model. In our evaluation, the algorithm is trained at a learning rate of 0.12, with 120 leaves and 2,000 trees. The model produced by the ranker can be used to determine a ranked list of the features on which the ranker was trained. We use the ranked list of features in the model generated by LambdaMART to identify the importance of features that we introduce in this chapter.

3.2.3 Baseline features

We use the features described by Raghavan et al. [76] as a baseline against which to compare the gains offered by the additional features that we propose. Raghavan proposes 19 features: query length and 6×3 features obtained by computing the following: *i*) word unigram overlap, *ii*) word bigram overlap, *iii*) character unigram overlap,

Table 3.1. Query features computed for each (query, ad) pair using creative, landing page, and organic results for the ad keyword associated with the ad.

Feature Type	Details	Count
Creative	ad title ∩ ak. titles	6
	ad text ∩ ak. desc.	6
	ad text ∩ ak. snip.	6
Landing Page	title ∩ ak. titles	6
	snippet ∩ ak. desc.	6
	snippet ∩ ak. snip.	6
Query	query ∩ ak. titles	6
	query ∩ ak. desc.	6
	query ∩ ak. snip.	6
Total		54

iv) character bigram overlap, *v*) ordered word bigram overlap, and *vi*) cosine similarity between the query and each of the title, description, and display URL of the ad creative for each (query, ad) pair.

We compute each of the overlap features as the overlap coefficient of the corresponding sets, computed as:

$$\text{overlap}(X, Y) = \frac{|X \cap Y|}{\min(|X|, |Y|)}.$$

For example, word unigram overlap coefficient between "black shoes" and "shoes at contoso inc" would be 0.5.

3.2.4 Query features

The first set of features that we introduce rely on using only the information that can be computed based upon the ad and the query in each (query, ad) pair. In the next section we discuss more advanced features that can be computed using organic results for the query which allows us to better measure the similarity between user and advertiser

intent.

For each ad, we compute features to determine whether the ad creative and landing page are consistent with the ad keyword the advertiser supplies. Specifically, we compute the same six similarity features as Raghavan et al. [76], but between the organic results returned for the ad keyword and aspects of the ad creative and landing page. For the creative, we compare the ad title to the search result titles, and the ad text to both the search result description and snippets. For the landing page, we compare its title to the search result titles, and the snippet to both result descriptions and snippets.

We further compute features that measure the similarity of query to the results of searching for the ad keyword. We compute the same six similarity features, but this time between the query and the titles, snippets and descriptions associated with the ad keyword search results, respectively. These features are easy to implement because organic results for ad keyword can be precomputed, and when the query is received in the online system, feature construction is a matter of computing overlap features. While in each case we use the same overlap and cosine similarity features as in Section 3.2.3, there is no limitation against using other similarity measures like Jacquard index or edit distance. In total, we add the 54 features shown Table 3.1. We call these query features.

3.2.5   Query search features

We also consider features that can be computed if the ads pipeline can interpret user intent in the same way the search engine does to generate organic results for query. We use the organic results generated for query as a proxy to capture user intent in much the same way as we use organic results for the ad keyword to capture advertiser intent. Once we have the organic results for the query, we compute six overlap features for each pair of titles, snippets and descriptions obtained from results of query and ad keyword.

Table 3.2. Query search features constructed using organic results for query and ad keyword capturing user and advertiser intent.

Feature Type	Details	Count
Query Search	query titles ∩ ak. titles	6
	query desc. ∩ ak. desc.	6
	query snip. ∩ ak. snip.	6
Category overlap	query ∩ ak. categories	1
Domain count	domain in query results	1
	domain in bk results	1
Total		21

Category feature. Each web page in the search index is classified using ODP data [4] into categories by an internal classification engine at indexing time as described by Bennett et al. [33]. Each organic result (which is chosen from the index) is thus classified into one of the 219 categories at the top two levels of the hierarchy. For the query and ad keyword, we obtain the categories to which the corresponding organic results belong. We then compute the cosine similarity between the categories of the two organic results:

$$\text{score} = \frac{\sum_c n_{qc} \times n_{kc}}{\sqrt{\sum_c n_{qc}^2} \times \sqrt{\sum_c n_{kc}^2}},$$

where, n_{qc} and n_{kc} are the number of times a result belonging to category c is present in organic results for the query and ad keyword respectively.

As has been argued by Broder et al. [40], the category feature allows us to identify scenarios when the query and ad keyword might not have a strong overlap but are relevant to each other because they belong to the same category. For example, for the query "shoes", the ad keyword "sneakers" does not result in a text overlap, but is very relevant.

Domain features. The last set of features that we introduce captures the presence of the ad domain (e.g., `contoso.com`) in the organic results for query and ad keyword. The ad domain of an ad is determined to be relevant to a query if the ad domain is present in the organic results for the query. Similarly, the the presence of the ad domain in organic results for the ad keyword indicates that the ad domain is relevant to the type of traffic the advertiser wants to attract by bidding on the particular ad keyword. We introduce two features to capture the relevance of ad domain to query and the ad keyword. Specifically, the features are computed as number of times the ad domain is present in organic results for both the query and ad keyword.

In sum, we call these additional $18 + 1 + 2 = 21$ features query search features and summarize them in Table 3.2. Together with the 54 features computed in Section 3.2.4 they form a total of 75 features that we consider.

3.3 Evaluation

In this section we quantify the improvements in relevance ranking obtained by incorporating advertiser intent. We present the gains in precision and recall over both a published baseline [76] and the production system for a large search engine.

The baseline system [76] is rudimentary and captures only the similarity between query and the ad creative for each (query, ad) pair. However, as far as we are aware the work by Raghavan et al is the best-performing published system. The features that we introduce use much richer information from the organic results to capture advertiser and user intent. As a result, we achieve extraordinary gains over the baseline. Our benefits over the production system—which uses hundreds of features—are less dramatic, but still significant in practice.

Table 3.3. Relative (%) improvement in precision-recall AUC over baseline for different types of ads.

| | | Ground-Truth Scores | | | |
		All	3+	4+	5
Query Features	All	27.5	70.3	60.1	21.6
	Exact	16.6	37.6	35.7	11.5
	Broad	32.2	90.3	82.6	58.8
Query Search Features	All	43.2	103.5	122.7	79.6
	Exact	22.7	50.0	67.9	52.2
	Broad	55.7	165.3	228.6	145.8

3.3.1 Datasets

The learning ranker that we use is trained using a sample of 1.28 million hand-scored (query, ad) pairs drawn from the ground-truth the production system uses. The scores range between one and five with five representing high relevance between the query and the ad. For offline testing of the model, the holdout validation dataset has 320,000 similarly sampled (query, ad) pair scores. The training and validation datasets are retrieved from the ad corpus using information retrieval methods used by the stages prior to the relevance stage [40]. They contain queries from all search frequency deciles.

We obtain the organic results corresponding to the query and the ad keyword by submitting each of them to the search engine. We also use the ODP [4] categorization of the organic results returned for query and ad keyword.

3.3.2 Baseline comparison

We start by evaluating the gains that the new features provide over the baseline [76]. For now, we consider an ad to be not relevant to the query if the (query, ad) pair is judged to be a one. We return to consider more stringent cutoffs in Section 3.3.2
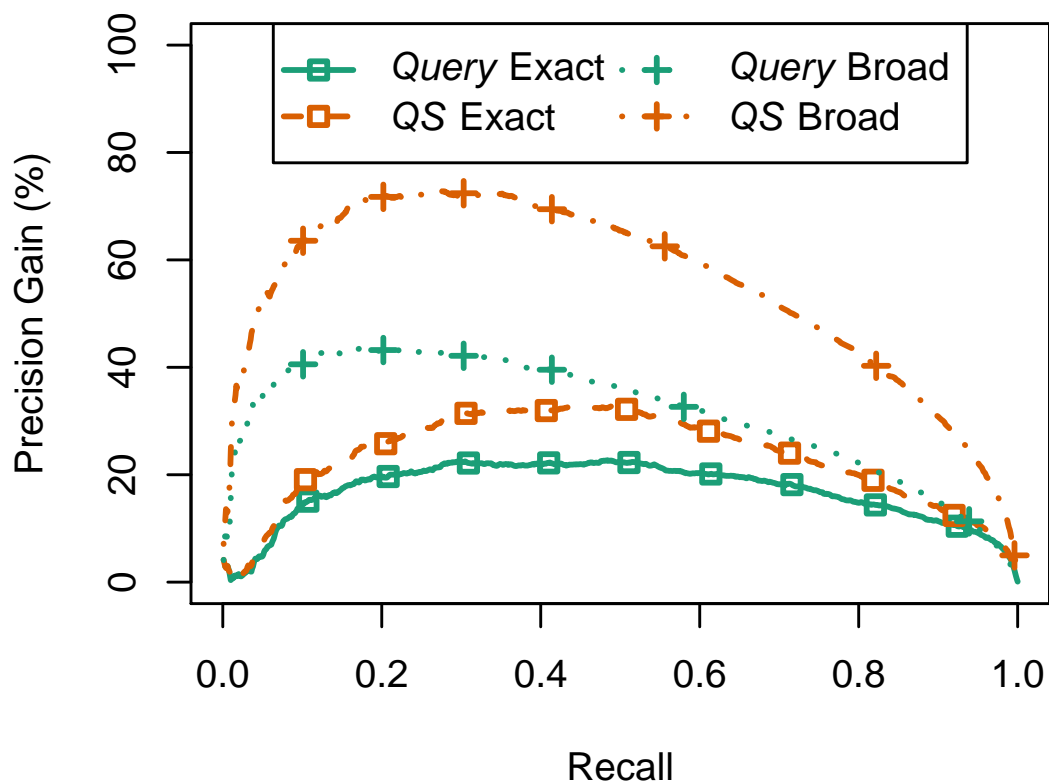
Figure 3.4. Improvement in precision at different recall values using the ranker trained on baseline and query features over a ranker trained only on baseline features.

Query features. As discussed in Section 3.2, for each (query, ad) pair in the training and the validation sets, we add 54 new features. Among these, 18 features compute the similarity of query to the titles, snippets and descriptions associated with organic results for the ad keyword. Similarity between creative, landing page and ad keyword organic results is captured in another 36 of these features. We compare the performance of a ranker trained with these features to a ranker trained using only the 19 baseline features.

The "Query Exact" and "Query Broad" lines in Figure 3.4 show the relative improvement in precision at different recall values over the baseline obtained by using the ranker trained on query features for (query, ad) pairs matched through exact and broad match, respectively. The relative change in the area under curve for the precision-recall

curves is presented in the All column of Table 3.3.

The results show that adding information from organic results for the ad key-word provides a large improvement in precision over the baseline. Also, note that the improvement is higher for (query, ad) pairs matched through broad match. Intuitively, the improvement is because information from organic results for the ad keyword increases the possibility of a match between the query and ad when the ad is relevant to the query. In the case of exact match, the baseline features already capture the overlap between the ad and the query because the creative likely contains the ad keyword and hence contains the query.

Query search features. Here we measure the benefit of adding new features which capture the similarity between user and advertiser intent by using organic results for the query and ad keyword. We introduce a total of 21 features which capture the similarity between user intent and the advertiser intent along with 54 features introduced earlier.

As described in Section 3.2 we use query organic results as a proxy for interpreting the user intent. Note that in an online system generating both organic and sponsored results, the results for query themselves may not be needed, instead techniques used to process the query and identify organic results would be enough to interpret the query.

The organic results of a query give us titles, snippets and the descriptions asso-ciated with the query. For each of these fields we compute six features which measure similarity to the corresponding field from organic results for ad keyword—giving us a total of 6×3 new features. In addition to these 18 features, we also add one feature which captures the similarity between the categories of results for the query and ad keyword. Two additional features use organic results to capture how relevant the ad domain itself is to the query and ad keyword.

The "QS Exact" and "QS Broad" lines in Figure 3.4 show the improvement in

precision at different recall values using the model obtained by training the ranker with query search features (along with query and baseline features) for exact and broad match types, respectively. These results show that using search results for the ad keyword to interpret advertiser intent provides a large improvement in the accuracy of the relevance ranker over using just the baseline features. Again, the relative change in precision-recall AUC is presented in the All column of Table 3.3.

Identifying good ads   While achieving high overall precision-recall numbers is important to distinguish between relevant and irrelevant ads, a good relevance ranker should be especially adept at identifying high-quality ads accurately. So, techniques which lead to gains in overall precision should not negatively impact the ability of the search engine to identify ads scored three or higher. The ability to identify good ads accurately is a desirable feature for the model used by search engines because it allows search engine to show good ads to the users and not just suppress bad ads. While suppressing bad ads is good for user experience, a ranking model which does not identify good ads would lead to lower revenue.

We measure the ability of the new features to distinguish between ads scored three or higher and ads scored lower than three. We consider ad scored one or two as irrelevant to the corresponding query in the (query, ad) pairs. The relative improvement in precision-recall AUC over the baseline model is shown by the 3+ column in Table 3.3. We perform a similar analysis for ads scored four or higher, and five by considering the remaining ads to be irrelevant respectively.

We see that the understanding of advertiser intent works especially well for identifying high-quality ads. Ads scored four or better and those scored three or better see greater improvement than the overall pool of ads. The behavior is expected because the ad keyword organic results of good ads are more likely to have pattern overlap with

Table 3.4. Relative (%) improvement in precision-recall AUC over the production ranker for different types of ads.

		All	3+	4+	5
Query Features	All	0.5	1.0	0.8	0.6
	Exact	0.4	0.6	0.1	0.1
	Broad	0.6	1.8	2.7	1.9
Query Search Features	All	2.7	3.7	5.7	8.9
	Exact	0.7	1.6	2.6	5.4
	Broad	3.9	7.3	12.3	13.2

the query organic results.

3.3.3 Production comparison

The results in the previous sections demonstrate that the new features provide significant gains over a published baseline proposed by Raghavan et al. [76]. It stands to reason, however, that highly engineered production systems employed by major search engines exhibit better performance and may, in fact, already consider many of the features we suggest. In this section, we consider the gains our new features bring to the production pipeline of a large search engine. We add the above mentioned features to the existing production features and train the production ranker on the combined feature set. The production features capture a variety of attributes about the query and ad, including quality of the ad domain, historical click-through-rates of the ad and landing page attributes [15].

Precision gain over the production system at different recall values with the introduction of query features for exact and broad match are shown using "Query Exact" and "Query Broad" lines in Figure 3.5. The results show that there is a small improvement in precision-recall curves due to the introduction of query features. The relative change in the precision-recall AUC is presented in the All column of Table 3.4. The gains are smaller than those over baseline but are nevertheless significant in a production system.
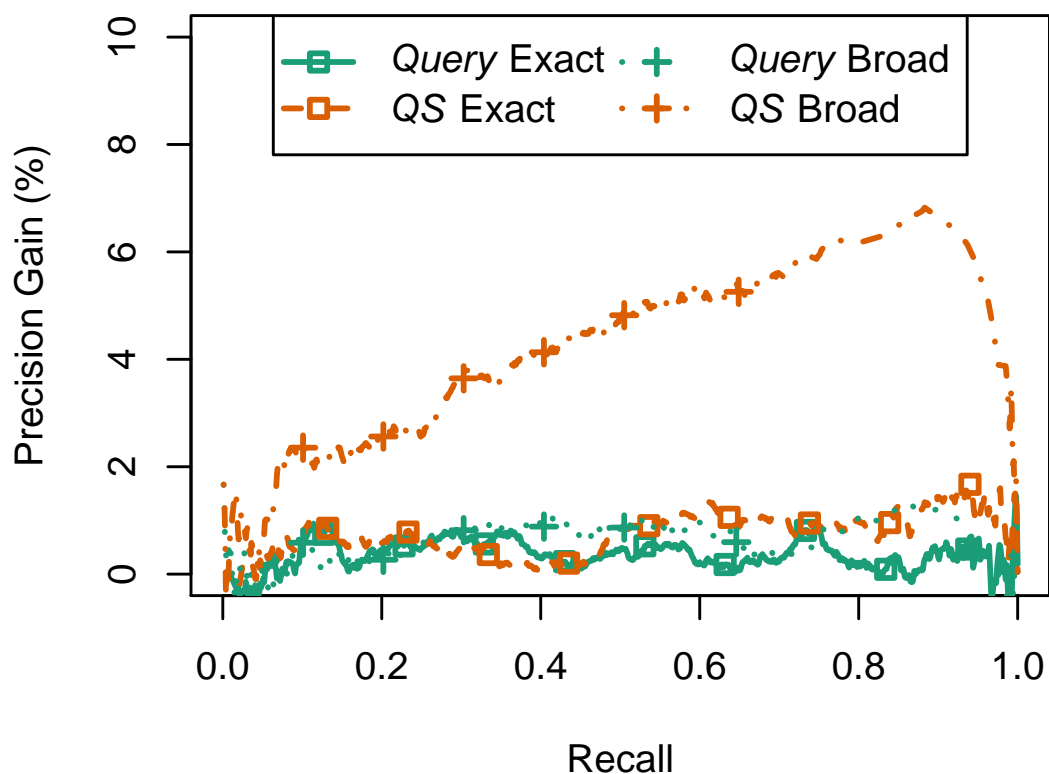
Figure 3.5. Improvement in precision at different recall values using the ranker trained on baseline, query and query search features over a ranker trained only on current production features.

As before, the new features perform best for highly-relevant ads in broad searches.

Moreover, we find that the production ranker improves significantly with the addition of query search features. The improvement in precision at different recall values is shown in Figure 3.5 by the lines labeled "QS Exact" and "QS Broad". The relative improvement in AUC is presented in the All column of Table 3.4. As before, the new features work even better for high-quality ads.

To capture improvement in the accuracy of the ranker, we measure precision-recall values at max F-score. Table 3.5 shows the precision and recall values for the production ranker and the ranker trained using query search features at max F-score.

Table 3.5. Gains achieved in precision, recall and max F-Score compared to the production system of a large search engine.

Ranker	Precision	Recall	Max F-Score
Query	-1%	+2.3%	+0.3%
Query Search	+1.7%	+2.9%	+2.2%

3.3.4   Feature importance

The value of the features we introduce can also be seen in the importance given to them in the final model generated by the LambdaMART ranker. The tree based model that is produced after training the ranker allows us to determine the ranked list of features. The importance of query search features is very clear in the ranking of the features. We find that the tree created by training the ranker on a combination of existing features and query search features ranks the following as the top three features: *i*) ad domain count in query organic results, *ii*) ordered bigram overlap between snippets of organic results for query and ad keyword and *iii*) ad domain count in ad keyword organic results. These features rank higher than many other highly engineered features.

The benefits of adding query features are smaller than those when compared to the baseline as we mention in Section 3.3. But, the value of interpreting the ad keyword is reflected in the ranking of the new features. Two of the 54 features are among the top 30 features in the final tree produced by the ranker. These are: *i*) word unigram overlap between query and snippets in organic results for ad keyword and *ii*) order word bigrams between query and titles of the organic results for ad keyword.

3.4   Discussion

The gains we see over the production system are naturally lower than the gains over the baseline. However, they are significant in production [14]. Moreover, the

features that we propose can be obtained using the datasets and learning experience already at the disposal of the search engine.

Quantifying the impact of these improvements on the revenue of the search engine and user experience is complicated. We see significant improvements in ranking accuracy. However, it may be possible that the ads which have been more accurately scored will not be shown to the user for a host of other reasons including low click probability, low bid values by the advertiser, and so on. In such a scenario, improvements to the relevance ranker would not enhance the user experience. Increase in recall values, however, would lead to the identification of more ads that are truly relevant to the user query, leading to greater competition in the auction—and higher revenue for the search engine.

3.5   Summary

The ad keyword is the only field in a sponsored search ad that allows an advertiser to express the type of traffic that they would like to attract. At the same time the ad keyword, like the user query, is very short which makes the task of interpreting advertiser intent hard. In this chapter, we leverage the ability of modern search engines to interpret the intent behind a user's query to similarly understand the advertiser's intent as conveyed in the ad keyword.

We make three main contributions in this chapter. First, we show that using organic search results to expand the ad keyword provides us with a rich source of information from which we can interpret advertiser intent. Second, we identify the features to be extracted from these organic results which can be used to improve the relevance ranker. Among these, 54 features can be implemented with few changes to the existing system and another 21 features would require user intent information as well. Finally, we evaluate the benefits of these features using training and validation datasets of 1.28M and 320,000 samples sampled from a corpus of ground-truth (query, ad)

pair scores respectively. We show that using features which capture user and advertiser intent leads to 43.2% improvement in precision-recall AUC over the baseline and a 2.7% improvement over the production ranker for a large search engine.

3.6   Acknowledgements

Chapter 4

Evaluating Advertiser Strategies

In the previous chapter, we discussed ways in which a search engine can better interpret the advertisers' targeting criteria by using organic search results for the ad keywords on which the advertisers place their bid. In this chapter, we evaluate the effectiveness of different targeting strategies used by advertisers using a month of user-click data from a large search engine. We find that the advertising landscape is quite nuanced: a given targeting strategy may be effective for some advertisers but not others. For instance, advertisers with well-established brands may find little benefit in targeting search queries mentioning their brand, while less well-known brands may need to do so to protect themselves from competitors looking to poach their customers. Similarly, targeting mobile devices may be more a effective strategy for businesses with brick-and-mortar storefronts than for businesses with purely online presence.

Measuring the effectiveness of advertising is, in general, a hard problem. As John Wanamaker famously quipped, "half the money I spend on advertising is wasted; the trouble is I don't know which half." Sponsored search holds out the promise of addressing this longstanding challenge by focusing spend on the right consumers at the right time. The sheer scale of fine-grained, user-activity data that can be brought to bear (e.g., tracking every ad click and every user action on advertisers' sites) allows advertisers to reach populations of particular interest. Moreover, search queries capture the intent

of the user allowing more direct connections between the ad and the user action [63]. This tight relationship contrasts with traditional brand advertising where connections between ads and purchases are more nebulous. Hence, it is no surprise that sponsored search accounts for over 50% of the $49.5 billion spent in online ads and growing at 20% annually [58].

Despite its ever-increasing prevalence, very little is publicly known about the effectiveness of online search advertising. Indeed, large advertisers have presented conflicting anecdotal evidence. As recently as April 2013, a study claimed that the estimated $51 million eBay spends on search ads is ineffective since they essentially cannibalize clicks from organic search results: in the absence of eBay's ad the user would have clicked the eBay page in the organic results [34]. Similarly, three Indian online apparel retailers found that poaching each other's users by advertising on their competitors' brand names was counterproductive; while they might succeed in getting their competitors' customers to click their ad, the customers would typically not convert. For the few customers who do end up making a purchase, the cost paid by the advertiser is often too high [26]. Despite these reports, ad networks and ad agencies maintain that poaching and cannibalizing organic clicks both have a net positive return on investment [26]. Opinions are similarly mixed in the mobile space, where many advertisers complain about accidental clicks while ad networks defend the effectiveness of mobile search ads [19, 23].

One factor contributing to these seemingly contradictory reports is the current inability of anyone other than the advertisers themselves to determine if a particular advertising campaign is effective. Publishers and marketing firms deal in terms of metrics like click-through ratio (CTR), which reports the number of clicks on an ad as a fraction of the number of times the ad is shown, and cost per acquisition (CPA), which reports an advertiser's total spend divided by the number of resulting purchases. Unfortunately,

neither of these metrics lends insight to the key question—namely whether a particular advertising campaign is profitable for the advertiser or not, which is captured by the metric profit per impression (PPI). However, determining an ad campaign's PPI fundamentally requires knowledge of an advertiser's cost and revenue structure—information they are likely loathe to share, even with their marketing agencies.

Instead, we present a simple metric—net acquisition benefit (NAB)—that leverages information about how much an advertiser bids for ad placement (which, for rational advertisers, is a lower bound on the profit they expect) to estimate profit per impression (PPI). Using data about billions of clicks from a large search engine we conduct a month-long measurement study of the effectiveness of three search ad campaign strategies—cannibalization, poaching, and ad extensions—that are widely employed by advertisers today.

4.1   Dataset

The dataset we use for analyzing the effectiveness of different targeting strategies is several terabytes in size, consisting of billions of search and ad clicks on a large search engine in the United States English market. We report upon clicks capturing all actions taken by hundreds of millions of users who issue hundreds of millions of unique English-language queries over a period of four contiguous weeks starting in early April 2015 (Analysis of a different four-week time period from August 2014 obtains qualitatively similar results). Desktops and laptops account for 81% of clicks while phones and tablets account for around 12% and 7% of clicks, respectively. Our dataset covers many millions of dollars in advertising spend[1] by hundreds of thousands of advertisers. Our dataset does not cover specialized search verticals like image, video and map, or product listings.

---

[1]We are obliged to report only the magnitude or normalized values for some sensitive quantities when doing so does not compromise the scientific value of our results.

For each click our dataset includes the following anonymized information: *i*) the normalized search query issued by the user and the search ad network's internal query classification; *ii*) information about the browser including version and operating system, device form factor *iii*) the list of organic search results and paid search ads presented to the user; *iv*) the details of the associated ad campaigns including bid amounts, keywords targeted, and ad extensions; *v*) the organic search results or paid search ads on which the user actually clicked (if any) including clicks on multiple results and ads; and, lastly, for ads clicked, *vi*) the second-price bid charged to the advertiser along with any advertiser-reported conversion event(s) for that click along with the URL of the pages for which the user conversion was reported and the (opaque) conversion types.

Along with click data, the analytics system collects user activity data on advertiser websites to track the performance of the ad campaigns on behalf of the advertisers. Whenever a user performs an action that the advertiser wants to track, JavaScript embedded in the page sends information about the action—along with a user cookie allowing the data to be connected to the user's search behavior—to the search engine. This system allows advertisers to declare which user actions constitute conversions.

4.2   Metric

In this section we present NAB, a simple metric we use for measuring the effectiveness of online search ads. Our primary goal is to design a metric that can be used by ad networks, ad agencies or any entity that manages large search advertising campaigns. We discuss the challenges in choosing the right metric to measure the performance of ad campaigns. In particular, we address the fact that details of customer transactions (e.g., profit margins) may not be available to the entities running the search engine marketing campaigns for the advertisers. We also discuss how our metric compares with other metrics commonly used by advertisers. We hope that NAB will allow not only advertisers
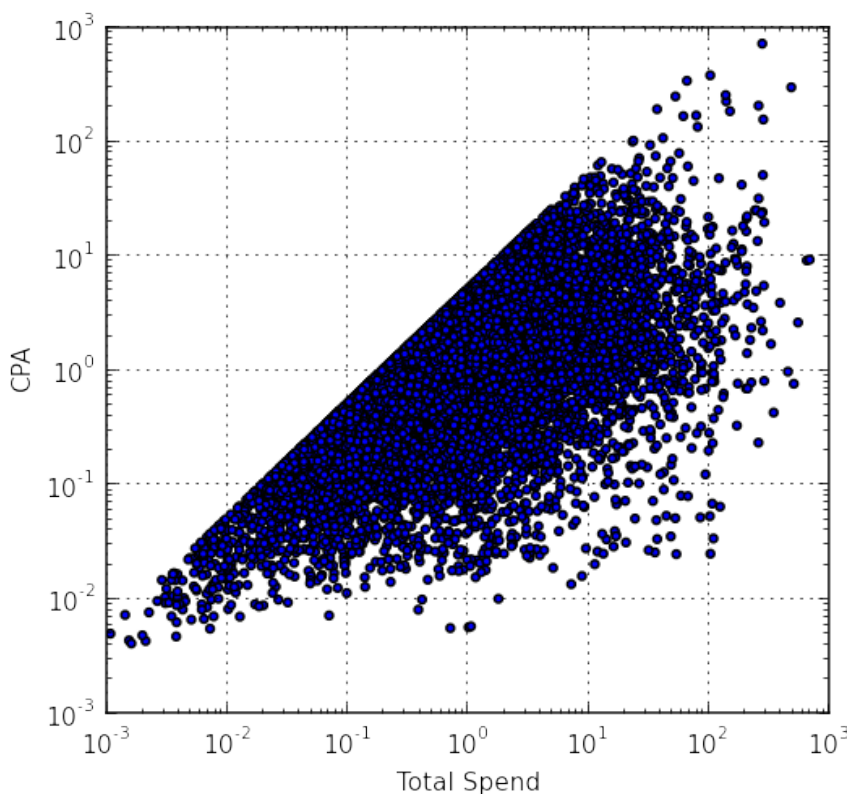
Figure 4.1. Large spread in the advertiser ad spend and the price they pay per conversion.

but ad agencies and other third parties to replicate our methodology to measure and compare other advertising strategies.

We would like to design a metric that is general enough to apply to the diverse ad campaigns seen at a large search engine. We face three key challenges that we discuss below: first, scaling to the large diversity of advertisers; second, balancing granularity of analysis with statistical significance; and finally third, being robust to large amounts of noise that are inherent in data at scale. We take a quantitative approach to better understand these challenges and to build intuition for our resulting metric.

4.2.1 Challenges

Our choice of metric for measuring the performance of different advertising strategies is driven by the following challenges we faced.

Advertiser Diversity. Advertisers span orders of magnitude in terms of their scale. The effective cost per acquisition (CPA) for an advertiser, computed as the total amount charged divided by the total number of conversions, is a rough proxy for the monetary utility the advertiser expects to derive from advertising [54]. Figure 4.1 plots the CPA for advertisers reporting at least 10 conversion events in our dataset vs. the total money spent by the advertiser; the values are normalized to the median value along each axis. (Because each advertiser has at least 10 conversions, there can be no points above the CPA = TotalSpend/10 diagonal.) As one might expect, different advertisers have vastly different budgets. Both the total money spent by advertisers (x-axis domain) and effective cost-per-acquisition (y-axis range) span almost six orders of magnitude. Thus, our metric must take into account an advertiser's target cost per acquisition and allow comparison between different advertising strategies of an advertiser.

Statistical significance. As mentioned, ad targeting can be extremely fine-grained focusing on specific keywords, device types, geographic regions, etc. Figure 4.2 plots the probability density function of the share of total advertising money spent on the most popular user queries (ordered by the amount of money spent on them). To illustrate, note that the most frequent user query attracts less than 0.2% of the total spend. There is a heavy tail of queries with the top 200 most-popular together accounting for less than 7% of the total spend. Performing analysis at query granularity results in poor statistical significance. The statistical significance is lower still if the data is further sliced by user, device type, advertiser and other targeting parameters.
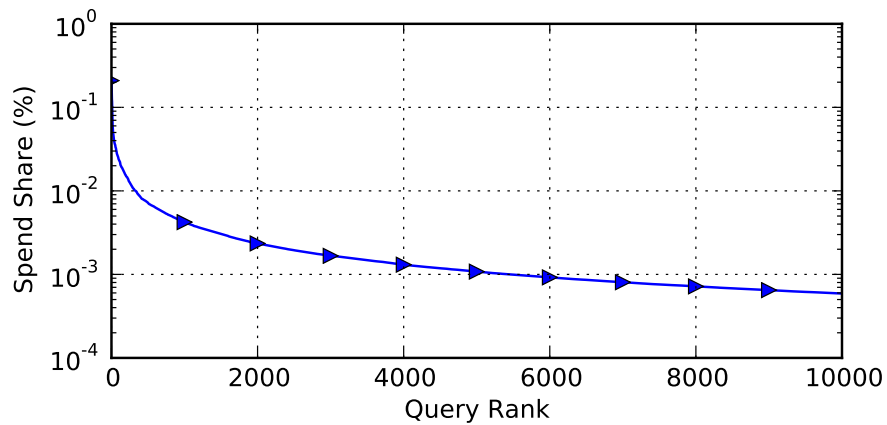
Figure 4.2. Top queries account for a small share of overall spend, illustrating the query diversity.

Hence, any useful metric must be defined over traffic aggregates, which we refer to as slices of search requests. The dimensions along which the data should be aggregated depends upon the strategy being evaluated. In general, we consider slices that capture a particular advertising strategy. For instance, aggregating data by device type may help evaluate the effectiveness of advertising on mobile devices, while aggregating data by query classification to, e.g., consider only ads placed on competitors' brand names, may help evaluate the impact of poaching.

Noise and uncertainty.   Consider briefly a preliminary design that we reject because it is extremely ill-conditioned to noise: incremental cost per incremental conversion, defined as:

$$\frac{\Delta \text{cost}}{\Delta \text{conversions}}.$$

To illustrate, we compare the effectiveness of Contoso Inc. advertising on the keyword "contoso" even though they are prominently ranked in the organic search results vs. not advertising. Δcost is the total money spent on advertising, and Δconversions is the number of conversions when the ad campaign is running vs. when the ad campaign is not

running over a comparable length of time. Say Δcost $= \$100$, and 100 people converted while the ad was running and 95 when paused, i.e., Δconversions $= 5$. The incremental cost per incremental conversion metric is $\$20$/conversion. Note this already is a more accurate reflection of effectiveness over the traditional cost per acquisition (CPA) metric, which would report $\$1$/conversion ignoring the organic conversions Contoso would anyway have received.

The ill-conditioned nature of the above design becomes apparent in the presence of noisy data. By noise we mean both systemic fluctuations (e.g., changes in the search ranking algorithms) as well as noise from sources external to the system (e.g., sudden user interest). Assume that for some unknown reason Contoso measured 101 conversions while the ad campaign was paused in the above example. Δconversion would now be -1 and the ill-conditioned output of the metric is ($\$100$), a nonsense result that computes the incremental cost per incremental conversion at negative $\$100$ (a -600% change in the output for a 6% change in the input). Inverting the metric makes it ill-conditioned with respect to slight variations in ad costs.

4.2.2   Net acquisition benefit (NAB)

Intuitively, the net acquisition benefit (NAB) is the conversion probability of a traffic slice adjusted by its cost. We define NAB for a traffic slice x as follows:

$$\mathrm{NAB}(\mathrm{x}) = \pi_x - \frac{v_x}{\lambda},$$

where:

x, n : Traffic slice x consisting of n impressions

π_x = Conversion probability, i.e., $\frac{\#conversions}{n}$

v_x = Average cost, i.e., $\frac{cost}{n}$

$$\lambda \quad = \quad \text{Advertiser's target cost-per-acquisition.}$$

λ is the maximum amount an advertiser would be willing to pay for a conversion, which is well captured, for example, by the bid they place in the ad auction or equivalently, the minimum amount advertiser would be willing to save for forgoing a conversion. Obviously, a rational advertiser would not want to pay more for an ad than they stand to make in profit on the conversion, so we argue that λ serves as a lower bound for the profit an advertiser expects to capture from a conversion. Note that if λ precisely equals the advertiser's profit margin on the product without accounting for the search advertising cost, then NAB is proportional to net profit per impression (PPI).

NAB $= 1$ for an optimally beneficial traffic slice — where every case results in a conversion ($\pi = 1$) and there is no cost ($v = 0$). NAB $= 0$ for traffic slices that have no net benefit, e.g., where the traffic slice is so expensive that the advertiser is willing to forgo every conversion and save the entire cost (i.e., effective cost-per-acquisition is λ and $v = \lambda \pi$). For detrimental traffic slices, e.g., there are no conversions ($\pi = 0$) and the advertiser is losing money ($v > 0$), NAB is negative. In practice, NAB is on the order of 10^{-2} in our real-world dataset. (Intuitively, this makes sense, as CTRs are typically of the same order.)

4.2.3 Incremental NAB

The incremental net acquisition benefit (INAB) measures the relative improvement in NAB of one traffic slice over another, i.e., the effectiveness of one ad campaign vs. another. Intuitively, it is the change in conversion probability ($\Delta \pi$) adjusted by the change in cost (Δv). We define INAB for traffic slice x over slice y as follows:

$$\text{INAB}(x \mid y) \quad = \quad \frac{\text{NAB}(x) - \text{NAB}(y)}{|\text{NAB}(y)|}$$

(a) Click-through rate

(b) Cost per acquisition
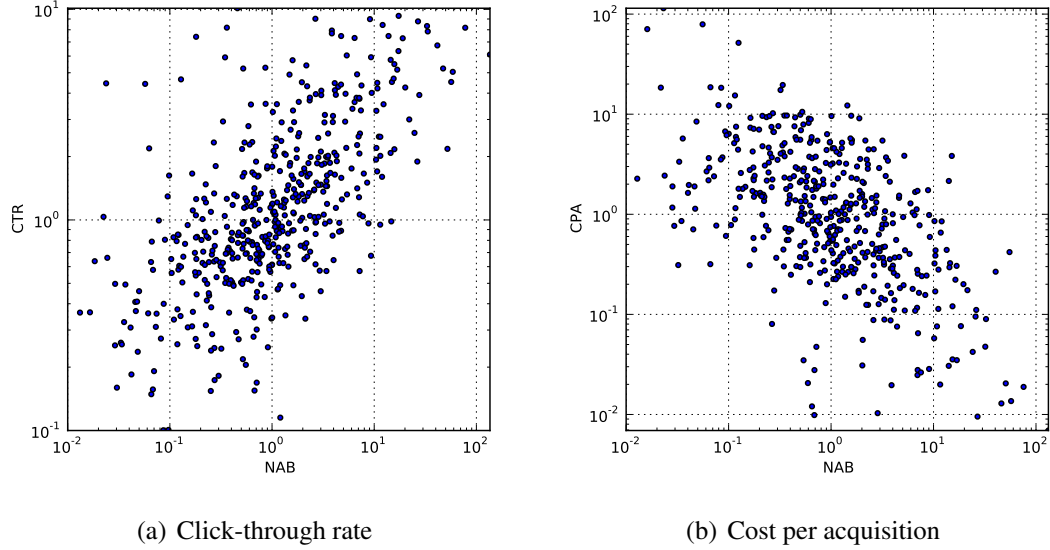
Figure 4.3. Comparing traditional metrics to NAB.

where $|\text{NAB}(y)|$ represents the absolute value of NAB(y).

INAB is defined only for two comparable traffic slices. That is, INAB can be computed for two traffic slices x and y belonging to the same advertiser but not across advertisers. Traffic slice x is more beneficial than y if and only if x has more net acquisition benefit than y, i.e., INAB($x \mid y$) is positive. Slice x is less beneficial than y if INAB($x \mid y$) is negative (or equivalently, INAB($y \mid x$) is positive). Both are equivalent if INAB($x \mid y$) is zero.

4.2.4 Validation

We validate the NAB metric by comparing it to click-through ratio (CTR) and cost per acquisition (CPA), the two most commonly used metrics to evaluate the effectiveness of ad campaigns. We compute NAB, CTR and CPA for each advertiser over the entire traffic that they get through sponsored search during the month for which we have the data. We describe our methodology for computing NAB in greater detail in the next section. Figures 4.3(a) and 4.3(b) are scatter plots of CTR and CPA respectively vs. NAB;

all values are normalized to the median for easier comparison. We note first that NAB is (weakly) positively correlated ($r = 0.43$) with CTR, and (weakly) negatively correlated ($r = -0.10$) with CPA. This correlation is consistent with expectations that advertisers have today: ad campaigns with high CTR or low CPA are, in some sense, beneficial to the advertiser as reflected by their higher NAB values. We manually investigate outliers to build confidence that NAB better captures the true effectiveness of advertising for the outliers in the figures above.

Misleading CTRs. For a sampling of advertisers with a high CTR we find that those with a low NAB have anomalously high cost per conversion. For a well-known cellular provider, for instance, their CPA is over an order of magnitude higher than their competitors' in our dataset. NAB for these advertisers is, thus, justifiably low and the simplistic CTR metric that fails to take their (lack of) conversions into account is misleading. For a sampling of advertisers with low CTR we find those with a high NAB have a high conversion rate and low CPA. The high NAB is typically because conversions for these advertisers refer to email sign-ups and file downloads, and while few click their ad (low CTR), a majority of those who do end up converting, resulting in an effective ad campaign consistent with a high NAB.

Misleading CPA. For a sampling of advertisers with high CPA we find that those with high NAB are in high-margin industry segments, e.g., insurance, and there is little disparity in the CPAs of competing advertisers in the same industry. Contrast the insurance company with the cellular provider mentioned earlier that is the only one in its industry with an anomalously high CPA. NAB correctly assigns a high effectiveness score to the former, and a low score to the latter. For a sampling of advertisers with low CPA we find that those with low NAB are ones that have an anomalously low CTR, i.e.,

their ad campaigns are poorly targeted to attract users.

4.2.5 Summary

NAB approximates profit per impression (PPI) when λ is equal to profit margin on conversion. But, unlike profit per impression, NAB does not require information about revenue derived and cost of the products. This allows ad networks and ad agencies to use the NAB metric to compare effectiveness of advertising campaigns. Note that both NAB and PPI, being impression based, can be sensitive to impression counts (impressions are cheap, one may argue). However, an impression represents the most basic intervention to user experience on behalf of advertisers, and in absence of a better denominator, impression-based metrics are still considered industry standard [82].

Target cost per acquisition. One subtlety with NAB is that advertisers could have different target costs per acquisition for different ad campaigns. We discuss our methodology for inferring target cost per acquisition in next section.

4.3 Methodology

This section describes the methodology we follow for measuring effectiveness of various ad targeting strategies in the subsequent sections. In order to compute NAB we infer target cost per conversion (λ) from the data. We aggregate data over queries that identify traffic representing different advertising strategies.

4.3.1 Conversions

In order to identify the conversions that an advertiser obtains from a slice of traffic we have to attribute the conversion to a specific prior search. To attribute a conversion to a prior user search, we identify the user actions on search engine prior to the conversion

event on the advertiser website. We then attribute the conversion to the latest user click (regardless of whether the click was on an organic result or a search ad) that led the user to the advertiser's website—as long as the click happened in the 24 hours prior to the conversion event.

Not all advertisers report conversion signals, or not in significant numbers. Unless otherwise mentioned, we omit advertisers for whom we have less than 30 conversion reports in our dataset.

4.3.2   Inferring target cost per acquisition

Recall from Chapter 2 that advertisers bid the maximum amount they are willing to pay for a click. We infer the maximum amount the advertiser is willing to pay for a conversion (λ) by dividing their total bid amount for the ads clicked by the number of conversions they received. Since the bid values are always more than the actual cost of advertising, overall NAB for any advertiser is always positive. Note that by making this choice we consider all conversions that the advertiser receives in US English market—irrespective of campaign—equivalent.

4.3.3   Aggregating queries

NAB must be computed over a significant aggregation of traffic. As mentioned earlier, individual search queries are too granular. We follow the search-ad network's internal classification scheme [83] to aggregate queries into the following four classes of particular interest: navigational (24% of all queries), local (9%), commercial (9%), and other, which includes informational queries.

Since the internal query classifier relies on heuristics, we verify the correctness of classification by manually investigating a representative sample. In all we manually verify 200 queries and find that in the large majority of cases (94%) our manual label matches

Table 4.1. Traffic features used to define traffic slices.

Query
nav	Navigational query; user seeks specific site
com	Commercial query; user has purchase intent
all	All queries

Device
phone	Mobile smartphones
pc	Desktops and laptops
all	All devices

Organic
yes	Present in first page of results
no	Not in first page of results
top	Top-most organic search result
poor	Ranked 2 or worse, or not on first page
n	Ranked n
$n+$	Ranked n or worse
all	All cases whether present or not

Ad
yes	Ad present
no	Ad not present

Ext. (set)
ad:call	Ad has call button
org:call	Organic result had call button
ad:comp	Competitor has an ad

the classifier's; in the remaining 6% of the cases we believe the classifier misclassified the query. We compute the sensitivity and specificity measures for classification of navigational queries and find that that 77% of the time, a navigational query is classified as navigational, whereas 4% of the time, a non-navigational query is classified as navigational. Overall, the query classification, while not perfect, seems sufficiently accurate for the purposes of our study.

Table 4.2. Traffic slices used in this chapter.

Slice	Query	Device	Org.	Ad	Ext.
Section 4.4: Cannibalizing Organic					
org-n-noad	all	pc	n	no	
org-n-ad	all	pc	n	yes	
nav-noad	nav	pc	top	no	
nav-ad	nav	pc	top	yes	
nav-comp-noad	nav	pc	top	no	ad:comp
nav-comp-ad	nav	pc	top	yes	ad:comp
noorg-ad	all	pc	no	yes	
Section 4.5: Poaching					
poach-ad	nav	pc	poor	yes	
poach-noad	nav	pc	poor	no	
commerce-ad	com	pc	all	yes	
Section 4.6: Ad Extensions					
phone-orgcall	all	phone	yes	no	org:call
phone-org	all	phone	yes	no	
phone-ad	all	phone	all	yes	
phone-adcall	all	phone	all	yes	ad:call
phone-noorg-ad	all	phone	no	yes	
phone-noorg-adcall	all	phone	no	yes	ad:call
phone-orgcall-adcall	all	phone	all	yes	org:call, ad:call

4.3.4 Traffic slices

In the subsequent sections we compare the effectiveness of various campaign strategies by comparing the profitability observed by an advertiser over different traffic slices. Each slice of traffic is defined by the query classification, device type, the position of the advertiser in the organic results (if at all), whether the advertiser's ad is shown or not, and whether the call button was present for the advertiser's ad or organic result. Table 4.1 lists these features and describes the values they take. Table 4.2 labels the various combinations of these features with a name that we use to refer to that traffic slice in subsequent sections. We discuss our choice of traffic slices further in Section 4.7.

4.4 Cannibalizing Organic Traffic

We refer to the scenario where advertisers show ads for queries where they have an organic presence as cannibalization: in the absence of ads, users could have navigated to the advertiser by clicking on the organic result of the advertiser. We consider three particular scenarios of cannibalization and use NAB to measure the benefits of each. We find that, *i*) for queries where the advertiser is the top result 56% of the advertisers derive no benefit from advertising, *ii*) the incremental benefit of advertising increases as the organic rank of advertiser decreases and *iii*) over 61% of the advertisers achieve limited benefit by advertising on navigational queries.

4.4.1 Improving visibility

In this section we consider the general case of searches where the user does not already have a particular destination website (advertiser) in mind. In such an instance, even though an advertiser's website appears in the organic search results, the advertiser may wish to increase its visibility to induce the user to visit its website as opposed to a competitor's. We consider an advertiser to be attempting to improve their visibility if the advertiser chooses to display an ad despite already being included in the organic results likely to be considered by the user (i.e., ranked in the top ten results based on relevance).

Same-query visibility.   The main challenge in measuring the impact of advertising is obtaining comparable search impressions. In the ideal scenario we would compare user actions in the presence and absence of an ad while everything else remains the same. Since we do not have such data, we compare the performance of ads by comparing benefits of advertising over two sets of impressions for the same query, one with ads and one without ads. Comparing user actions over impressions for the same query ensures that the most significant variable—the user query—remains the same in both sample sets.
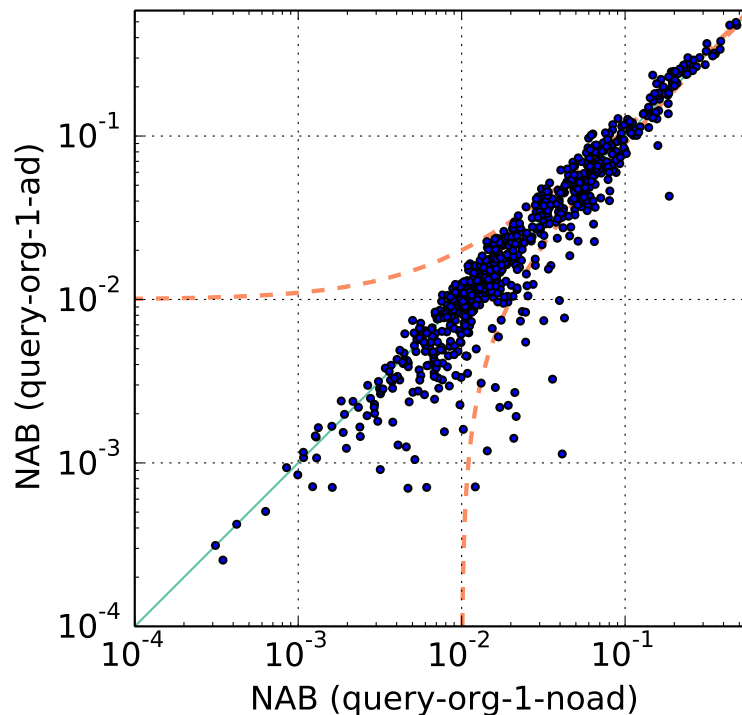
Figure 4.4. Most advertisers see little to no benefit in advertising for queries where they are the top result.

We begin by considering the effectiveness of advertising when an advertiser is already the top organic search result. For each (advertiser, query) pair, where advertiser is the top organic result for that query, we identify impressions with and without ads by the advertiser. I.e., a single advertiser may be considered multiple times if they are the top search result for more than one query. Figure 4.4 plots the NAB of advertising (query-org-1-ad) vs. not advertising (query-org-1-noad) for all (advertiser, query) pairs where we have more than 30 conversions (as discussed in Section 4.3). In our dataset, we have 824 (advertiser, query) pairs covering 345 distinct advertisers. The diagonal line represents an INAB of zero, i.e., the effectiveness of advertising is equivalent to the effectiveness of not advertising for that advertiser. Advertising is more effective than not for advertisers above the diagonal. A slight majority are below the line where the inverse
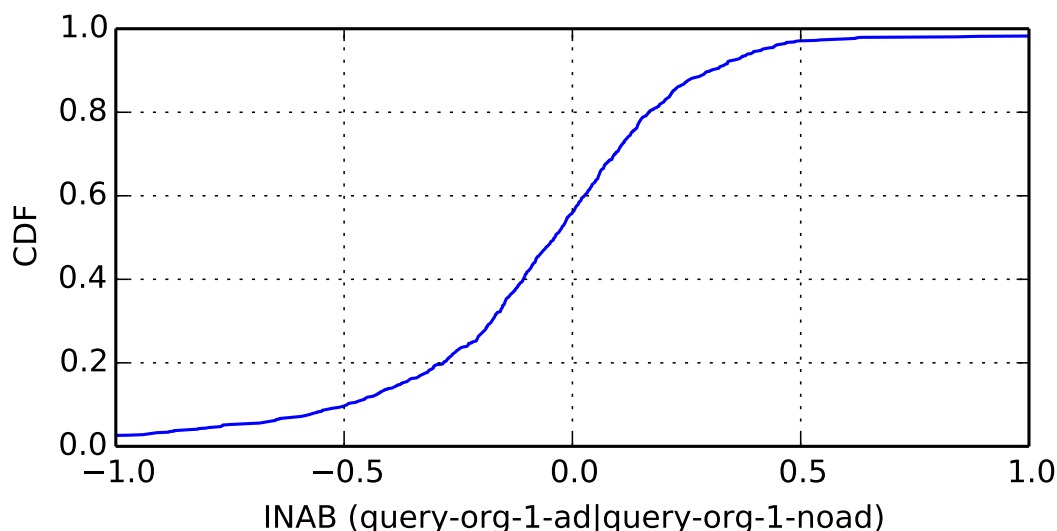
Figure 4.5. 56% advertisers achieve zero or negative INAB by advertising on queries where advertiser is the top result.

is true—i.e., it is not worth advertising for that query. The dashed red curves above and below the diagonal are intended to aide comparison as the graph spans four orders of magnitude on each axis: they represent a ΔNAB (NAB(ad) - NAB(noad)) of $+0.01$ and -0.01 respectively.

As an alternative representation of the same data, Figure 4.5 plots the CDF of the corresponding INAB metric. Recall that INAB is normalized to the NAB of the base strategy, so a value of 1.0 represents a $2\times$ improvement. For 56% of the (advertiser, query) pairs the corresponding advertiser gets zero or negative incremental benefit by advertising on the query (i.e., fall below the diagonal in Figure 4.4). The results mean that these advertisers should reconsider advertising on those queries.

Over all queries. The constraint of comparing the benefits of advertising by holding the user query a constant limits our analysis to a few advertisers who have significant number of conversions for the same query. Even for those advertisers, these few queries contribute to a small proportion of their entire ad spend. Hence, there is reason to believe

(a) Rank 1

(b) Rank 3

(c) Rank 5
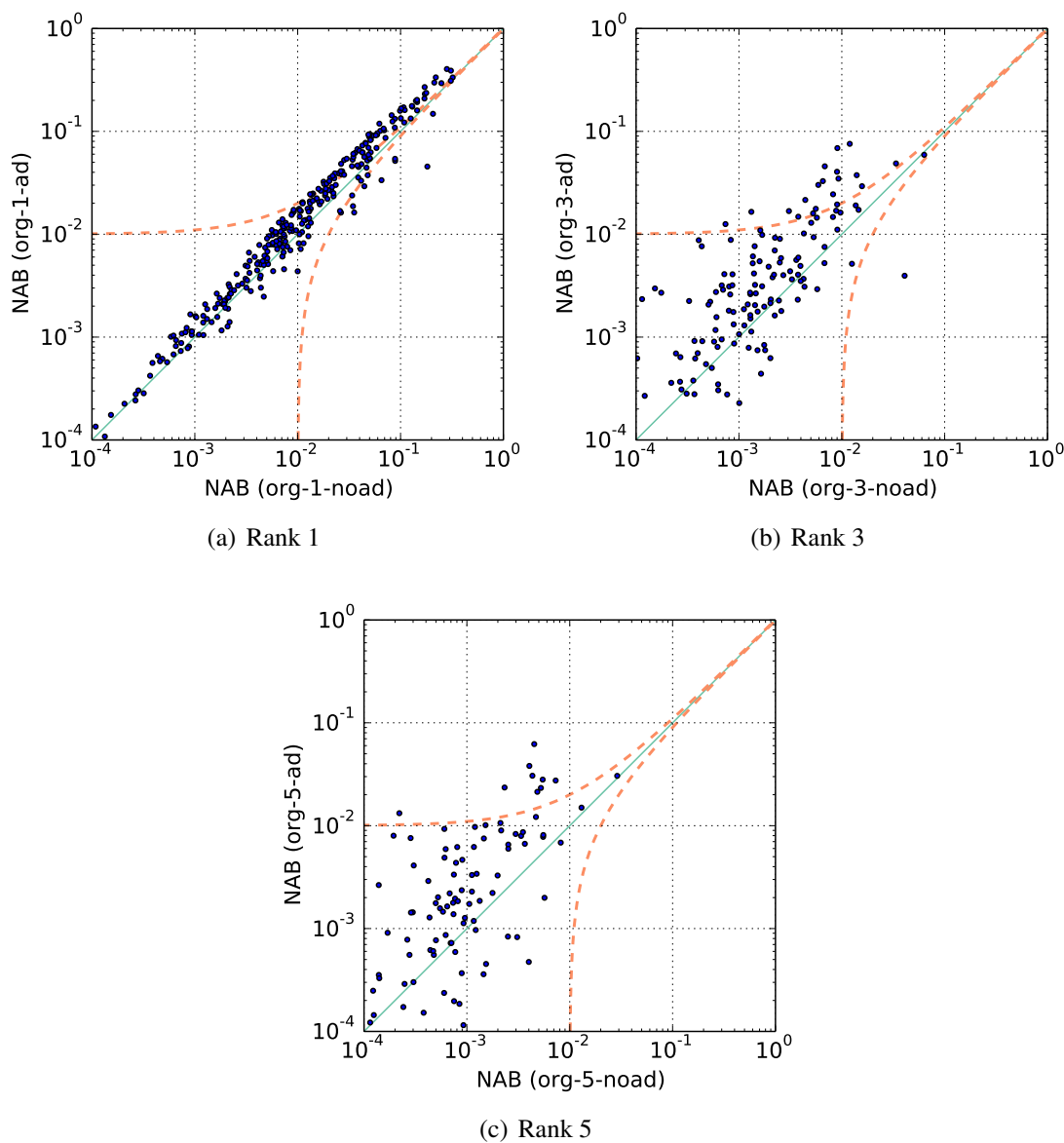
Figure 4.6. The benefits of cannibalizing organic clicks increase with decreasing organic search ranking.

the same-query results may not be representative. By design, we can apply our metric to measure effectiveness of ads on an arbitrary slice of traffic. Here we expand our analysis to compare traffic slices where the advertiser has a particular rank in organic results, irrespective of user query.

Figure 4.6 considers only non-navigational queries (i.e., those where the user
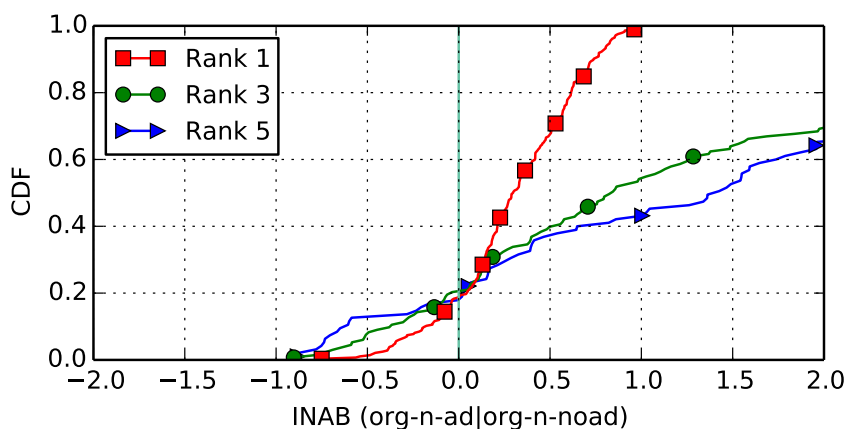
Figure 4.7. As rank increases in organic results benefits of cannibalistic advertising turn positive, albeit only slightly.

likely does not have a particular destination website in mind) and plots NAB of advertising (org-n-ad) vs. not advertising (org-n-noad) aggregated across all queries for each advertiser (i.e., each advertiser is a single data point) where we have sufficient data as described in Section 4.3. We separate campaigns based upon their rank in organic results. Figure 4.6(a) considers the aggregate performance of all non-navigational queries for which an advertiser anyway obtains the top organic result position (cf. Figure 4.4 which shows a distinct point for each unique top-result-producing query an advertiser may obtain). Figures 4.6(b) and 4.6(c) show the same comparison for campaigns where the advertiser appears third or fifth in organic search results, respectively.

Figure 4.7 plots the CDFs of the corresponding INAB metrics for each of the three classes of campaign considered in Figure 4.6. In contrast to the same-query results above, most of the advertisers gain by advertising for non-navigational queries despite being present in the organic results. The gains from advertising are likely because non-navigational queries tend to be competed for more aggressively and users are flexible with choosing any business that meets their needs. Also, while the benefits of advertising increase as the organic rank drops, the absolute benefits of advertising are very low—likely because of the lower relevance of the advertiser to user query.

Interestingly, for 32% of the top-ranked websites, there appear to be significant benefits to advertising (INAB $> 50\%$). Manual investigations find that these advertisers belong to two categories. One set of advertisers have less recognized brands. If a better recognized brand advertises for the query, and is thus placed above the top-ranked organic result, the latter lose out. The other category of advertisers who gain are those who share their brand with other advertisers. For example, car manufacturers lose potential converting users to competing ads placed by car dealers trying to attract the same users. For 19% of the top-ranked websites, advertising is a drain on their ad spend (INAB < 0) since they end up competing on hotly contested queries resulting in higher cost per acquisition than their overall average.

4.4.2   Navigational queries

A subset of queries (explicitly excluded above) for which an advertiser is the top organic result is known as navigational queries. We consider a given query navigational for an advertiser if: *i*) the search query is classified as navigational, i.e., the user query includes branded terms or keywords (e.g., Amazon or Facebook) that suggest the user is seeking a specific website, and *ii*) the advertiser is the top-most organic result for that query. Here, we consider whether it is effective for an advertiser to advertise for such queries where the user is explicitly seeking to navigate to the same advertiser's website.

Figure 4.8 plots the NAB of advertising (nav-ad) versus not advertising (nav-noad) for all advertisers where we have at least 30 conversions attributed to clicks following queries matching the above criteria. Figure 4.9 plots the same data as a CDF of the INAB of advertising over not advertising on these queries. As is evident from points clustering along the diagonal in the scatter plot, the majority of advertisers (61%) receive very little incremental benefit (less than 10%) from advertising on navigational queries where they are the top organic result. Moreover, for 32% of advertisers, advertising on
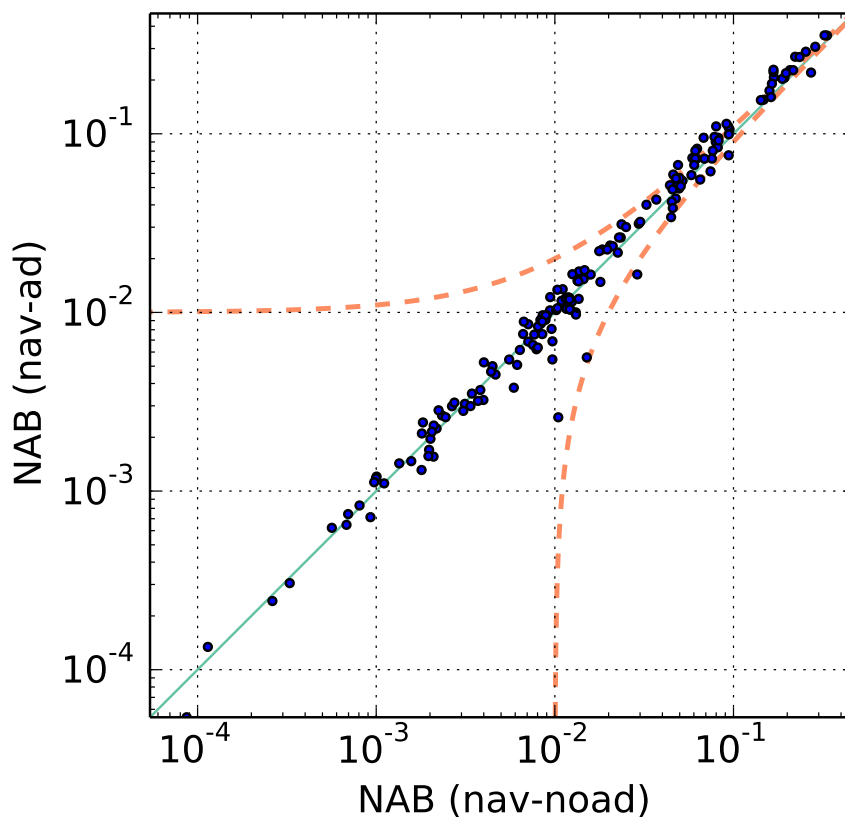
Figure 4.8. There appears to be little difference in returns for the advertiser between the scenarios of advertising vs not advertising for navigational queries.

navigational queries for which they are the top organic result is a net loss when compared to not advertising (i.e., INAB is < 0). In 9% of the cases, however, such advertising bears significant fruit (INAB is $> 25\%$).

Our finding that 61% of advertisers receive limited benefit (INAB approximately zero) from advertising on navigational queries vs. not advertising squarely contradicts reports from other ad networks that suggest 89% of ad clicks are incremental, and would be lost without advertising [44]. We reconcile these results by observing that the previous study does not consider conversions and focuses solely on clicks. Indeed, the study's authors explicitly state that advertisers should consider conversions since relying on clicks alone may be misleading. We show below that advertising on navigational queries
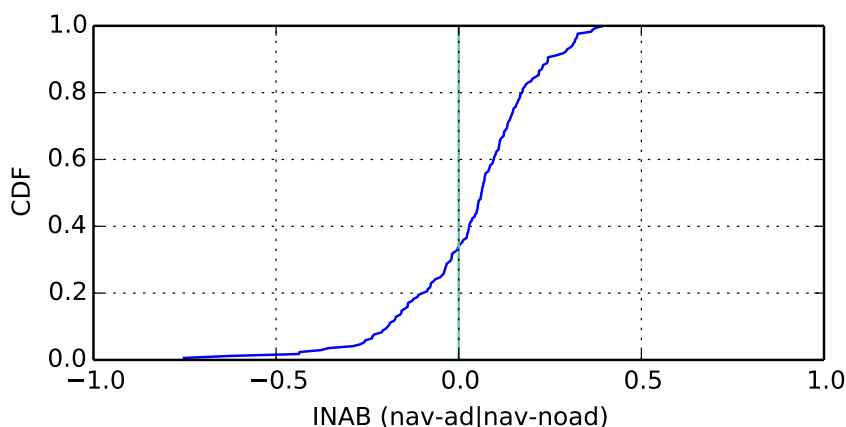
Figure 4.9. The incremental benefit of advertising for navigational queries is very small for most advertisers.

does, in fact, inflate click counts without increasing conversions for the vast majority of advertisers. Another reason for the divergent result is that the previous study considers campaigns that are paused due to budget shortage. Choosing advertisers who face budget shortages would bias the choice of campaigns towards very small advertisers who might not have strong presence for navigational queries.

For advertisers that value conversions, however, our data discredits conventional wisdom that promotes advertising on navigational queries for which the advertiser is the top organic result. Finally, while our general finding is consistent with experimental evidence from large advertisers including eBay [34], we nevertheless encourage advertisers to conduct their own experiments and track conversions to determine if they belong to the small (9%) set of advertisers for whom advertising on navigational queries brings significant benefits. These advertisers, as we discussed in Section 4.4.1, either have a weaker brand or are competing against other advertisers who can legitimately advertise on their brand.
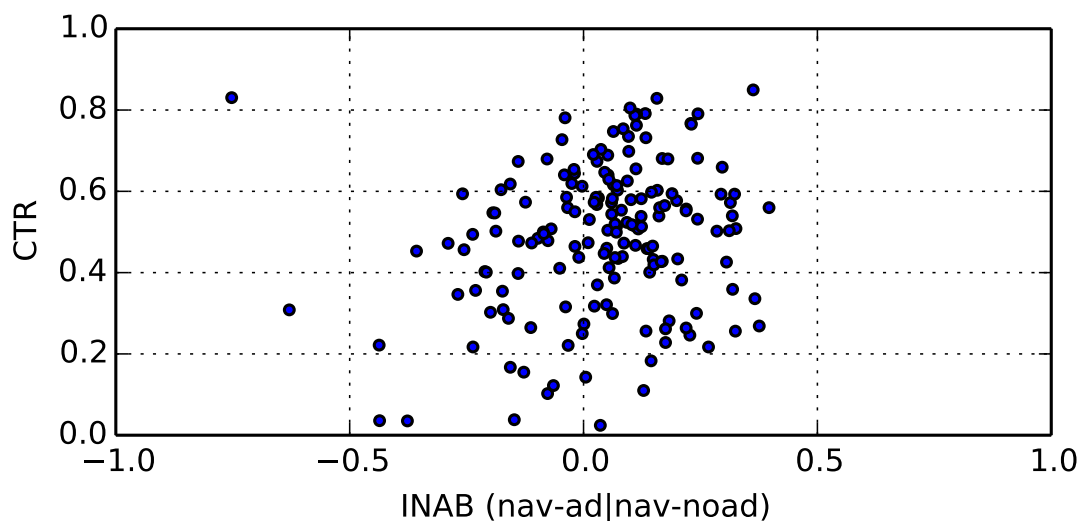
Figure 4.10. INAB has no correlation with CTR and most advertisers receive a high click through rate on navigational queries.

4.4.3 Click count inflation

Many advertisers appear to optimize for clicks rather than conversions. While we can only speculate as to why they choose to do so, anecdotal evidence ranges from naïveté, e.g., unawareness of metrics other than the click-through rate, to financial, e.g., ad agencies that collect commission per-click and advertisers that specify CTRs and minimum click counts (rather than conversions) in contracts with such agencies [17]. One of the easiest ways for advertisers to inflate ad click counts is to cannibalize the advertiser's own navigational queries.

Figure 4.10 plots achieved click-through rate of campaigns that advertise on an advertiser's own navigational queries as a function of their INAB. While the INAB of such a strategy is close to zero for the vast majority of campaigns as we saw earlier, the click-through rate for these campaigns is extremely high as compared to typical search ad CTRs of around 1-2% [21]. Advertisers should run carefully calibrated A/B tests to determine the incremental benefit of advertising for queries where the user is already looking to navigate to the advertiser.
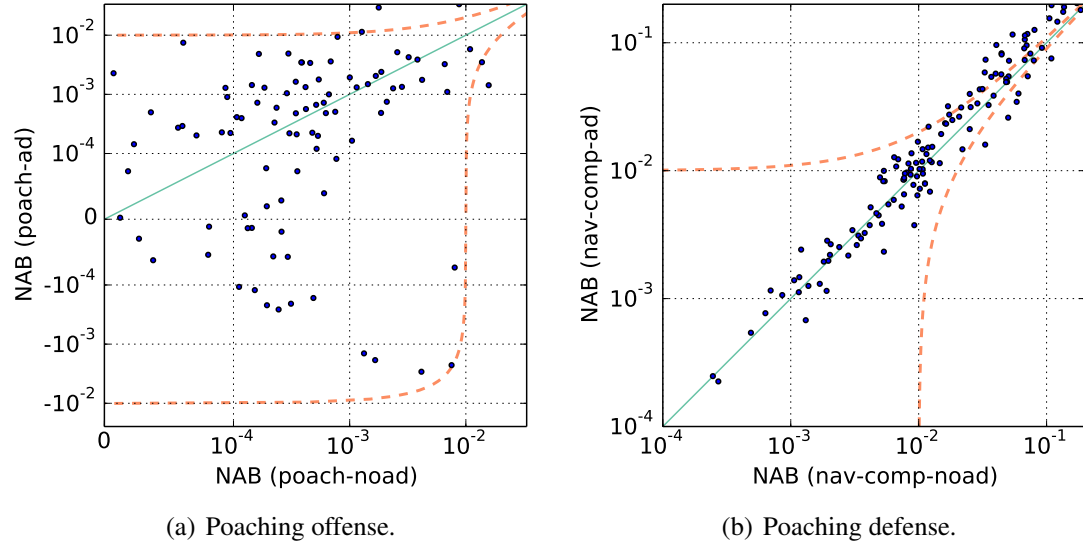
4.5 Poaching Competitor Traffic

Poaching refers to advertising strategies that involve bidding on navigational queries specifically seeking a competitor's website (e.g., queries with competitor's trademarked terms or brand names). Initially, Google's trademark usage policy prevented advertisers from bidding on trademarked keywords they did not own or from using them in their ad content; the restriction on bidding on competitors' trademarked keywords was relaxed in 2004—even if the trademark owner explicitly objected—and restrictions on using them in ad content were relaxed in 2009 [11, 12]. The relaxed policy effectively increased the cost per click of trademarked keywords by allowing competition from third parties, thereby increasing trademark owners' costs by compelling them to bid defensively to protect their trademark.

We find that poaching may help smaller advertisers get users who would not navigate to the advertiser in the absence of ads. But, the costs of attempting to gain such users may be too high for some advertisers when compared to the cost they pay for an average conversion. Moreover, given the varied effectiveness of poaching, defending against poaching has uneven results and can, in fact, occasionally lead to negative results due to the high cost of defense.

4.5.1 Offense

Here, we analyze the benefit of trying to poach users seeking to navigate to a competitor. Figure 4.11(a) compares the NAB for poaching ads (poach-ad, i.e., ads on navigational queries where the advertiser is not the top-most organic search result) against the NAB of not advertising on the same queries (poach-noad). Each data point represents an advertiser where we have sufficient conversion data. About half the advertisers fall below the diagonal showing that poaching is actually detrimental. For the half that do see

(a) Poaching offense.
(b) Poaching defense.

Figure 4.11. Poaching often leads to wasteful spend and defending against it is not worthwhile either.
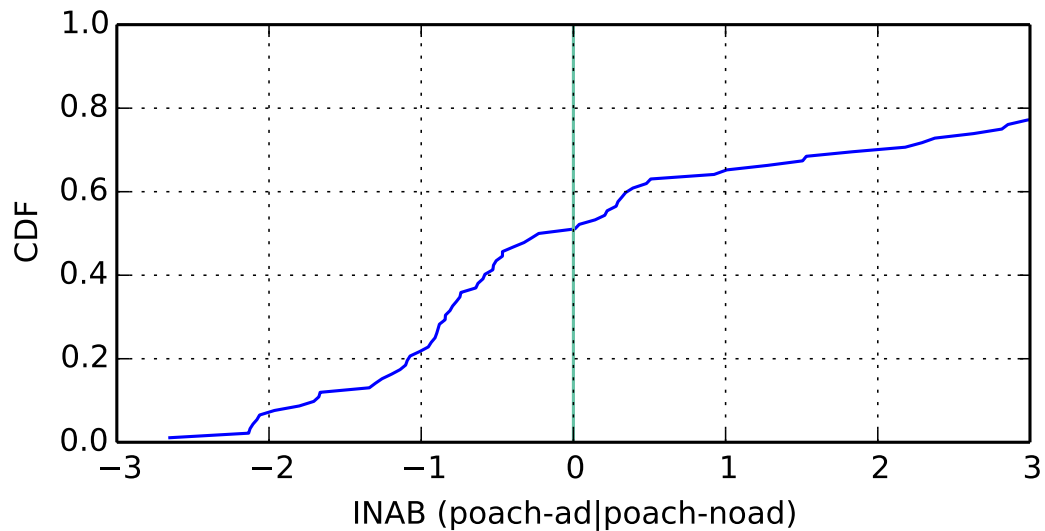
Figure 4.12. For most advertisers, poaching competitor's navigational queries yields little benefit, while some see extreme gains.

relative gains over not advertising, the absolute benefits are negligible (note the absolute value of the NAB is generally well below 0.01).

Figure 4.12 plots the CDF of the INAB for these advertisers. For 50% of ad-
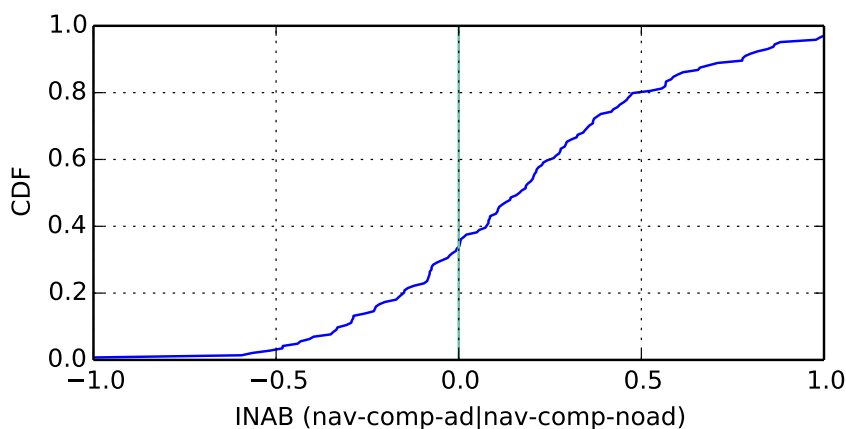
Figure 4.13. Benefits of defending against poaching vary dramatically across advertisers.

vertisers, poaching is of negative value—likely due to few conversions and high costs associated with such ads. There are however, a minority that see significant gains relative to their performance when ads are not shown. The significant gains are due to very poor rate at which they obtain traffic when they do not show ads; said another way, while the relative gains are substantial, in absolute terms they still do not receive many conversions.

4.5.2 Defense

Regardless of how effective poaching is for the advertiser, competitors may still be harmed because they value lost conversions more dearly. Here, we consider whether it is useful for an advertiser to defend against potential poaching by out-bidding competitors for ad space despite being the top organic result for a navigational query. We consider an advertiser a potential victim of poaching for a given query if: *i*) the search query is classified as navigational, i.e., the user query includes branded terms (e.g., Amazon or Macy's) that suggests the user is seeking a specific website, *ii*) the advertiser is the top-most organic result for that query, and *iii*) another advertiser advertises on the query.

Figure 4.11(b) compares the NAB of defensive ads (nav-comp-ad, i.e., ads on navigational queries where the advertiser is the top-most organic search result and a
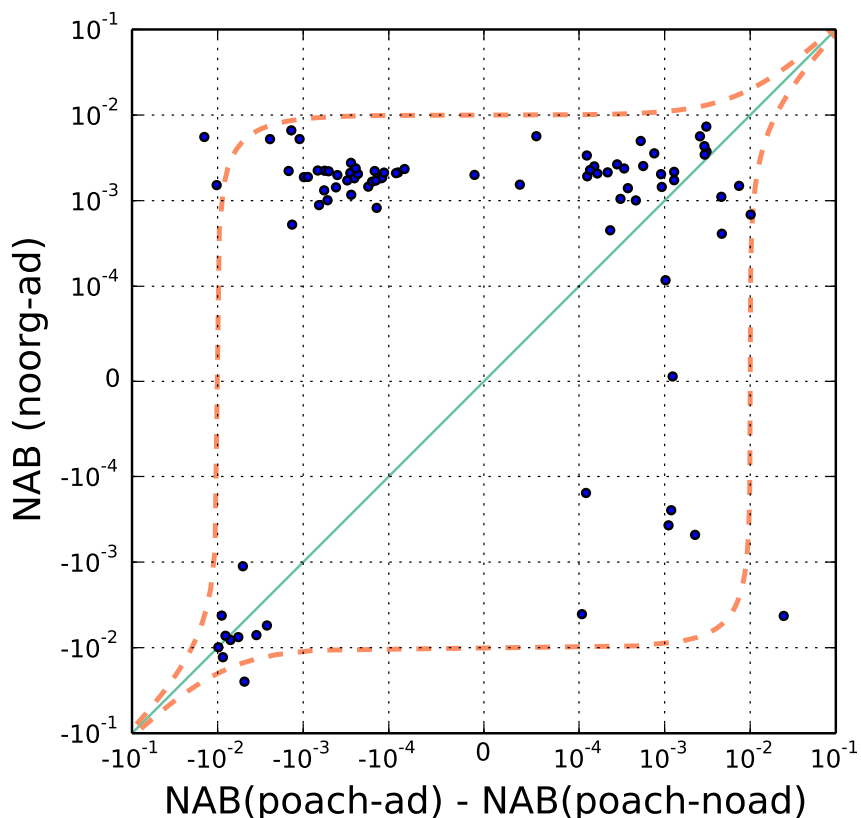
Figure 4.14. A better alternative is to try to gain users on commercial queries where there is no organic presence.

competitor ad is present) against the NAB of not advertising on the same queries (nav-comp-noad). Figure 4.13 plots the CDF of the INAB for these advertisers. The results are mixed: 20% of advertisers realize significant (INAB > 25%) benefits, while a 32% are negatively impacted by attempting to defend these queries.

4.5.3 Spending smarter

As an alternative to poaching a competitor's customers (i.e., users who have issued a navigational search query for a competing website), an advertiser might try instead to recruit customers who are likely to convert somewhere, but have not yet decided on a particular vendor. Here we consider an advertiser deciding between spending money

on poaching ads vs. spending that money to compete on commercial queries (i.e., those that are likely to lead to conversions) where they may not be prominently ranked in the organic results, Figure 4.14 plots the additional benefit of advertising on commercial queries where the advertiser does not have an organic presence i.e. NAB(noorg-ad) vs. the benefit of poaching ads over not poaching (NAB(poach-ad) - NAB(poach-noad)). For 78% of the advertisers, advertising on commercial queries where the advertiser has no organic presence dominates poaching by a wide margin (i.e., the are significantly above the diagonal).

4.6   Displaying Ad Extensions

Call extensions allow advertisers to explicitly add a call button to their ads when rendered on mobile phones. Call buttons are also shown for business listings in the organic search results. These features are provided by the search engine for no additional cost to the advertisers. We have limited data for different advertising strategies on mobile devices. However, our preliminary results show that creating a business listing, which would allow the search engine to show a call button in organic results, is beneficial to the advertisers. The effect of adding a call button to an ad when an organic result with call button is already present is mixed.

4.6.1   Organic business listings

Businesses can create a (free) listing that includes their location, phone number, store hours, parking information, and payment methods accepted through the the search engine itself [3, 10] or third-parties like Yelp [25]. The search engine uses this structured information to enhance the presentation of organic results including showing the call button, map directions, and so on.

We first look at the effectiveness of the call button for organic business listings

(a) Benefit of listing.

(b) Ad call when organic call is present.

(c) Benefit of call ad over regular ad.
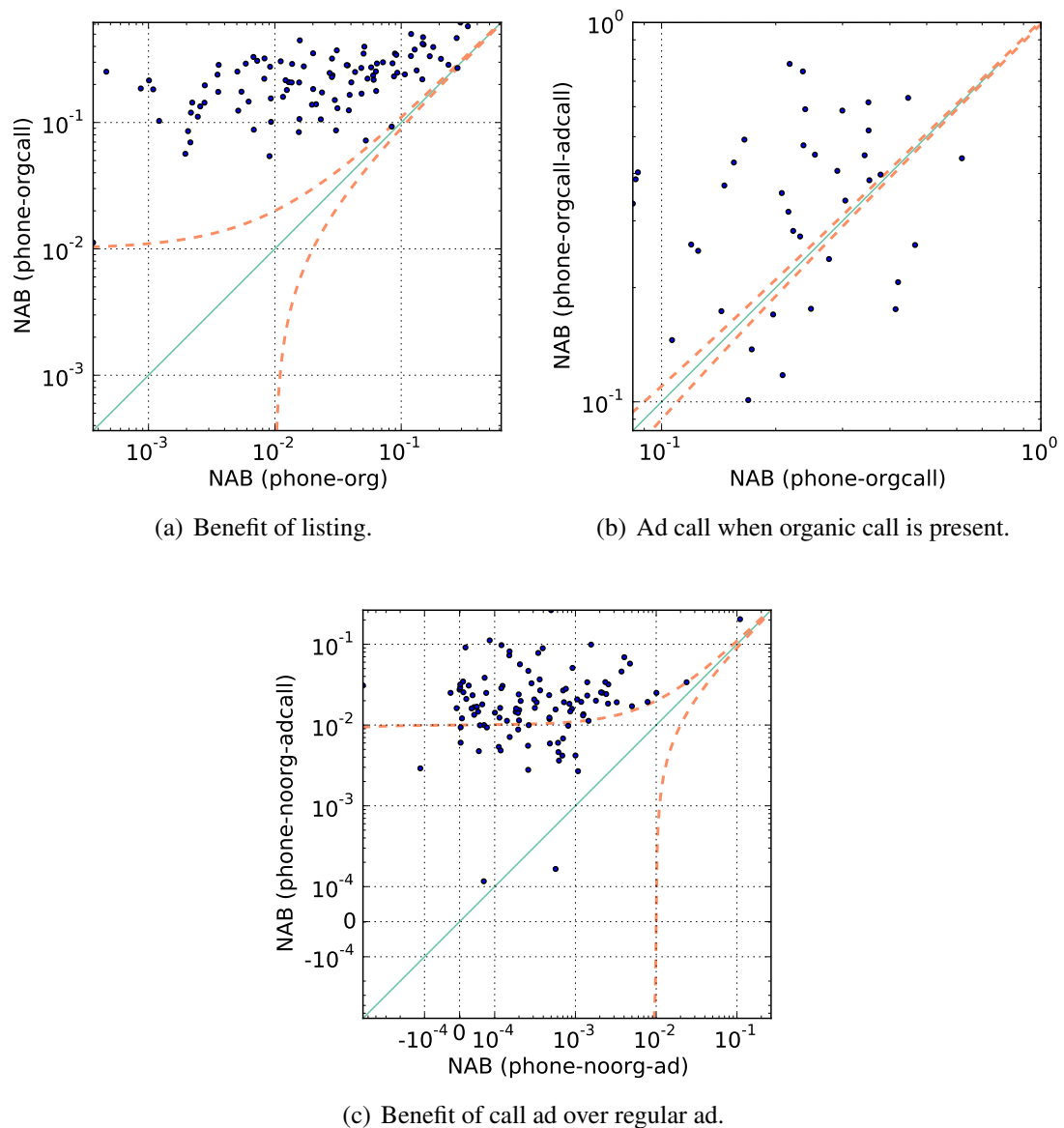
Figure 4.15. Listing business with the search engine benefits advertisers and in the presence of organic call options, ad call extensions yield mixed results. But, showing an ad with extension is more beneficial than a regular ad.

vs. plain search results. It is challenging, however, to define conversion rates in these scenarios. We obviously have no way of knowing from the search logs how many users that call the business end up converting in a way that is equivalent to the conversions in previous sections; here, we consider the simple act of a user contacting the business by
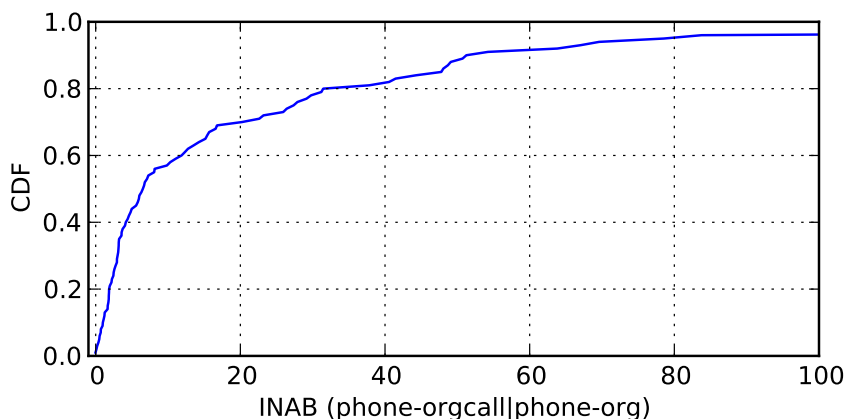
Figure 4.16. Listing a business in local results brings clear benefit to advertisers.

clicking the call button as a conversion event for the purposes of computing the NAB. For organic listings without call buttons, however, no similar data is available. Instead, we substitute the conversion rate the advertisers obtain when they advertise on local queries. While conversions are potentially an undercount when compared to calls, we hope that using the conversion rate for ads as opposed to organic clicks restores some of the balance, offering a reasonable baseline of user engagement.

Figure 4.15(a) plots the NAB of having an organic call button (phone-orgcall) vs NAB of plain search results (phone-org) for searches performed on mobile phones. Note that since both NABs are for organic results—for which businesses do not pay—the cost term in the NAB computation is zero, and NAB reduces to conversions per impression. Figure 4.16 plots the CDF of the INAB of call button over plain results. As evident from the figures, median conversions per impression increases by a factor of 10 when a organic call result is present.

4.6.2 Ads with call extensions

When an advertiser chooses to place a call button in an ad, it is possible the ad results in a conversion off-line; i.e., rather than navigating to the website and converting,
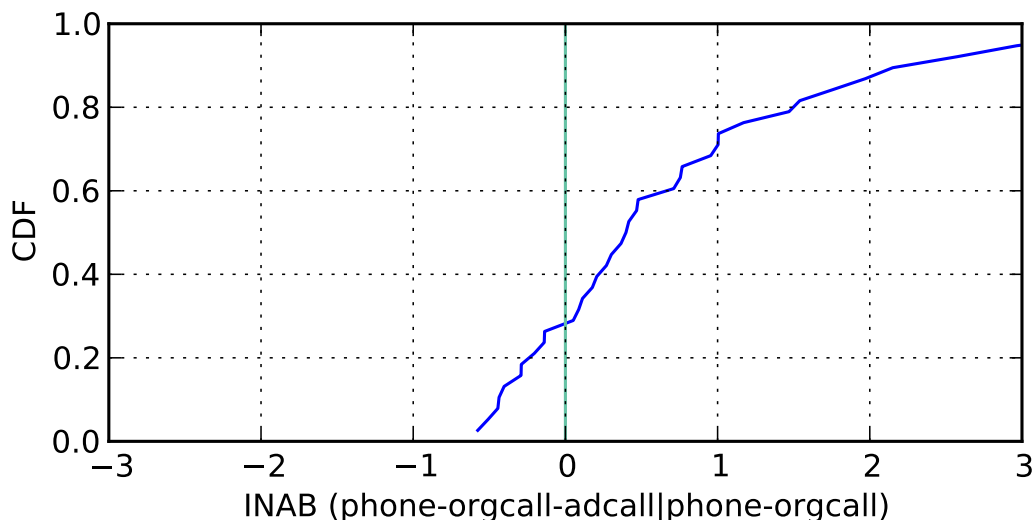
Figure 4.17. When organic call is already listed, advertising with a call ad has mixed results.

the user may instead call the advertiser and "convert" by making a purchase or similar analogous activity without further web interaction. Hence, when considering conversions for ads with call extensions, we define a conversion event to be either a call or an advertiser-reported conversion. For the very few impressions that result in both, we report only one conversion.

Figure 4.15(b) compares the NAB of mobile search ads with call extensions (phone-orgcall-adcall) vs. the NAB of organic business listings with call buttons (phone-orgcall) for the very few businesses in our dataset that both advertise call extensions as well as created an organic listing, as well as reported a sufficient number of conversion events. Figure 4.17 plots the CDF of the corresponding INAB. We find that even if organic search results contain a call button extension, ads employing the same are effective for 74% of advertisers, but ineffective for the remaining 26%. The orders-of-magnitude larger NAB (and INAB) values in this section can be attributed to the relative paucity of data for results using the call extension.

For businesses that do not have a rank high enough to be listed in the organic
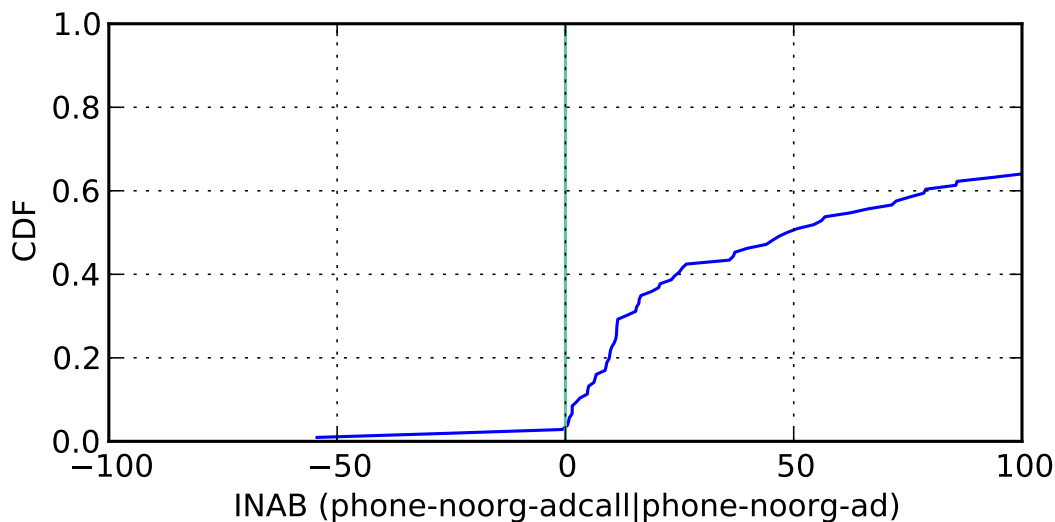
Figure 4.18. When advertiser does not have organic presence, advertising with a call button is more beneficial.

results, Figure 4.15(c) compares the NAB of mobile search ads with call extensions (phone-noorg-adcall) vs. the NAB of ads without the call extension (phone-noorg-ad). Figure 4.18 plots the CDF of the corresponding INAB. We find that ads with a call extension are universally more effective.

Preliminary data suggests that there are mixed benefits to using call extensions for businesses where organic listings include call button. For businesses that do not have a high-enough organic rank for their listing to appear in the first page of ads, however, there is a consistent boost from call extensions in search ads. That being said, since call extensions and business listings are both recent features and very few advertisers have opted-in to both, our results are preliminary and we encourage advertisers to conduct their own experiments and compute their respective INABs to assess the effectiveness of the call extension in their specific case.
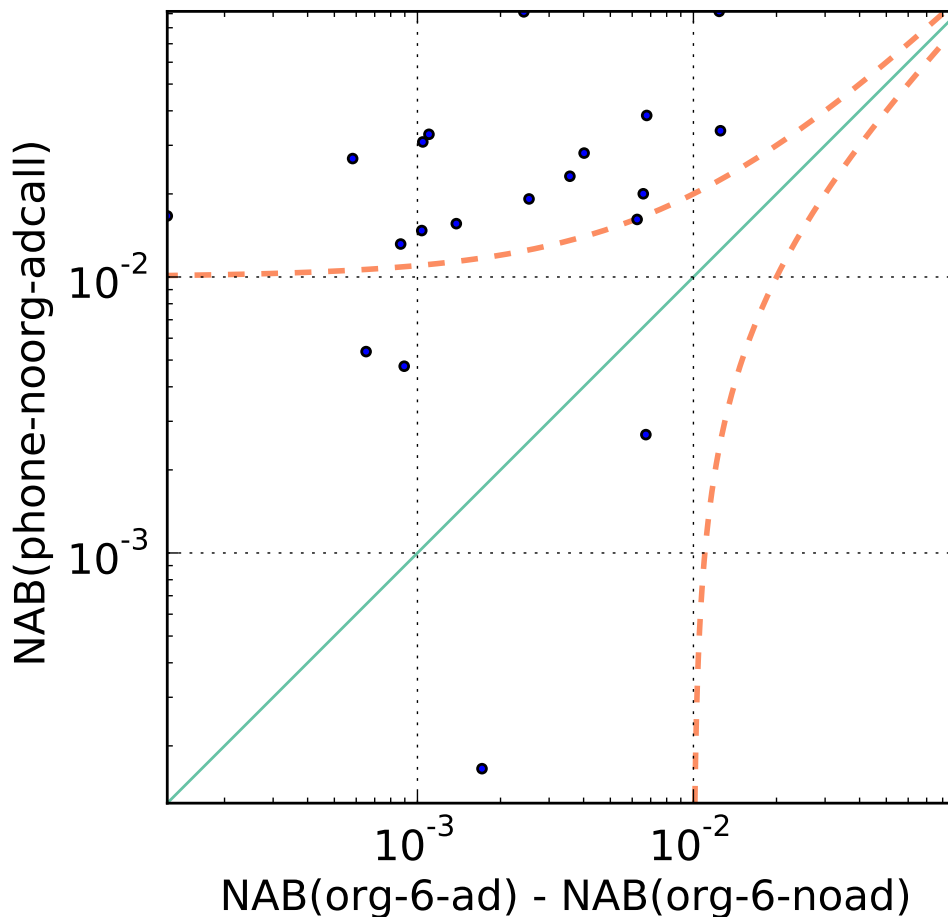
Figure 4.19. Call ads on mobile are better than regular desktop ads in the absence of organic presence.

4.6.3 Spending smarter

In closing, we compare the marginal benefit of call extensions in mobile ads to the benefit of traditional advertising on computers. As we discussed previously, the benefit of the latter depends tremendously on the position of the advertiser in the organic search results. As an optimistic estimate, we focus on an advertiser who appears sixth in the organic search results. Specifically, Figure 4.19 plots the marginal benefit of running mobile ads with call extension over organic business listings with call button (phone-noorg-adcall) vs. advertising on desktops and laptops. For 75% of the advertisers

in our dataset, it is (modestly) more beneficial to focus on mobile call ads vs. desktop ads.

While we have limited data in our dataset, our initial assessment indicates greater benefits can be obtained by adding call buttons to ads on mobile compared to traditional advertising on desktop when organic presence is poor.

4.7   Discussion

In this section, we discuss our choice of using aggregated traffic slices and the underlying systemic bias that exists in the way the traffic slices are chosen for comparison.

4.7.1   Choosing traffic slices

In order to measure the effectiveness of advertising, we identify similar search impressions with and without ads. For this, an ideal comparison would require impressions where the presence of an ad is the only distinguishing attribute. However, even the largest search engines have only sparse data when aggregated at the query level [56]. In order to have reasonable samples of impressions, we aggregate impressions by the category of user query.

4.7.2   Systemic biases

A source of bias in our analyses is that queries for which ads are shown are often more representative of the advertiser than queries where ads are not shown for comparable traffic slices. The bias is because whenever ads are shown, both the search engine and advertiser find the query relevant. But, when ads are not shown, the search engine finds query relevant to advertiser but advertiser does not. So, whenever results returned by the search engine are poor, users are less likely to choose the advertiser from organic results hence lowering NAB for the traffic slice without ads.

The exception to the bias towards queries for which are shown is when the advertiser tries to poach users looking for a competitor discussed in Section 4.5.1. This exception is because in the case of poaching, while the impressions with advertiser ad are likely to be more attractive to the advertiser, they also cost more.

Finally, in Section 4.6 queries for which the search engine presents a result with a call button are often more relevant to a mobile user than queries where a web result is delivered; hence, the NAB for organic search results with call buttons may be an overstated.

4.8   Summary

In this chapter we use net acquisition benefit (NAB) metric to approximate profit per impression in order to measure the effectiveness of an ad campaign—as defined by the targeting criteria chosen by the advertiser, and the incremental net acquisition benefit (INAB) to measure the marginal benefit of one ad strategy over another. Using these metrics and extensive search and ad click data from a major search-ad provider, we find that cannibalizing organic traffic and poaching a competitor's traffic are frequently ineffective while call extensions on mobile phones show promise.

4.9   Acknowledgements

Chapter 4 includes material that is submitted for publication as "Empirical Analysis of Search Advertising Strategies". Bhanu C. Vattikonda, Vacha Dave, Saikat Guha, Alex C. Snoeren. The dissertation author was the primary investigator and author of the paper.

Chapter 5

Conclusions

Efforts to improve sponsored search have focused on the systems used by the search engine and the role of advertisers in sponsored search. As discussed in Chapter 2, a large body of work has focused on improving the performance of sponsored search. For example, researchers have considered query expansion techniques [41, 68, 81] to identify variations of the user query to increase the number of ads that can be matched to the query. Auction systems have been studied to improve search engine revenues and identify the true value of a sponsored search ad [30, 50, 53, 77, 80]. Several authors have used data from individual advertisers to measure the benefit of online advertising strategies for the respective advertisers [34, 66, 86].

This dissertation complements these prior approaches. We propose an advertiser centered approach to improve the effectiveness of sponsored search. We argue that the performance of ad campaigns and user experience can be improved by taking advantage of the control advertisers have in targeting their ads. The focus of our work is to address the inefficiencies of a search engine and provide tools for advertisers in a way that would increase the value of sponsored search for advertisers. We now present the contributions of this dissertation and then conclude with a brief discussion of possible future directions.

5.1 Contributions

We find that accurately interpreting the ad keywords that advertisers target improves the overall relevance of ads shown to users. We use the inherent abilities of a search engine to find web results for a query to interpret the ad keywords. We then use features extracted from the web results of the ad keyword to improve the relevance ranker used by a large search engine. An improved relevance ranker would allow the search engine to identify possible relevant ads in response to user query and could lead to increased revenues for the search engine. It could also identify irrelevant ads that could have been shown to the users and eliminate them enhancing the user experience.

The significant role that advertisers play in crafting their ad campaigns also requires advertisers to get their targeting strategy right. To assist advertisers in this effort, we present a simple metric—net acquisition benefit (NAB)— to measure the success of different ad targeting strategies. Using the metric over a month long data from the same search engine, we analyze three common strategies: cannibalization, poaching and the use of call extensions on mobile devices. We find that advertisers often run campaigns which are not profitable to them. They spend their advertising budgets to target users who might already be familiar with their offerings and would have availed their products or services irrespective of the ads. Our analysis of the use of mobile call extensions using NAB tells us that advertisers can benefit from using the latest features offered by the search engine to tailor the ads they show to users on mobile devices.

To improve sponsored search effectiveness by taking advantage of the role of advertisers, this dissertation makes the following specific contributions:

1. We propose features that can be used by the relevance ranker at a search engine to improve the accuracy of the relevance ranker. We achieve a 43.2% improvement in the area under the precision-recall curve over the best published baseline and 2.7%

improvement in the production system at a large search engine.

2. We introduce a metric that can be used by advertisers and ad agencies to measure the performance of the ad campaigns that they are running with the search engine.

3. We find that for a majority of advertisers, advertising on queries where they already have an organic presence is a losing proposition. While poaching may help some of the smaller advertisers, poaching and defending against poaching provides questionable value to most advertisers.

4. Finally, advertisers could improve the performance of their ad campaigns on mobile devices by using the features offered by the search engine to customize ads for mobile device users.

5.2   Future directions

We hope that our findings encourage advertisers and search engines to take advantage of the role advertisers play in sponsored search to improve both ad effectiveness and user experience. We have evaluated the ideas we presented in Chapter 3 using offline experiments. An immediate extension of our work is to study the benefits of incorporating the features we presented over a slice of live traffic handled by a production system. Such a study would allow us to measure the impact of new features on metrics (e.g., ads per query, share of queries with ads and clicks per search impression) that have a direct bearing on user experience and search engine revenue.

Modern search engines are very good at interpreting a user query and finding content related to it. Due to their unbiased nature and the ability to look for content from a large corpus of web pages, search engines return content that is highly related to the user query. We harness this ability to interpret ad keywords on which advertisers

place their bids. We can however, go even further. For example, The ability of a search engine to identify news related results for a user query can be used to identify when the user query is related to tragic current events. If the user query is related to tragic current events, not showing any ads at all would be a prudent choice—something advertisers would prefer.

Finally, we hope that advertisers go beyond metrics like click-through-rate and cost-per-acquisition to measure the performance of campaigns. Online advertising, despite the promise of being data driven, still appears to largely rely on heuristics [77]. Search engine marketing companies peddle various metrics which justify the campaign strategies that they follow and present an inflated view of the performance of sponsored search campaigns. The results of our work and other studies, news reports [34, 26] which highlight that search advertising could lead to a false impression of effectiveness, should encourage advertisers to conduct bold experiments to measure the value delivered by sponsored search.

Bibliography

[1] Ad Copy Testing for the Google Adwords Beginner. http://www.lunametrics.com/blog/2012/06/21/ad-copy-testing-for-beginners/.

[2] Bing. http://www.bing.com/.

[3] Bing Places for Business. https://www.bingplaces.com/.

[4] DMOZ - the Open Directory Project. http://www.dmoz.org.

[5] Dynamic text: Customize your ads based on searches. https://help.bingads.microsoft.com/#apex/3/en-us/51001.

[6] English Stop Words (CSV). http://www.textfixer.com/resources/common-english-words.txt.

[7] Enhance your ad with extensions. https://support.google.com/adwords/answer/2375499.

[8] Google. https://www.google.com.

[9] Google — Year in Search 2014. http://www.google.com/trends/2014/.

[10] Google My Business. https://www.google.com/business/.

[11] Google Settles With Rosetta Stone Over Trademark Infringement Within Search And Ads Lawsuit. http://www.reuters.com/article/2012/10/31/us-usa-court-rosettastone-google-idUSBRE89U1GE20121031.

[12] Google:Update to U.S. ad text trademark policy. http://adwords.blogspot.com/2009/05/update-to-us-ad-text-trademark-policy.html.

[13] Online Ad Effectiveness Depends on Time of Day. https://advertising.yahoo.com/Blog/SEARCH-DAYPARTING-STRATEGY.html.

[14] Personal communications with Bing Ads Relevance Team.

[15] Relevance and quality guidelines - Bing Ads. http://advertise.bingads.microsoft.com/en-us/editorial-relevance-quality-guidelines.

[16] Search Engine Use 2012. http://www.pewinternet.org/files/old-media//Files/Reports/2012/PIP_Search_Engine_Use_2012.pdf.

[17] Search marketing agency price models. http://searchengineland.com/search-marketing-agency-pricing-models-11039.

[18] Set up ad customizers. https://support.google.com/adwords/answer/2375499.

[19] Smartphone user study shows mobile movement under way. http://googlemobileads.blogspot.com/2011/04/smartphone-user-study-shows-mobile.html.

[20] What are ad extensions? https://help.bingads.microsoft.com/#apex/3/en-us/51001.

[21] What average ads click through rate (ctr) is good for text ad? https://www.en.adwords-community.com/t5/Tracking-Reporting/What-average-Ads-click-through-rate-CTR-is-good-for-text-ad/td-p/150054.

[22] Wikipedia, the free encyclopedia. https://en.wikipedia.org/.

[23] Will useless clicks really cost mobile advertisers $6 billion+ by 2016? http://searchenginewatch.com/article/2207164/Will-Useless-Clicks-Really-Cost-Mobile-Advertisers-6-Billion-by-2016.

[24] Yahoo Search - Web Search. https://search.yahoo.com.

[25] Yelp Inc. http://www.yelp.com.

[26] Is google gobbling up the indian internet space? http://forbesindia.com/article/real-issue/is-google-gobbling-up-the-indian-internet-space/35641/1, July 2013.

[27] Gagan Aggarwal, Ashish Goel, and Rajeev Motwani. Truthful Auctions for Pricing Search Keywords. In *Proceedings of the 7th ACM Conference on Electronic Commerce*, EC '06, pages 1–7, New York, NY, USA, 2006. ACM.

[28] Eugene Agichtein, Eric Brill, and Susan Dumais. Improving Web Search Ranking by Incorporating User Behavior Information. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '06, pages 19–26, New York, NY, USA, 2006. ACM.

[29] Eugene Agichtein, Eric Brill, Susan Dumais, and Robert Ragno. Learning User Interaction Models for Predicting Web Search Result Preferences. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '06, pages 3–10, New York, NY, USA, 2006. ACM.

[30] Moshe Babaioff, Noam Nisan, and Renato Paes Leme. Price Competition in Online Combinatorial Markets. In *Proceedings of the 23rd International Conference on World Wide Web*, 2014.

[31] Lisa Ballesteros and W. Bruce Croft. Phrasal Translation and Query Expansion Techniques for Cross-language Information Retrieval. In *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '97, pages 84–91, New York, NY, USA, 1997. ACM.

[32] Hila Becker, Andrei Broder, Evgeniy Gabrilovich, Vanja Josifovski, and Bo Pang. What Happens After an Ad Click?: Quantifying the Impact of Landing Pages in Web Advertising. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, CIKM '09, pages 57–66, New York, NY, USA, 2009. ACM.

[33] Paul N. Bennett, Krysta Svore, and Susan T. Dumais. Classification-Enhanced Ranking. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 111–120, New York, NY, USA, 2010. ACM.

[34] Thomas Blake, Chris Nosko, and Steven Tadelis. Consumer Heterogeneity and Paid Search Effectiveness: A Large Scale Field Experiment. *NBER Working Paper*, pages 1–26, 2013.

[35] Christian Borgs, Jennifer Chayes, Nicole Immorlica, Kamal Jain, Omid Etesami, and Mohammad Mahdian. Dynamics of Bid Optimization in Online Advertisement Auctions. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, pages 531–540, New York, NY, USA, 2007. ACM.

[36] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(17):107 – 117, 1998. Proceedings of the Seventh International World Wide Web Conference.

[37] Andrei Broder. A Taxonomy of Web Search. *SIGIR Forum*, 36(2):3–10, September 2002.

[38] Andrei Broder, Massimiliano Ciaramita, Marcus Fontoura, Evgeniy Gabrilovich, Vanja Josifovski, Donald Metzler, Vanessa Murdock, and Vassilis Plachouras. To Swing or Not to Swing: Learning when (Not) to Advertise. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, CIKM '08, pages 1003–1012, New York, NY, USA, 2008. ACM.

[39] Andrei Broder, Evgeniy Gabrilovich, Vanja Josifovski, George Mavromatis, and Alex Smola. Bid Generation for Advanced Match in Sponsored Search. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, pages 515–524, New York, NY, USA, 2011. ACM.

[40] Andrei Z. Broder, Peter Ciccolo, Marcus Fontoura, Evgeniy Gabrilovich, Vanja Josifovski, and Lance Riedel. Search Advertising Using Web Relevance Feedback. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, CIKM '08, pages 1013–1022, New York, NY, USA, 2008. ACM.

[41] Andrei Z Broder, Marcus Fontoura, Evgeniy Gabrilovich, Amruta Joshi, Vanja Josifovski, and Tong Zhang. Robust Classification of Rare Queries Using Web Knowledge. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 231–238. ACM, 2007.

[42] Christopher J.C. Burges. From RankNet to LambdaRank to LambdaMART: An Overview. Technical Report MSR-TR-2010-82, June 2010.

[43] C.J.C. Burges, R. Ragno, and Q.V. Le. Learning to rank with non-smooth cost functions. In *Advances in Neural Information Processing Systems 19*. MIT Press, Cambridge, MA, January 2007.

[44] David Chan, Yuan Yuan, Jim Koehler, and Deepak Kumar. Incremental Clicks Impact Of Search Advertising. Technical report, Google, Inc., 2011.

[45] Yejin Choi, Marcus Fontoura, Evgeniy Gabrilovich, Vanja Josifovski, Mauricio Mediano, and Bo Pang. Using Landing Pages for Sponsored Search Ad Selection. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 251–260, New York, NY, USA, 2010. ACM.

[46] Massimiliano Ciaramita, Vanessa Murdock, and Vassilis Plachouras. Online Learning from Click Data for Sponsored Search. In *Proceedings of the 17th International Conference on World Wide Web*, WWW '08, pages 227–236, New York, NY, USA, 2008. ACM.

[47] Hang Cui, Ji-Rong Wen, Jian-Yun Nie, and Wei-Ying Ma. Probabilistic Query Expansion using Query Logs. In *Proceedings of the 11th international conference on World Wide Web*, pages 325–332. ACM, 2002.

[48] Cristian Danescu-Niculescu-Mizil, Andrei Z. Broder, Evgeniy Gabrilovich, Vanja Josifovski, and Bo Pang. Competing for Users' Attention: On the Interplay Between Organic and Sponsored Search Results. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 291–300, New York, NY, USA, 2010. ACM.

[49] Kushal S. Dave and Vasudeva Varma. Learning the Click-through Rate for Rare/New Ads from Similar Ads. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '10, pages 897–898, New York, NY, USA, 2010. ACM.

[50] Benjamin Edelman, Michael Ostrovsky, and Michael Schwarz. Internet Advertising and the Generalized Second Price Auction: Selling Billions of Dollars Worth of Keywords. Working Paper 11765, National Bureau of Economic Research, November 2005.

[51] Daniel C. Fain and Jan O. Pedersen. Sponsored Search: A Brief History. *Bulletin of the American Society for Information Science and Technology*, 32(2):12–13, 2006.

[52] Jon Feldman, S Muthukrishnan, Martin Pal, and Cliff Stein. Budget Optimization in Search-based Advertising Auctions. In *Proceedings of the 8th ACM Conference on Electronic Commerce*, EC '07, pages 40–49, New York, NY, USA, 2007. ACM.

[53] Sharad Goel, Sbastien Lahaie, and Sergei Vassilvitskii. Contract Auctions for Sponsored Search. In Stefano Leonardi, editor, *Internet and Network Economics*, volume 5929 of *Lecture Notes in Computer Science*, pages 196–207. Springer Berlin Heidelberg, 2009.

[54] Google. Adwords help – measuring return on investment (ROI). https://support.google.com/adwords/answer/1722066?hl=en.

[55] Layla Hasan, Anne Morris, and Steve Probets. Using Google Analytics to Evaluate the Usability of E-Commerce Sites. In Masaaki Kurosu, editor, *Human Centered Design*, volume 5619 of *Lecture Notes in Computer Science*, pages 697–706. Springer Berlin Heidelberg, 2009.

[56] Dustin Hillard, Stefan Schroedl, Eren Manavoglu, Hema Raghavan, and Chirs Leggetter. Improving Ad Relevance in Sponsored Search. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, WSDM '10, pages 361–370, New York, NY, USA, 2010. ACM.

[57] Kai Hui, Bin Gao, Ben He, and Tie-Jian Luo. Sponsored search ad selection by keyword structure analysis. *ECIR 2013, LNCS 7814, pp. 230241, 2013*, 2013.

[58] IAB. IAB internet advertising revenue report. http://www.iab.net/media/file/IAB_Internet_Advertising_Revenue_FY_2014.pdf.

[59] IAB. IAB Internet Advertising Revenue Report. http://www.iab.net/media/file/resources_adrevenue_pdf_IAB_PwC_2004full.pdf.

[60] Bernard J Jansen and Tracy Mullen. Sponsored search: an overview of the concept, history, and technology. *International Journal of Electronic Business*, 6(2):114–131, 2008.

[61] Bernard J Jansen and Marc Resnick. An examination of searcher's perceptions of nonsponsored and sponsored links during ecommerce Web searching. *Journal of the American Society for Information Science and Technology*, 57(14):1949–1961, 2006.

[62] Bernard J. Jansen and Marc Resnick. An Examination of Searcher's Perceptions of Nonsponsored and Sponsored Links During Ecommerce Web Searching. *Journal of the American Society for Information Science and Technology*, 57:1949–1961, 2006.

[63] Bernard J Jansen and Simone Schuster. Bidding on the buying funnel for sponsored search and keyword advertising. *Journal of Electronic Commerce Research*, 12(1):1–18, 2011.

[64] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. Accurately Interpreting Clickthrough Data As Implicit Feedback. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '05, pages 154–161, New York, NY, USA, 2005. ACM.

[65] Thorsten Joachims and Filip Radlinski. Search Engines That Learn from Implicit Feedback. *Computer*, 40(8):34–40, August 2007.

[66] Garrett A Johnson, Randall A Lewis, and David H Reiley. Location, Location, Location: Repetition and Proximity Increase Advertising E ectiveness. 2014.

[67] A. Joshi and R. Motwani. Keyword Generation for Search Engine Advertising. In *Data Mining Workshops, 2006. ICDM Workshops 2006. Sixth IEEE International Conference on*, pages 490–496, Dec 2006.

[68] Pawel Kowalczyk, Ingrid Zukerman, and Michael Niemann. Analyzing the Effect of Query Class on Document Retrieval Performance. In GeoffreyI. Webb and Xinghuo Yu, editors, *AI 2004: Advances in Artificial Intelligence*, volume 3339 of *Lecture Notes in Computer Science*, pages 550–561. Springer Berlin Heidelberg, 2005.

[69] RandallA. Lewis and DavidH. Reiley. Advertising Effectively Influences Older Users: How Field Experiments Can Improve Measurement and Targeting. *Review of Industrial Organization*, 44(2):147–159, 2014.

[70] Donald Metzler, Susan Dumais, and Christopher Meek. Similarity Measures for Short Segments of Text. In Giambattista Amati, Claudio Carpineto, and Giovanni Romano, editors, *Advances in Information Retrieval*, volume 4425 of *Lecture Notes in Computer Science*, pages 16–27. Springer Berlin Heidelberg, 2007.

[71] Mandar Mitra, Amit Singhal, and Chris Buckley. Improving Automatic Query Expansion. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '98, pages 206–214, New York, NY, USA, 1998. ACM.

[72] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank Citation Ranking: Bringing Order to the Web. Technical Report 1999-66, Stanford InfoLab, November 1999. Previous number = SIDL-WP-1999-0120.

[73] Beatriz Plaza. Google Analytics for Measuring Website Performance. *Tourism Management*, 32(3):477 – 481, 2011.

[74] Martin F Porter. An algorithm for suffix stripping. *Program*, 14(3):130–137, 1980.

[75] Filip Radlinski, Andrei Broder, Peter Ciccolo, Evgeniy Gabrilovich, Vanja Josi-fovski, and Lance Riedel. Optimizing Relevance and Revenue in Ad Search: A Query Substitution Approach. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, pages 403–410, New York, NY, USA, 2008. ACM.

[76] Hema Raghavan and Dustin Hillard. A Relevance Model Based Filter for Improving Ad Quality. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 762–763. ACM, 2009.

[77] Benjamin Rey and Ashvin Kannan. Conversion Rate Based Bid Adjustment for Sponsored Search. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 1173–1174, New York, NY, USA, 2010. ACM.

[78] Matthew Richardson, Ewa Dominowska, and Robert Ragno. Predicting Clicks: Estimating the Click-through Rate for New Ads. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, pages 521–530, New York, NY, USA, 2007. ACM.

[79] Paat Rusmevichientong and David P. Williamson. An Adaptive Algorithm for Selecting Profitable Keywords for Search-based Advertising Services. In *Proceedings of the 7th ACM Conference on Electronic Commerce*, EC '06, pages 260–269, New York, NY, USA, 2006. ACM.

[80] Amin Sayedi, Kinshuk Jerath, and Kannan Srinivasan. Competitive Poaching in Sponsored Search Advertising and Its Strategic Impact on Traditional Advertising. *Marketing Science*, 33(4):586–608, 2014.

[81] D. Sculley, Robert G. Malkin, Sugato Basu, and Roberto J. Bayardo. Predicting Bounce Rates in Sponsored Search Advertisements. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, pages 1325–1334, New York, NY, USA, 2009. ACM.

[82] D. Sculley, Robert G. Malkin, Sugato Basu, and Roberto J. Bayardo. Predicting Bounce Rates in Sponsored Search Advertisements. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, pages 1325–1334, New York, NY, USA, 2009. ACM.

[83] Yang Song, Hao Ma, Hongning Wang, and Kuansan Wang. Exploring and Exploiting User Search Behavior on Mobile and Tablet Devices to Improve Search Relevance. In *Proceedings of the 22Nd International Conference on World Wide Web*, 2013.

[84] Ellen M. Voorhees. Query Expansion Using Lexical-semantic Relations. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '94, pages 61–69, New York, NY, USA, 1994. Springer-Verlag New York, Inc.

[85] Qiang Wu, Chris Burges, Krysta Svore, and Jianfeng Gao. Adapting boosting for information retrieval measures. *Information Retrieval*, 13(3):254–270, June 2010.

[86] Sha Yang and Anindya Ghose. Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence? *Marketing Science*, 29(4):602–623, 2010.

[87] Wei Vivian Zhang, Xiaofei He, Benjamin Rey, and Rosie Jones. Query Rewriting Using Active Learning for Sponsored Search. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '07, pages 853–854, New York, NY, USA, 2007. ACM.