Characterizing Query Intent From Sponsored Search Clickthrough Data

Azin Ashkan, Charles L.A. Clarke
University of Waterloo, Canada
{aashkan, claclark}@cs.uwaterloo.ca

Eugene Agichtein, Qi Guo
Emory University, United States
{eugene, qguo3}@mathcs.emory.edu

ABSTRACT

Understanding the intention underlying users' queries may help personalize search results and therefore improve user satisfaction. If a commercial intent exists, and if an ad is related to the user's information need, the user may click on that ad. In this paper, we develop a methodology for using ad clickthrough logs from a commercial search engine to study characteristics of commercial intent. The findings of our study suggest that ad clickthrough features, such as deliberation time, are effective in detecting query intent. We also study the effect of query type and the number of displayed ads on ad clickthrough behavior.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms

Query Intent, Sponsored Search

Keywords

Ad Targeting, Query Log, Clickthrough

1. INTRODUCTION

Intent detection is one of the crucial long-standing goals of information access. Understanding the intent underlying user queries may help personalize search results and therefore improve user satisfaction. User intent may correspond to any of the standard categories of Web query [2]: *navigational*, *informational*, and *transactional.* On the other hand, in the context of sponsored search, information providers may also wish to know whether a user has the intention to purchase or utilize a commercial service, or what is called *online commercial intention* [3]. Sponsored search has evolved to satisfy the needs of users for relevant search results and the desires of advertisers for increased traffic to their Websites. It is now considered to be among the most effective marketing vehicles available [5]. It basically operates by matching ads to queries as they are received by a search engine. These ads are displayed to the user, along with organic search results. The most common model is pay per click, where advertisers are charged based on user clicks (if any) on the displayed ads [1]. Ideally advertisers wish to bid on multiple low-cost, highly targeted keywords that will generate high clickthrough rates for their ads.

In order to identify search query intention, implicit feedback techniques take advantage of user behavior to understand their interests and preferences. Amongst the implicit feedback techniques, clickthrough-based analysis considers the history of user-submitted queries and user-selected documents on the corresponding search result pages. Bringing the query intent detection into the context of sponsored search can help advertisers to automatically create more appropriate and relevant ad content, develop better ranking ads by matching the content of the ads with the users query intent, as well as contribute to the general understanding of user intent inference and web search behavior modeling. In this regard, we divide our motivations of this work into three parts: i) detecting the intentions of queries based on ad clickthrough features, ii) estimating the average ad clickthrough rate for each query type, and iii) studying the ad clickthrough behavior of newly arriving queries in different categories of query intent.

In the first part, we define features and train a decision tree classifier to categorize queries in two dimensions: commercial-noncommercial and navigational-informational. We define a *commercial* query as a query with the underlying intention to make an immediate or future purchase of a specific product or service, while anything else falls into the *noncommercial* category. Furthermore, a *navigational* query is defined as a query with the underlying intention to locate a specific Web site or page, while an *informational* query is everything else. Rose and Levinson [14] conducted a study, developing a hierarchy of query goals with three top-level categories: informational, navigational and resource. Under their taxonomy, a transactional query as defined by Broder [2] might fall under either of their three categories, depending on details of the desired transaction. In this paper, a transactional or resource query would be subsumed under one of the two categories of either navigational or informational, as appropriate.

We aim to distinguish between different query types according to their ad clickthrough behavior. For this reason, in

SIGIR-IRA, July 2008, Singapore.

the second part of the work, the average clickthrough rate is estimated for different query types. For each query type, this estimation is performed separately according to the number of displayed ads. The results of these estimates can be used as evidence to indicate how much the number of displayed ads determines the number of ad clicks for each query type. Finally, in the last part of our work, we use the obtained average clickthrough ratio for each type of query as a means for calculating the number of ad clicks for previously unseen queries.

The remainder of the paper is organized as follows: Section 2 discusses related work. Section 3 presents a general picture of the data set and provides some details on the post-processing of the data in order to prepare it for analysis. In Section 4, we study properties of queries with respect to deliberation time and the number of displayed ads. These properties are used in our classifiers and in our prediction model. Section 5 presents our decision tree based classifiers used to identify query intent based on ad clickthrough features. Section 6 describes the proposed ad clickthrough prediction model based on query intent and on ad numbers. Finally, we conclude the findings and discuss further research possibilities in Section 7.

2. RELATED WORK

Regelson et al. [12] estimate the clickthrough rate of new ads by using the clickthrough rates of existing ads with the same bid terms or topic clusters. Similar work by Richardson et al. [13] incorporated features that depend on more than just the bid terms, including information about the ad itself, such as the length of the ad, the page the ad points to, and statistics concerning related ads. On the other hand, in recent work by Debmbsczynski et al. [4] the authors did not have access to the ad contents and keywords. They approximated the title and the body of each ad by combining all queries for which a given ad was displayed. They also use features based on the search result page (the rank of the ad and result page number) and on the ad's target URL. They used these extracted features to build a prediction model based on decision rules that they used to generate recommendations on how to improve the quality of ads. All three of these works focus on ad-based features in order to predict the clickthrough rate of new ads that would help to predict the quality of these new ads. We study the average ad clickthrough rate for queries in terms of their underlying intent using the query-based features and ad clickthrough statistics.

In the area of sponsored search, Dai et al. [3] propose a commercial query detector. They train machine learning models from two types of data sources for a given query: content of the search result page(s) and contents of the top pages returned by the search engine. Their findings indicate that frequent queries are more likely to have commercial intent. In the general context of query intention based on clickthrough data, Lee et al. [11] predict user query goals in terms of navigational and informational intents. They show that the prediction can be performed on the basis of two types of feature sets: past user-click behavior and anchor-link distribution. In this paper, we consider two dimensions of query intent, commercial/noncommercial and navigational/informational, utilizing features extracted from the ad clickthrough data for search queries.

In [6], Ghose et al. study the effect of sponsored search at a keyword level on the ad clickthrough rate. Their results indicate that while retailer-specific ads (based on navigational queries) increase clickthrough rates, brand-specific ads (based on transactional queries) decrease clickthrough rates. However, they focus on data from one advertiser only. Our work studies differences amongst queries (with different underlying intents) in terms of average ad clickthrough rates by pooling data on ads from multiple advertisements. We show that, on average, commercial-navigational queries receive more ad clicks than commercial-informational queries.

Jansen et. al. [10] study the factors influencing the ad clicks by searchers. They report that searchers have a bias against sponsored links (ad results) as compared to non-sponsored links (organic results). In other work, Jansen [9] studies the relevance of sponsored results and non-sponsored results by submitting a set of previously collected commercial queries to three major search engines. Jansen concludes that average relevance ratings for sponsored and non-sponsored links are practically the same, although the sponsored links relevance ratings are statistically higher.

3. DATA SET

The results reported in this paper are based on a data set obtained from Microsoft adCenter search and ad click logs sampled over a few months. Personally identifying information was removed from this data set. The data includes a sample of roughly 100 million search impressions, where an impression is defined as a single search result page. There is also a set of ad clicks (about 8 million) that are associated with the impression data.

Each impression and each click is described by a set of attributes. The set of attributes from the impression data used in this paper are as follows: *date and time of the impression*, *user query*, *number of ads displayed in results of the impression*, *user session ID*, and *impression ID*. The set of attributes used from the clickthrough data is as follows: *date and time of the click*, *user query*, *the target host for the clicked ad*, *user session ID*, and *impression ID*.

3.1 Data Post-processing

Queries are assumed to be in the English language. We removed any extra space at the beginning and end of the queries, and between words of the queries for both the impression and the clickthrough files. We then case-normalized the queries. We found about 27 million queries occurring only once in the impression file, mostly with no ads. Such queries were removed from the impression data. Impressions with a duplicate combination of impression id and user session id were removed in order to filter out repeated queries from the same user. Consequently, we ended up with about 25 million unique queries in the impression data set (about 75 million unique impressions) and about 2.4 million unique queries for which there was at least one ad click recorded in the click data.

In order to prevent train-test contamination, we split the impression and clickthrough data into three equal-sized sets (train, test, validation) on a query-level. In other words, all the impressions and click data for a given query went into the same set. This process was achieved by randomly assigning each query (with all its impression and click information) into one of the three sets. All the three sets

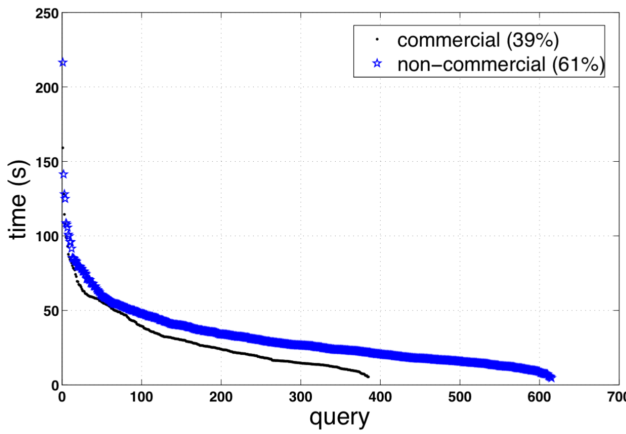

Figure 1: Deliberation Time between Entering a Query and Clicking on an Ad for that Query (for Queries Manually Labeled as Commercial/Noncommercial)

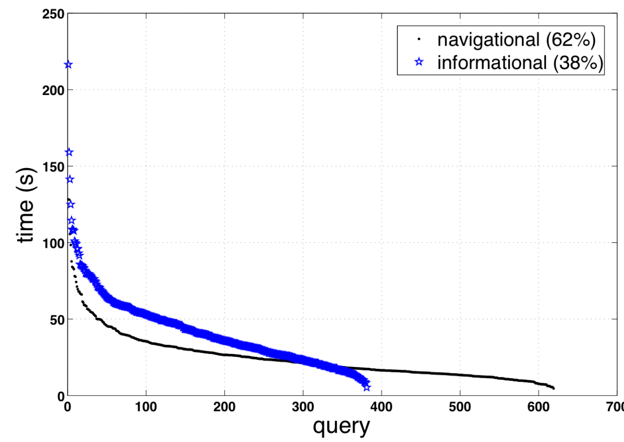

Figure 2: Deliberation Time between Entering a Query and Clicking on an Ad for that Query (for Queries Manually Labeled as Navigational/Informational)

contain approximately the same number of queries (about 8.3 million). As mentioned before, there are many queries with very small number of ad clicks. Similar to Richardson et al. [13], since our analysis deals with empirical ad clickthrough of queries, it may be wildly different from the true clickthrough rate for queries with few number of ads, leading to noise in the training and testing processes. Hence, we further filtered the three sets to include only those queries that have at least four ad clicks. After the filtering, we ended up with 44,941, 45,032, and 44,909 queries in the test, train, and validation sets respectively (134,882 queries in total).

In the remainder of the paper, we will refer to the case/ space/ user normalized impression data and its corresponding clickthrough information as the *original* data. Otherwise, by impression or clickthrough data, we mean one of the three sets (i.e. train, test, and validation) of impressions and their corresponding clickthrough data created from the original one.

3.2 Labeling Process

The original impression data was sorted based on the time of the impression. Starting from an arbitrary point in the file (approximately 1/5 of the length of the file from the beginning), 1000 queries were selected for which: i) the query was contained in the training data, and ii) the ad click frequency of the query was greater than 10. Each selected query was then manually labeled as commercial/noncommercial and navigational/informational by one of the authors.

If the assumed purpose of a query was to locate a specific Web site or page, the query was labeled as "navigational". Everything else was considered as "informational". We ended up with 62% of the queries labeled as navigational and 38% labeled as informational. The person who labeled the data was responsible for judging the assumed commercial intent of the search queries from the perspective of a user as well. If the assumed purpose of submitting a query was to make an immediate or future purchase of a product or service, the query was labeled as "commercial". Otherwise, if the purpose of the query was assumed to have little to do with commercial activity, it was labeled as "noncommercial". Since we focus on queries with at least a few number of ad clicks (11 for the labeled data and 4 in general for the three sets), one could consider that the data set construction favors including only commercial queries. In that case, commercial and noncommercial could be considered as "strongly commercial" and "slightly commercial" respectively. As a result of the labeling, 39% of the queries were labeled as commercial and 61% were labeled as noncommercial. Moreover, the focus of this paper is on the ad clickthrough-based features, where they should not be that meaningful for queries with no or relatively few displayed ads leading to noises in the classification. Therefore, we stick to classifying our filtered set of queries from the commercial/noncommercial perspective. It is also worth mentioning that the labeling result (specially for commercial/noncommercial) is subjective. In order to have a more confident result, a further exploration of this work could use multiple annotators in order to assign the final labels based on the maximum agreement among the annotators.

4. INITIAL ANALYSIS OF THE DATA

In this section, we study some properties of the data set for use in further experiments. One property is deliberation time, the average time between a query and an ad click. The other one is the average clickthrough rate for all impressions for which a particular number of ads are displayed.

4.1 Time Analysis on the Labeled Data

For each hand-labeled query, the average deliberation time for that query was calculated. The plots for each of the dimension of query type are depicted in Figures 1 and 2, where the queries are sorted by decreasing deliberation time. According to Figure 1, the deliberation time for a commercial query is generally less than for a noncommercial one. We explain this observation according to the intuition that ads basically target commercial queries more than noncommercial queries. In other words, it is more likely to find a related

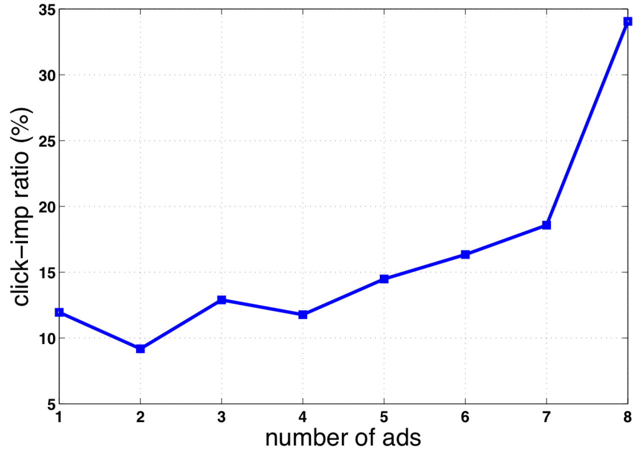

Figure 3: Average Click to Impression Ratio for Impressions with a Particular Number of Ads (lines do not imply interpolation)

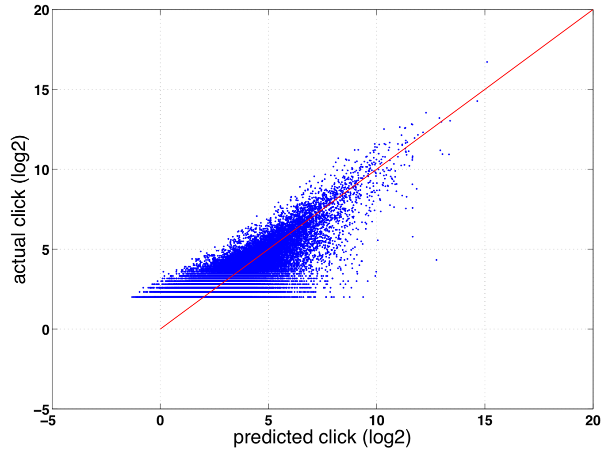

Figure 4: The Actual Number of Clicks vs. the Estimated Number of Clicks

answer among the ads for commercial queries comparing to noncommercial ones. Therefore, finding the related answers to commercial queries and thus clicking on them should take less time on average than the case for noncommercial queries. On the other hand, when it comes to navigational versus informational queries (as depicted in Figure 2), users would spend less time for navigational queries in comparison to informational ones. Because, there is usually one answer for a navigational query, often presented at the top of the result list. Users would click on their desired target as soon as they find it among the displayed ads.

Generally speaking, it appears that deliberation time can be considered as an important feature for distinguishing commercial queries from noncommercial ones, and navigational queries from informational ones. Therefore, we provide it as a feature to the classifiers described later in the paper.

4.2 Click to Impression Ratio for Varying Number of Ads

The average number of clicks per impression (clickthrough rate) for queries with a particular number of ads was calculated for the training set. In order to do that, the impressions are sorted according to the number of ads displayed for each. The number of ads in the impression file varies from one to eight. Thus, impressions are divided into eight groups, each denoted as set A_i, where i is the number of displayed ads for the impressions in that set. The value $|A_i|$ indicates the number of impressions with i ads displayed. We use the unique id number for each impression (impression id) to find out whether it resulted in an ad click. Repeating this process for all impressions in the eight groups, we can calculate the total number of ad clicks resulting from the impressions in each group.

Let $id_i^j \in A_i$ denote the unique id for the j^{th} impression in A_i. We define c_i^j to represent whether there was an ad click resulting from such an impression. In other words, $c_i^j = 1$, if there is an ad click associated with id_i^j in the clickthrough data, and $c_i^j = 0$ otherwise. Hence, the average number of ad clicks per impression (clickthrough rate), CTR_i, for queries with a particular number of ads i is obtained as follows:

$$CTR_i = \frac{\sum_{j=1}^{|A_i|} c_i^j}{|A_i|} \qquad 1 \le i \le 8 \tag{1}$$

We calculated the average clickthrough rate for the eight ad-based groups of the train set which resulted in the plot depicted in Figure 3. For clarity of presentation, we connect the points for each particular number of ads, while the lines do not imply the interpolation of points.

We use the obtained rates for the training set in order to estimate the number of clicks for each query in all the three sets (train, test, and validation). Such an estimation is performed based on the number of ads displayed for each query (thus, the average clickthrough rate corresponding to that ad#) and the number of unique impressions in which the query appears.

For a given query q in each set (train, test, or validation), let imp_q^i denote the number of times query q appears in the impressions with i number of ads. In Equation 1, we estimated the average ad clickthrough rate for such a query as CTR_i. Thus, the estimated number of ad clicks for such a query is calculated as follows:

$$click_q = \sum_{i=1}^{8} CTR_i \times imp_q^i \tag{2}$$

We simply pass through all the impressions for a query and multiply the number of impressions with a particular number of ads by the value of the obtained average clickthrough rate corresponding to that particular number of ads. As the ads number varies from one to eight, we add the eight different values for each query together in order to estimate the total click number for the query.

Figure 4 depicts the actual number of clicks versus the estimated number of clicks in the *test* set. The plot is presented in log-log basis. As can be seen in the plot, 55% of the queries appear above the line $y = x$ in the plot, meaning that their actual number of clicks is greater than that which was estimated for them. The plots for the other two sets (train and validation) follow the same pattern as the test set, so we do not present them here. As is shown in the figure, the predicted number of clicks and the actual number of clicks are correlated, particularly as the number of clicks increases. This observation could indicate that the number of ads actually determines the number of clicks, at least in part. We further study this issue later in the paper, when we discuss query intent.

5. CLASSIFYING QUERY INTENT

As mentioned previously, two dimensions of query intent are studied in this work: commercial/noncommercial and navigational/informational. We utilize decision trees for our intent classification process and stick to a set of features extracted from our large sets of impression and ad clickthrough data. In similar work with sponsored search data by Richardson et. al [13], some features of the ad itself, such as the structure of the landing page and bid terms, were used in order to estimate the clickthrough rate for a given ad. Since we are concerned with the query intent, we limited our feature set to those related to queries and their ad clickthrough information.

As Lee et. al in [11] suggest, predicting user query goals can be performed based on the two type of feature sets, past user-click behavior and anchor-link distribution. We considered similar types of features based on what we had available in the given ad clickthrough data. Some of the features were first normalized and then fed to the classifiers for both types of query intents. The set of features used for each query are as follows:

- The query length in terms of the number of characters in a query.
- A feature, namely *URL-element*, is set to 1 if the query has any URL element, such as .com, .org, .ca, and etc, otherwise it is set to 0.
- Number of target hosts is calculated as the number of different ad links that were clicked for the query.
- Average number of clicks per target host (namely *avg*, which is equal to the number of ad clicks for the query divided by the number of different hosts clicked for that query).
- Significance of the query's most frequent target host (the number of times a click happens on the most frequent target host as a result of the query, divided by *avg*).
- The level of decrease in clicks between the top two frequent target hosts for the query (the number of times click happens on the most frequent target host as a result of the query divided by the number of times the top second frequent target host receives click as a result of the same query).
- Click rate defined as the ratio of the total number of ad clicks resulting from the query against the total number of impressions in which the query appeared.
- Number of target hosts of which the query string is a substring divided by the total number of different hosts clicked for that query.

Table 1: Prediction Accuracy

Classifier	Query Intent	Precision	Recall	Accuracy
A	Commercial	0.77	0.61	72.5%
	Noncommercial	0.70	0.82	
B	Navigational	0.82	0.87	83.27%
	Informational	0.85	0.79	

- Total number of clicks on target hosts of which the query string is a substring divided by the total number of ad clicks for that query.
- The difference between the query impression time and its ad click time on average (the deliberation time).

The above features have been extracted for the 45,032 queries of the test set, and the 44,909 queries of the validation set, and also the 1000 labeled queries selected from the train set. The 1000 labeled queries along their features were first fed to a C4.5 decision tree (using the WEKA tool [7]) in order to train the classifier (separately for each dimension of query intent). We applied the 10-fold cross validation method on the labeled queries to measure the accuracy of our classifiers. A report of the prediction accuracy for the commercial/noncommercial classifier (A) and the navigational/informational classifier (B) is presented in Table 1. Afterwards, the test and validation queries (total of 89,941 unlabeled queries) were passed to each classifier in order to predict their types.

As a result of this intention classification, each query will fall into one of the following four intention categories: i) commercial and navigational, ii) commercial and informational, iii) noncommercial and navigational, and iv) noncommercial and informational. We will consider the first two of these 2-dimensional intentions for each query in our future analyses in order to characterize different queries commercial intent according to their ad clickthrough information.

6. CLICK PREDICTION BY NUMBER OF ADS AND QUERY INTENTS

The average clickthrough rate for particular number of ads should be different for various types of queries. In this section, we study this issue and use it as a basis for predicting the number of ad clicks for a given query.

6.1 Click to Impression Ratio by Query Intent

At this point, we follow a similar approach to what we did in calculating the average click to impression ratio for all the impressions with particular number of ads in the training set. However, this time, we consider only the impressions for which their associated queries are of a particular type. Note that we calculate the ratio values for the queries in the *training* set, and later we will use these values associated with particular number of ads and query intent in order to estimate the number of ad clicks for queries in the *test* set.

The average clickthrough rates for the four general types of queries (i.e. commercial, noncommercial, navigational, and informational) with the particular number of ads are plotted in Figure 5. The same analysis is performed for two pairs of query types (queries that are either commercial-navigational or those that are commercial-informational),

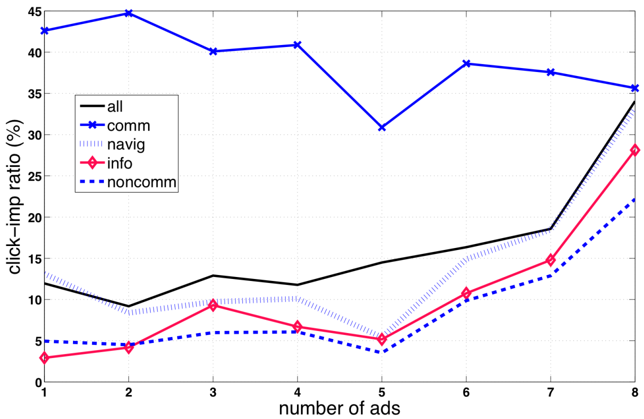

Figure 5: Average Click to Impression Ratio for Commercial, Noncommercial, Navigational, Informational, and All Types of Queries at a Particular Number of Ads (lines do not imply interpolation)

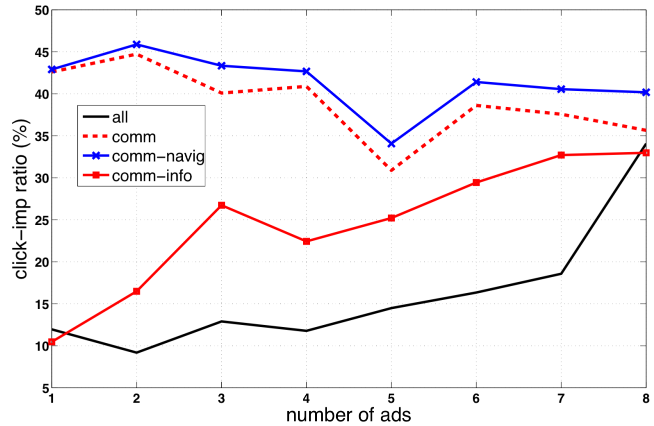

Figure 6: Average Click to Impression Ratio for Commercial, Commercial-Navigational, Commercial-Informational, and All Types of Queries at a Particular Number of Ads (lines do not imply interpolation)

which results in the two plots depicted in Figure 6 along with the plot for general commercial queries. The plot from Figure 3 is also placed in Figures 5 and 6 to provide a baseline for comparison. The plots for noncommercial-navigational and noncommercial-informational queries are not presented, because we are interested in studying the behavior of commercial queries and the effects of navigational and informational intent on user behavior for these commercial queries.

According to both figures, for most of the query types, the more ads are displayed as results of a query, the more clicks they receive. Moreover, note that the commercial query type is the leader (Figure 5) in terms of click frequency for all number of displayed ads. This frequency is larger (Figure 6) when the commercial query is also navigational rather than informational.

There are some peaks and valleys in the plots that could be because of the location of different ads (top or side of the result page) for which the clicks are recorded. According to Jansen [8], top-listed ads are assumed to be more relevant than organic results and side-listed ads. This could affect the frequency of clicks for ads at different locations and could be the cause of bumps at some points of the plots. The location of the ads is not available to us, however we believe it should be the subject for further study on this issue.

As depicted in Figure 5, navigational queries receive more ad clicks than informational queries on average. Similarly, Figure 6 conveys that commercial-navigational queries receive more ad clicks than commercial-informational queries on average. Our intuition for explaining both observations would be the fact that a query is navigational restricts the top result links (either ad or organic results) to a particular website. Therefore, the top results for a navigational query would more likely match with what user is seeking. This could result in more ad clicks for navigational queries in comparison to the informational queries. The difference is even larger when the user's intent is also commercial, because the target of click for a commercial query (in this case, commercial-navigational) is most likely for an ad. This could make a larger difference in number of ad clicks between the commercial-navigational queries and the commercial-informational queries as compared to the navigational queries and the informational queries in general.

A good example for illustrating this difference is "American airlines" as a commercial-navigational query against "airline tickets" as a commercial-informational query. The chance that user would find a related ad for the former query is greater than the later one, because the former query is restricted by the airline name. As Jansen suggests [8], searchers approach e-commerce searching from two major perspectives, one to look for a specific product or service, and the other to detect information. We believe that the commercial-navigational queries fall mostly into the former category, while the commercial-informational queries fall into the latter one.

Figure 6 shows similar clickthrough rates for commercial and commercial-navigational queries with one displayed ad (and also for two ads), while the rate for the commercial-informational queries is much smaller. We looked into this issue and found that out of 160,040 impressions for the commercial queries with one displayed ads in the training set, only 1,462 belong to commercial-informational queries while the rest (158,578 impressions) belong to the commercial-navigational queries. Moreover, 153 clicks were recorded for 1,462 impressions with one displayed ads for the commercial-informational queries (i.e. 10.5 % clickthrough rate). These numbers are 68,025 out of 158,578 (i.e. 42.9 % clickthrough rate) for the commercial-navigational queries which are relatively high.

We hypothesize that entering a navigational query (in this case, commercial-navigational query) results in a specific highly-related page. Hence, if only one ad is supposed to be displayed for such a query, it will most likely be the same as (or highly related to) that single page. Therefore, the impressions for which only one ad is listed correspond most closely to the commercial-navigational queries rather than commercial-informational queries. In other words, comparing to the commercial-navigational queries, the commercial-informational queries provide more chance so that various ads will be displayed as the result of the queries.

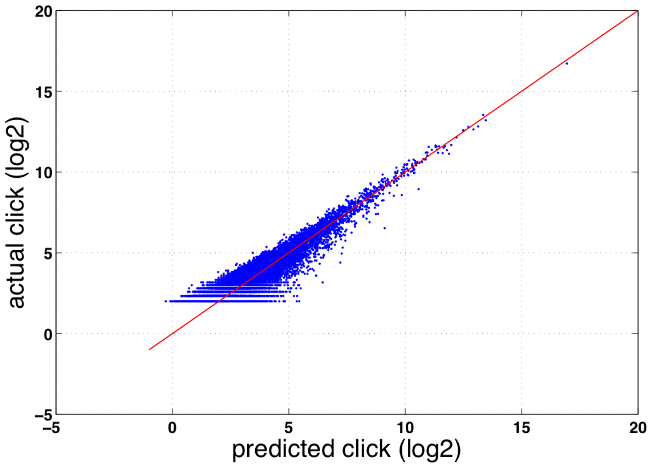

Figure 7: The Actual Number of Clicks vs. the Estimated Number of Clicks for the Commercial Queries

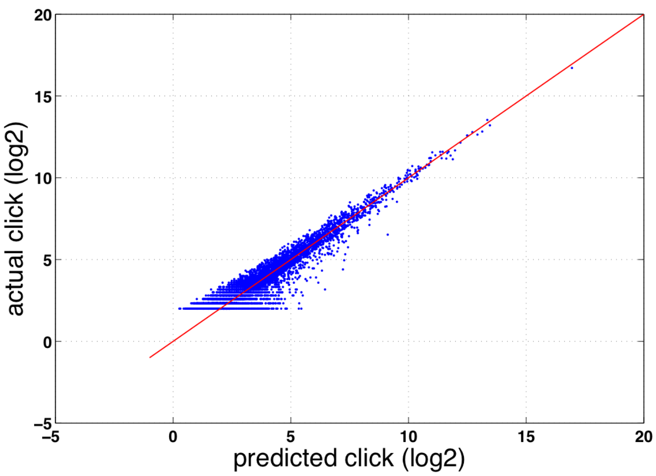

Figure 8: The Actual Number of Clicks vs. the Estimated Number of Clicks for the Commercial-Navigational Queries

6.2 Click Prediction

Using the average clickthrough rate obtained for the two dimensions of query types from the training set, we now focus on the test set. Recall that all the queries in this set have 4 or more clicks recorded for them in the click data. Also note that we are generally interested in behavior of commercial queries in the domain of sponsored search (commercial-navigational and commercial-informational queries, more specifically).

Let CTR_i^{cn} and CTR_i^{ci} be the average clickthrough rates for impressions with i number of ads that belong to commercial-navigational and commercial-informational queries respectively. Similarly, let CTR_i^{nn} and CTR_i^{ni} be the average clickthrough rates for impressions with i number of ads that belong to noncommercial-navigational and noncommercial-informational queries respectively. For a given query $q \in Q$, where Q is the set of all queries, we define function t

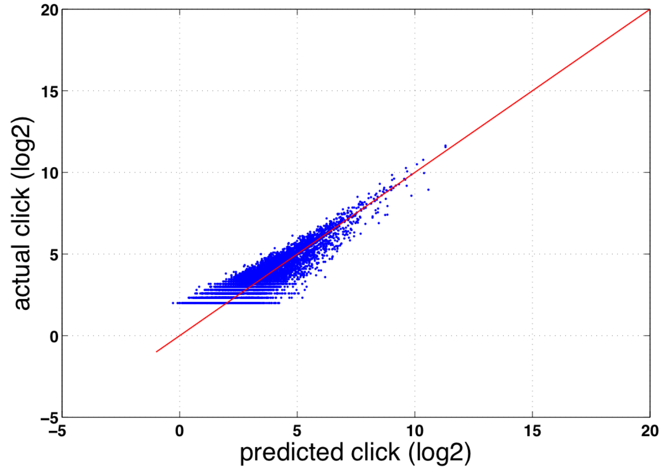

Figure 9: The Actual Number of Clicks vs. the Estimated Number of Clicks for the Commercial-Informational Queries

as $t : Q \rightarrow T$. T is the set of pairs of query intents which are among the followings: commercial-navigational, commercial-informational, noncommercial-navigational, and noncommercial-informational. According to our previous notation, we consider $T = \{\text{cn, ci, nn, ni}\}$. Based on what we just defined, the proposed prediction strategy obtains the query intents using the function t. It then uses the average rate corresponding to that category of intents in order to calculate the estimated number of clicks by going through all the impressions (similar to the Equation 2) of the query:

$$click_q^{int} = \sum_{i=1}^{8} CTR_i^{t(q)} \times imp_q^i \tag{3}$$

where $click_q^{int}$ is the estimated number of clicks based on the proposed prediction model which considers the average clickthrough rate for different query intents.

The plots for the commercial queries are presented in Figure 7. As is shown in the figure, the predicted number of clicks and the actual number of clicks are more correlated than the baseline depicted in Figure 4. We measured the correlation for each plot by calculating the covariance of the two data sets (the predicted clicks versus the actual clicks), where a perfect prediction with all the points on the line $y = x$ would result in correlation equal to 1. The correlation for the plots in Figure 4 is calculated as 0.786, while the one for the commercial queries (Figure 7) is 0.927. This may indicate that the number of ads represents a major factor in determining the number of clicks for commercial queries.

To further study the effectiveness of the number of ads in such an intention-based prediction, we plotted the actual number of clicks versus the predicted clicks for commercial-navigational and commercial-informational queries in Figures 8 and 9 respectively. These two plots confirm our previous statement that commercial-navigational queries receive on average more ad clicks than commercial-informational queries. Moreover, as depicted in Figure 8, the actual number of ad clicks are greater than the number predicted by our model for most of the commercial-navigational queries. This could indicate that the number of ads determine the number of ad clicks for commercial-informational queries more effectively than queries that are commercial and navigational. However, the covariance measure reports a slightly lower correlation for the former one compared to the later one (0.903 versus 0.950). Investigating the reason behind these observations is a direction for future work.

7. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, we develop a methodology to use ad clickthrough logs from Microsoft adCenter Search ad click logs in order to study characteristics of different query intents. The findings of our study suggest that ad clickthrough features, such as the deliberation time between entering a query and clicking on an ad for that query, are effective in detecting different query intents. Utilizing this feature, along with other query-based and ad clickthrough features, we trained a decision tree to classify queries in two dimensions: commercial/noncommercial and navigational/informational. The average clickthrough rate is then estimated for different query types. The obtained rates were used to predict clickthrough rate for a given query with particular intentions and various number of ads (one to eight) displayed as the result of the query. All in all we can list our findings as follows:

- Users spend more time on average for noncommercial queries than commercial queries to find the related ad (if any) to click on.
- Users spend more time on average for informational queries than navigational queries to find a related ad to click on.
- According to the accuracy of our decision tree based classifiers (Table 1), ad clickthrough-based features correlate with query categories: commercial/noncommercial and navigational/informational.
- For most of the query types, the more ads displayed as results of a query, the more clicks they receive. However, commercial-navigational queries receive more ad clicks for all number of displayed ads compared to other types of queries.
- The number of displayed ads affects the number of ad clicks for each category of query intents differently. It seems this factor is more effective in predicting the ad clickthrough rate for commercial queries, especially the commercial-informational ones, compared to the others.

A possible future direction for this work is studying the organic clickthrough behavior as to whether it follows the behavior we have reported in this paper. As mentioned before, the average click to impression ratio for some query types has some bumps at particular number of ads (usually five). It is worth looking into this issue, to determine if the location of the clicked ads has anything to do with this observation. Another possible direction for this work would be to study the possibility of whether the clickthrough data can be exploited in the labeling process. Finally, the reasons behind these behaviors can be further explained.

8. ACKNOWLEDGMENTS

We would like to thank Microsoft Research and Microsoft adCenter for the Beyond Search data sets, and for partially supporting this work.

9. REFERENCES

[1] R. Briggs and N. Hollis. Advertising on the web: Is there response before clickthrough. *Journal of Advertising Research*, 37(2), 1997.

[2] A. Broder. A taxonomy of Web search. *ACM SIGIR Forum*, 36(2):3–10, 2002.

[3] H. Dai, L. Zhao, Z. Nie, J. Wen, L. Wang, and Y. Li. Detecting Online Commercial Intention (OCI). *Proceedings of the* 15^{th} *International Conference on World Wide Web*, pages 829–837, 2006.

[4] K. Debmbsczynski, W. Kotlowski, and D. Weiss. Predicting ads clickthrough rate with decision rules. *Workshop on Target and Ranking for Online Advertising*, WWW 2008.

[5] D. C. Fain and J. O. Pedersen. Sponsored search: A brief history. *Bulletin of the American Society for Information Science and Technology*, 32(2):12–13, 2006.

[6] A. Ghose and S. Yang. An empirical analysis of sponsored search performance in search engine advertising. *Proceedings of the International Conference on Web Search and Web Data Mining*, pages 241–250, 2008.

[7] G. Holmes, A. Donkin, and I. Witten. WEKA: A machine learning workbench. *Proceedings of the* 2^{nd} *Australian and New Zealand Conference on Intelligent Information Systems*, pages 357–361, 1994.

[8] B. Jansen. The comparative effectiveness of sponsored and nonsponsored links for Web e-commerce queries. *ACM Transactions on the Web*, 1(1), 2007.

[9] B. Jansen, D. Booth, and A. Spink. Determining the user intent of Web search engine queries. *Proceedings of the* 16^{th} *International Conference on World Wide Web*, pages 1149–1150, 2007.

[10] B. Jansen, A. Brown, and M. Resnick. Factors relating to the decision to click on a sponsored link. *Decision Support Systems*, 44(1):46–59, 2007.

[11] U. Lee, Z. Liu, and J. Cho. Automatic identification of user goals in Web search. *Proceedings of the* 14^{th} *International Conference on World Wide Web*, pages 391–400, 2005.

[12] M. Regelson and D. Fain. Predicting clickthrough rate using keyword clusters. *Proceedings of the* 2^{nd} *Workshop on Sponsored Search Auctions*, 2006.

[13] M. Richardson, E. Dominowska, and R. Ragno. Predicting clicks: estimating the clickthrough rate for new ads. *Proceedings of the* 16^{th} *International Conference on World Wide Web*, pages 521–530, 2007.

[14] D. Rose and D. Levinson. Understanding user goals in Web search. *Proceedings of the* 13^{th} *International Conference on World Wide Web*, pages 13–19, 2004.