Exploring Searcher Interactions for Distinguishing Types of Commercial Intent

Qi Guo
Emory University
qguo3@mathcs.emory.edu

Eugene Agichtein
Emory University
eugene@mathcs.emory.edu

ABSTRACT

An improved understanding of the relationship between search intent, result quality, and searcher behavior is crucial for improving the effectiveness of web search. While recent progress in user behavior mining has been largely focused on aggregate server-side click logs, we present a new search behavior model that incorporates fine-grained user interactions with the search results. We show that mining these interactions, such as mouse movements and scrolling, can enable more effective detection of the user's search intent. Potential applications include automatic search evaluation, improving search ranking, result presentation, and search advertising. As a case study, we report results on distinguishing between "research" and "purchase" variants of commercial intent, that show our method to be more effective than the current state-of-the-art.

Categories and Subject Descriptors

H.3.3 [**Information Systems**]: Information Storage and Retrieval

General Terms

Design, Experimentation, Performance

Keywords

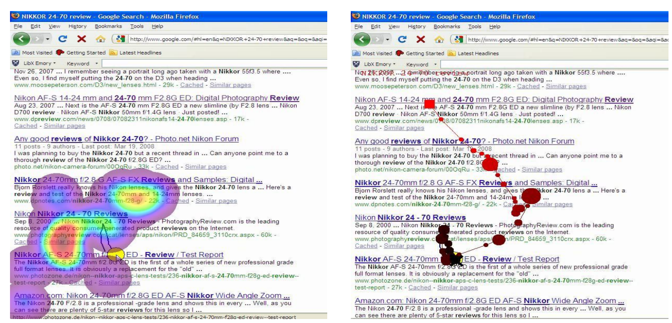
user intent inference; search behavior modeling

1. INTRODUCTION

An improved understanding of searcher needs and interests is crucial for search engines to generate satisfactory search results, with applications ranging from search evaluation to improving search ranking, presentation, and usability. While previous studies have shown the effectiveness of eye tracking to identify user interests, unfortunately, eye tracking requires expensive equipment, limiting its applicability to the laboratory setting. In this paper we attempt to infer web searcher intent by modeling mousing, clicking, scrolling, and other client-side behavioral clues that are available on a wide range of computing devices. For this, we attempt to learn a correspondence from eye tracking data obtained in a lab, to the real behavior of users "in the wild" - that is, when engaged in natural web search tasks.

What makes the problem particularly daunting is that the same query may reflect different intents not only for different users [6], but even for the same user at different times [3]. As a concrete example, consider how users with research intent examine the search engine result page (SERP) for a query "nikkor 24-70 review". This query is commercial (the searcher is probably considering whether to buy this digital camera model), but could also be research-oriented (the searcher is interested in reviews, and not yet in making an immediate purchase). Figure 1 (a) shows the gaze position "heat map" (different colors represent amount of time spent examining the corresponding page position). Figure 1 (b) shows the mouse movements performed by the subject as they were examining the SERP. This example illustrates the possible connection between user interactions on the SERP and interest in specific results or types of results, and can be more revealing than the keywords of the query or clicks on the result links.

Previous work (e.g., [1]) addressed the detection of commercial intent in the aggregate. Another dimension of work somewhat similar in approach to ours considered query chains and browsing behavior to infer document relevance (e.g., [5]). The work that is closest to ours in spirit (e.g., [4]), attempted to capture and identify user goals based on the query context. In contrast, our work explores the value of automatic mining of client-side interactions – contextualized within a search session – for detecting searcher intent.



(a) (b)

Figure 1: Searcher gaze position and corresponding mouse trajectory (Research intent)

Our Task: the goal is to detect, given a user's behavior on a SERP, whether the query had research or purchase intent. That is, each search and corresponding SERP was labeled as part of "research" or "purchase" intent; therefore, this is a test of our model to be able to correctly recover the original intent of the searcher.

Next, we introduce our techniques (Section 2) followed by preliminary experimental results and discussion (Section 3).

2. SYSTEM DESCRIPTION

We represent the user as a non-deterministic state machine where the emissions are the observable actions, and the user intent and search goals as the hidden states in the model. Therefore, searcher actions such as queries, result/ad clicks, mouse movements and button presses are observations generated by the hidden states. For example, ad clicks are more likely to be emitted if the user is in a

WWW 2010, April 26–30, 2010, Raleigh, North Carolina, USA.
ACM 978-1-60558-799-8/10/04.

purchase state, and less likely if the user is in a research state. Similarly, the emission probabilities of other observations would vary over different hidden states. Hence, we hypothesize that observations including the search context and user interactions are related to the intent states of the users.

Specifically, we consider the following information sources and the corresponding feature representations. Each feature group is described below to the extent permitted by space limitations. Complete feature specification and the dataset for our experiments is available online[1] for reproducibility.

Query group: This group of features is designed to capture the same information as was used in the previous studies of capturing clicks in the context of previous and subsequent queries [4, 5]. Specifically, we include features such as query length, and unigram (token) list for query and query URL tokens.

SERPContent: These features represent the text (and markup) content of the SERPs, specifically, the unigram (token) set of terms and markup in the SERP.

ResultQuality: These features are to capture coarse information about the SERP connection to the query, namely: how many words in result snippets match the query terms, how many words in search ad text match the query terms, as well as the number of ads.

Interaction: The features include time from page load until the first move or UI event, dwell time, mouse trajectory length, vertical and horizontal range. We also capture more precise physiological characteristics, such as speed, acceleration, rotation of the mouse movements.

Click: Captures the types and properties of result clicks and SERP revisits, features include: tokens in the clicked URL, whether the click is a Bounce click[7], a SAT click[2], number of URLs visited after a result click, average and total dwell times on each visited result URL, click type, etc.

Context: Captures where the search belongs within a task context: for example, position of search in the task (e.g., initial query vs. 3rd one), whether the query was issued within same session, and whether the query is the initial query in session, whether the query is identical to previous query, whether the query is reformulation, expansion or contraction.

The compared methods are as following:

- **Baseline**: always guess the majority class (Research).
- **QC**: similar to the state-of-the-art query chains models (e.g., [5, 4]), implemented using Query group features and trained using the SVM classifier.
- **SVM (All)**: implemented using all group features and trained using the SVM classifier.

3. RESULTS AND DISCUSSION

We performed a user study where 10 subjects were asked to have two intents: purchasing (that is, to attempt to "buy" an item of immediate interest to the subject), and research (that is, to "research" an item of interest to the subject, for some future purchase). The subjects were not restricted on time, and all the interactions, as well as gaze position, were tracked using the EyeTech TM3 integrated eye tracker, with actual sampling rate of approximately 30Hz. All the subjects were graduate and undergraduate students and university staff, that is, were technically savvy and had some experience with web search. To collect the data, we developed a simple Firefox plugin to capture the mouse movements and other user action events on search result pages.

We now report the results for detecting whether the searcher had research or purchase intent. We used the searches from 7 of the 10 subjects as training and the searches from the other 3 subjects as test. Also, we used standard-defined Precision, Recall and Macro-averaged F1 as metrics. Table 1 shows that our system SVM (All) substantially outperforms both baselines, including the QC (query chains) baseline, resulting in accuracy of almost 82% and precision (on identifying Research intent) of 92%.

To identify the most important features contributing to our performance, we performed feature ablation by removing one feature group at a time from the classifier (Table 2). All the feature groups provide significant contributions. The most important features appear to be ResultQuality and Context features: with these features removed, accuracy degrades to 70.5% from 82% with these features included. We conjecture that ResultQuality and Context features help interpret user behavior by indicating whether the searcher is succeeding in the search intent. Further investigation and additional user study is needed to fully understand the connection between various feature groups; our work is just a first step in the right direction.

Method	*Acc. (%)*	*Research*		*Purchase*		*F1*
		Prec.	*Recall*	*Prec.*	*Recall*	
Baseline	65.9	65.9	100	0	0	39.7
QC	56.8	85.7	41.4	43.3	86.7	56.8
SVM (All)	**81.8**	**92.0**	**79.3**	**68.4**	**86.7**	**80.8**

Table 1: Main Results for Predicting Research vs. Purchase

Method	*Acc. (%)*	*Research*		*Purchase*		*F1*
		Prec.	*Rec.*	*Prec.*	*Rec.*	
SVM (-Query)	72.7	81.5	75.9	58.8	66.7	70.6
SVM (-SERPContent)	75	90.9	69.0	59.1	86.7	74.3
SVM (-ResultQuality)	70.5	86.4	65.5	54.5	80.0	69.7
SVM (-Click)	72.7	81.5	75.9	58.8	66.7	70.6
SVM (-Interaction)	75	90.9	69.0	59.1	86.7	74.3
SVM (-Context)	70.5	90.0	62.1	54.2	86.7	70.1

Table 2: Feature Ablation Results for Predicting Research vs. Purchase

In summary, we have shown that by modeling interactions in context we can substantially improve the prediction of research vs. purchase intent of the user. The presented work is just a first step in contextualized user interaction modeling, eventually enabling more accurate intent inference, personalization, and targeted content delivery for the next generation of search. In the future, we plan to expand our model to consider user interactions and page context other than the search result pages. In particular, we plan to incorporate the interactions on the intermediate result pages visited between successive searches, which may provide additional contextual information.

4. REFERENCES

[1] H. K. Dai, L. Zhao, Z. Nie, J.-R. Wen, L. Wang, and Y. Li. Detecting online commercial intention (oci). In *Proc. of WWW*, 2006.

[2] S. Fox, K. Karnawat, M. Mydland, S. Dumais, and T. White. Evaluating implicit measures to improve web search. *ACM Transactions on Information Systems*, 23(2), 2005.

[3] Q. Guo, E. Agichtein, C. L. A. Clarke, and A. Ashkan. In the mood to click? towards inferring receptiveness to search advertising. In *Proc. of WI*, 2009.

[4] B. Piwowarski, G. Dupret, and R. Jones. Mining user web search activity with layered bayesian networks or how to capture a click in its context. In *Proc. of WSDM*, 2009.

[5] F. Radlinski and T. Joachims. Query chains: learning to rank from implicit feedback. In *Proc. of KDD*, 2005.

[6] D. E. Rose and D. Levinson. Understanding user goals in web search. In *Proc. of WWW*, 2004.

[7] D. Sculley, R. G. Malkin, S. Basu, and R. J. Bayardo. Predicting bounce rates in sponsored search advertisements. In *Proc. of KDD*, 2009.

[1] At `http://ir.mathcs.emory.edu/intent/data/`.