In the Mood to Click? Towards Inferring Receptiveness to Search Advertising

Qi Guo Eugene Agichtein
Mathematics & Computer Science Department
Emory University
Atlanta, USA
{*qguo3,eugene*}@*mathcs.emory.edu*

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

***Abstract*—We present a method for modeling, and automatically inferring, the current interest of a user in search advertising. Our task is complementary to that of predicting ad relevance or commercial intent of a query in the aggregate, since the user intent may vary significantly for the same query. To achieve this goal, we develop a fine-grained user interaction model for inferring searcher receptiveness to advertising. We show that modeling the search context and behavior can significantly improve the accuracy of ad clickthrough prediction for the current user, compared to the existing state-of-the-art classification methods that do not model this additional session-level contextual and interaction information. In particular, our experiments over thousands of search sessions from hundreds of real users demonstrate that our model is more effective at predicting ad clickthrough within the same search session. Our work has other potential applications, such as improving search interface design (e.g., varying the number or type of ads) based on user interest, and behavioral targeting (e.g., identifying users interested in immediate purchase).**

I. Introduction

Search advertising is one of the most lucrative industries on the Web. As Web search engines are supported largely by search advertising, the ability to serve ads at the right time and with the right content is crucial for both increasing revenue and improving user satisfaction.

In the literature and practice, most work on search advertising has focused on ad relevance and quality. However, the complementary aspect of the problem - the "right time" to serve an ad - has not yet been well studied. For example, a searcher issuing a seemingly commercial query (e.g., "Nintendo Wii") may not be interested in the search ads if the organic (non-sponsored) search results are sufficient for their needs. In this case, showing ads could annoy the searcher, and contribute to "training" them to ignore the ads. Conversely, there are cases when a (normally) non-commercial query is issued by a searcher who is in fact interested in advertising, but no ads were shown. In both situations, the searcher intent and context supersede ad relevance or quality.

In this paper, we begin to explore the modeling of the searcher receptiveness to search advertising: that is, predict whether a searcher would be interested in advertising for *future* searches in the current session, *before* these queries are even issued. Given the predictions, the search engine could adjust ad display strategy dynamically. That is, before generating the results for a search, the search engine could consider the prediction provided by our system to decide how many, and which search ads to include with the results. We introduce the problem more formally in Section II.

To model such short-term user interest, we introduce a novel session-level model of searcher advertising receptiveness using search context and fine-grained, client-side interactions. Our hypothesis is that finer grained interactions such as mouse movements, contextualized by the page content, provide information about user interest. As shown in previous work, mouse movements correlate with eye movements (Rodden et al. [23]), making mouse movements a lightweight and convenient proxy for eye tracking. Like eye movements, mouse movements can reflect short-term user interest and attention. For example, if the user is looking for a deal on a specific product, she may move the mouse faster, but when researching a product category, she may move the mouse more slowly. These factors can, in turn, influence the final user decision to click on an ad or not. To capture these interactions, we run JavaScript in a browser toolbar, allowing our system to track mouse movements, clicks, page scrolls, and other fine-grained interaction events. Note that this level of instrumentation does not require a user download or installation: JavaScript code similar to what we run in the toolbar can be easily returned in the header of a Search Engine Result Page (SERP). Therefore, all the data we collected would be available to the search engine via light-weight instrumentation on a SERP.

The implications of our new technology are substantial for the search engine industry. Knowing the searcher receptiveness to advertising would allow search engines to target ads better; and for advertisers to better target the appropriate population of "receptive" searchers, potentially paying a premium to have ads shown to receptive users. Indeed, previous studies have shown that displaying ads inappropriately is likely to annoy users [7]. So, if we could infer a user's current interests based on her search context and behavior, a search engine may then show more or fewer ads (or none at all) if the user is not "in the mood" for advertising - thereby conserving advertising budget and

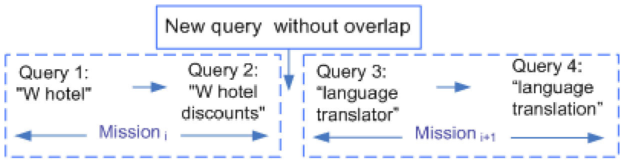

Figure 1. An example search session, consisting of two missions.

increasing searcher satisfaction.

In summary, our contributions include:

- A model of searcher receptiveness to search advertising, that naturally maintains session-level *state* for behavior modeling. We show that maintaining state across sessions substantially improves performance over stateless classification (Section III).
- Investigation of client-side interaction features for this task, including mouse movement-based features. We show that these new features can provide substantial improvement over previously explored query- and search result-based features (IV).
- A working implementation of our prototype system, **C**onte**x**tualized **I**nteraction (**CxI**) Model, and a thorough experimental evaluation of **CxI** over a dataset of hundreds of real users and thousands of search sessions, demonstrating substantial improvements in predicting ad clickthrough (Section VI).

Before we introduce our solution in detail, we briefly describe and illustrate the search advertising setting.

II. Problem Statement

We first define precisely what we mean by a *search session* and *search mission*, which we will use as a primary unit of organizing searches. Then, we formally state our problem of predicting advertising receptiveness.

Definition A *search session* is a sequence of searches, such that there is no time gap of greater than 30 minutes between any two consecutive searches.

Definition A *search mission* is a sequence of searches within a *search session*, such that a query of each search shares at least one non-stopword query term with *at least one* previous (but not necessarily consecutive) query.

An example search session consisting of two consecutive missions is reported in Figure 1.

Now that we have defined a search mission, we are ready to state the problem we address in this paper:

Predicting Advertising Receptiveness: *Given* the first i searches in a search mission $S(s_1, ..., s_i, ..., s_m)$, and the searcher behavior on these first i search result pages, *predict* whether the searcher will click on an ad presented on the results for any of the future searches $s_{i+1}, s_{i+2}, ..., s_m$, within the current search mission S.

Note that future ad clickthrough here is used as a *rough proxy* of searcher receptiveness to advertising. Also, while addressing this problem, we would like to put more emphasis on more accurate behavioral targeting: that is, we attempt to obtain higher precision in inferring searcher receptiveness (predicting future ad clickthrough), knowing that the ad clickthrough of current search engines is relatively low (e.g., Google CTR in our dataset is about 1%).

III. Modeling Searcher Context and Behavior

In this section we describe two approaches to modeling session-level behavior: a stateless/page-level approach, where ad clickthough is predicted independently of the session context (Section III-A), and our proposed contextual session-level model (Section III-B).

A. Page-Level Model

We start with a simple approach of intent inference as supervised classification performed independently for each search, where the goal is to predict the ad receptiveness for future searches. The assumption for this simple model is that a user's receptiveness to advertising is either present or absent during a search mission. Thus, we can predict for each search (and behavior on the search result page) *independently*, and then combine predictions to classify the rest of the mission as likely receptive or not. We will relax this assumption in our more general contextualized model in the next section.

We represent each search result page and associated interactions as feature vectors and predict over these feature vectors. The predictions over the initial searches are combined to predict advertising receptiveness in the rest of the search mission. A variety of classifiers could be used for this task, as we describe in Section IV. There are many ways to combine the predictions, such as majority voting, maximum or minimum prediction, average of the predictions. We experimented with a variety of these, and settled on using the *average* of the predictions as the combination function, with an empirically tuned threshold.

B. Session-Level Contextual Model

We now present our session-level contextual model that predicts future ad clickthrough based on the context and behavior of the observed sequence of searches during a mission. We represent ad receptiveness as two or more hidden states (e.g., "R" for receptive and "N" for non-receptive) and predict future ad clickthrough based on the most likely sequence for the portion of the mission observed so far. That is, given a sequence of related searches, we infer the most likely sequence of hidden states and predict *future* ad clickthrough based on the last inferred state of the mission. We illustrate this model in Figure 2.

Note that the predictions in this richer model are dependent on the *previous* inferred state of the user, thus naturally allowing us to maintain the user's "mental state" across individual searches. Thus, the receptiveness inference problem is transformed into the task of assigning label sequences to a set of observation sequences. Furthermore, this formalism allows for arbitrary number of hidden states which may correspond to different types of ad receptiveness, as we explore empirically in Section VI.

IV. THE CXI SYSTEM

We now introduce our system for **C**onte**x**tualized **I**nteraction (**CxI**) modeling. First we describe the client-side instrumentation for collecting the data. Then we describe the specific context and interaction features used in the current implementation.

A. Data Collection

We modified the Firefox version of the OpenSource LibX toolbar [1] to instrument mouse movements and other user action events on search result pages. The events are encoded in a string and occasionally sent to the server as HTTP request for later analysis or playback. The instrumented Web browsers were installed on 250 public-use computers (mostly, Windows PCs) at the Learning Commons at the Emory University Libraries. The usage was tracked only for users who have explicitly opted in to participate in our study. No identifiable user information was stored.

In our prototype implementation we sample mouse movements and scroll events at every 5 pixels moved, or every 50 ms, whichever is more frequent, and all other events (mouse down, keystrokes, etc.) are preserved at the original rate.

B. Search Context and Interaction Features

We now describe our information sources captured and the corresponding feature representations.

Our current implementation focuses on the *Context Features* (including search queries and other server-side information) and *Interaction Features*. In particular, we capture the mouse events (moves, clicks, down/up), scroll events, and button and menu events, in addition to the standard information such as page dwell time, and clickthrough on results.

1) Context Features: These features have been explored in previous work, and consist of word unigram features for the text of search query, SERP URL, clicked URL and SERP content. We also use the number of northern (top) ads and the number of eastern (side) ads as two additional features.

2) Interaction Features: This group consists of time-before-first-move, dwell time, number of different client-side events, the ways of entering the page (either via issuing a new query or clicking on the "back" button), the previous page type, and most importantly: a variety of representations of the mouse move trajectories. Again, our hypothesis is that mouse movements, like eye movements, can reflect short-term user interest and attention. Specifically, searchers may behave differently, varying the mouse's speed and range, as their interest and satisfaction level changes.

- *Global features*: First, we consider a coarse features such as the length, vertical and horizontal ranges of trajectories. For example, a user interested in ads might have wider horizontal range of mouse movements to hover over the ads.
- *Trajectory Features*: To capture additional information hidden in the mouse movements, we also model more precise physiological characteristics, following the work of Phillips and Triggs [19]. In particular, we capture properties such as *speed*, *acceleration*, *rotation* and other precise characteristics of the mouse movements. Specifically, to distinguish the patterns in different stages of the user interactions with the search results, we split each mouse trajectory into five *segments*: initial, early, middle, late, and end. Each of the five segments contains 20% of the sample points of the trajectories. Then, for each segment of the mouse trajectory we compute the average speed, average acceleration, slope and the rotation angle between the current segment and the segment connecting the beginning and the end (the click position) of the trajectories.
- *Hovering Features*: Since our focus is on search advertising prediction, we include two specialized features focusing on mouse hovering over the commonly displayed "north" and "east" ad regions on SERPs. These are boolean features that are triggered if the user hovers over the ad regions for more than 500ms.
- *Ad Clickthrough Features*: We also include the ad clickthrough on the current search page as binary feature. Recall, that our task is predicting *future* ad clicks, so ad clickthrough on current search result page is legitimate.

C. Classifier Implementation

We now describe the details of classifier implementation for both page-level and session-level contextual models.
Page-level classification: For the page-level model we tried various standard supervised machine learning classification techniques. In particular, we ended up with the Weka[2] implementation of Support Vector Machines.
Session-level classification: To implement this formalism, we use Mallet[3], an implementation of the Conditional Random Fields graphical model. Specifically, we define a conditional probability over hidden state sequences given a particular observation sequence of searches. We illustrate the configuration in Figure 2. It has two hidden states, *R*, and *N*, corresponding to "Receptive/Expect Ad Click in Some Future Search in this Mission", and "Not Receptive: Do

[1] Available at www.libx.org

[2] Available at http://www.cs.waikato.ac.nz/ml/weka/.

[3] http://mallet.cs.umass.edu/

Not Expect Ad Click in Future Searches". At training, the hidden state is assigned according to whether an ad click was observed in the future searches within the mission. Note that an ad click on *current* page is simply an observation, and is not necessarily an indication whether a user remains likely to click on future search ad in the mission. During prediction, given a sequence of observations (context and interactions), our system attempts to recover the label sequence that maximizes the conditional probability of the observations.

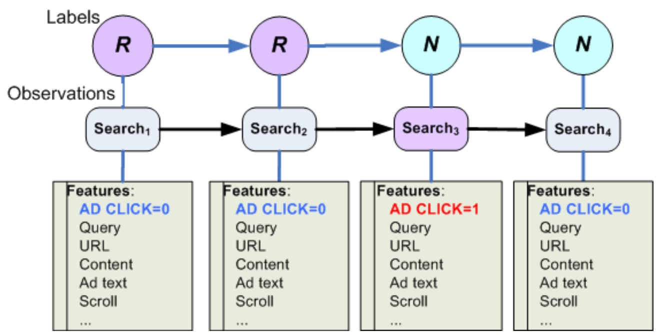

Figure 2. CRF Implementation of Session-Level Model with two hidden states, R and N, with labels assigned according to the observed ad clickthrough on the third search result pages within the mission.

V. Experimental Setup

This section describes our experimental setup. First, we describe the dataset, collected over a period of four months; then we define our evaluation metrics. Finally, we summarize the methods compared, including our implementation of the previous state-of-the-art prediction models.

A. Dataset

The data was gathered from mid-August through mid-December 2008 from public-use machines at the libraries of a major university. To ensure data consistency, we generated a longitudinal dataset for the 440 users who have clicked on a search ad using our instrumented machines at least once during the data collection period. For this universe of users we include all the search sessions attempted during this period. The resulting dataset contains 4,377 login sessions, comprising 6,476 search sessions, 17,123 search missions and 45,212 searches.

B. Metrics

To focus specifically on the ad click prediction task (positive class), we will report Precision (P), Recall (R), and F1-measure (F1).

- **Precision (P)**: Precision of prediction is computed with respect to the positive (receptive) class, as fraction of true positives over all predicted positives. Specifically, for each mission, the precision is computed as the fraction of correct positive predictions over all positive predictions in the mission. We then average this value over all the missions with positive predictions.
- **Recall (R)**: Recall is a measure of the coverage of the correct predictions: the faction of true positives correctly identified. For each such mission, the recall is computed as the fraction of correct positive predictions over all positive labels in the mission. This value is then averaged over all the missions with positive labels.
- **F1-measure (F1)**: F1 represents the balance between P and R and is computed as $\frac{2P \cdot R}{P+R}$.

C. Methods Compared

We now describe the methods used.

- **QC**: Query Chains, similar to the state-of-the-art query chains models (e.g., [21], [20]), implemented using features extracted from the query strings, SERP URLs, and using the CRF model.
- **QCLK**: Variant of Query Chains, incorporating additional features extracted from clicked URLs.
- **PC**: Page-level Context, similar to the state-of-the-art page-level context models (e.g., [13]), implemented using page-level context features to independently infer ad receptiveness for each search, and then aggregated over the mission prefix predictions using the *average* heuristic with best-performance threshold of 0.5 (that is, at least half of the observed searches predicted as likely receptive).
- **CxI (C)**: CRF model, using all context features.
- **CxI (C+I)**: CRF model, using all context and interaction features.

VI. Results

All results are reported for 5-fold Cross Validation, with 80% of the sessions used for training, and remaining 20% used for test. Note that data was split by sessions so there are no missions in both training and test data. The results are reported in Table I. As we can see, all the variants of our system would potentially result in more accurate user behavior targeting (compared to 1% Google Ad CTR in our dataset), for a portion of the searches.[4]

Specifically, the **CxI** system substantially outperforms the **QC**, **QCLK** and **PC** systems on ad clickthrough prediction. Along the dimension of features, SERP content and interactions improve performance over the query and URL features substantially. Specifically, adding SERP content improves precision and F1, and interactions on top of SERP content gains further improvements. Notice that, we also treated organic result text separately from ad text

[4]Google CTR is not strictly comparable to Precision as we define it for our task. To estimate comparable Precision value for Google, we could use Google's rendering of ads on the following SERP as their "prediction" of advertising receptiveness, and estimate the Precision as we do for other methods. Using this metric, the Precision of Google prediction is approximately 5%, and in fact is somewhat lower as we have no way to evaluate the accuracy of "prediction" for the last search in each mission. Regardless, all of the methods except for the QC baseline exhibit substantially higher Precision than Google.

in our development experiments. We found that using the whole text of the SERP, results in the best performance, so we use the whole text for all subsequent experiments.

Method	*Precision*	*Recall*	*F1*
QC	0.045 (-)	0.358 (-)	0.079 (-)
QCLK	0.075 (+67%)	0.150 (-58%)	0.100 (+27%)
PC	0.081 (+80%)	0.117 (-67%)	0.096 (+22%)
CxI (C)	0.199 (+342%)	0.100 (-72%)	0.133 (+68%)
CxI (C+I)	**0.207 (+360%)**	**0.124 (-65%)**	**0.155 (+96%)**

Table I
PRECISION, RECALL, AND F1 FOR PREDICTING AD CLICKTHROUGH ON FUTURE SEARCHES WITHIN A SEARCH MISSION

As an attempt to increase Recall, we copy the positive training samples five times to reduce the effects of the skewed class distribution (few positive class instances) and assign a higher weight to the positive class - note that the composition of the test set remains unchanged. The results are reported in Table II. As we can see, copying positive training samples indeed could improve Recall, although with some sacrifice in Precision. Also, **CxI** still performs substantially better than the Query Chains baseline.

Method	*Precision*	*Recall*	*F1*
QCLK	0.040 (-)	0.278 (-)	0.069 (-)
CxI (C)	0.167 (+318%)	0.111 (-60%)	0.133 (+93%)
CxI (C+I)	**0.193 (+383%)**	**0.130 (-53%)**	**0.155 (+125%)**

Table II
PRECISION, RECALL, AND F1 FOR PREDICTING AD CLICKTHROUGH ON FUTURE SEARCHES WITHIN A SEARCH MISSION (COPYING POSITIVE TRAINING SAMPLES)

In summary, our results have shown substantial improvement on ad clickthrough for precision and F1 of our system compared to state-of-the-art baselines that do not consider session-level context and user interactions. We believe that our model, features, and problem formulations are novel and practical, as we discuss when putting our contribution in the context of the previous work.

VII. RELATED WORK

The origins of user modeling research can be traced to library and information science research of the 1980s. An excellent overview of the traditional "pre-Web" user modeling research is available in [4]. With the explosion of the popularity of the Web, and with increasing availability of large amounts of user data, the area of modeling users, user intent, and in general Web usage mining has become an active area of research in the information retrieval and data mining communities. In particular, inferring user intent in Web search has been studied extensively, including references such as [24], [18], [2]. Taxonomies of Web search and user goals have been relatively stable since Broder's classic paper classifying intent into navigational, transactional and informational [6]. Recently, topical commercial query classification was presented in [24].

Previous research on user behavior modeling for Web search focused on aggregated behavior or other general aspects of behavior [14]. Another approach is to model random walk on the click graph [11], which considers a series of interactions within a search page, but not session-level context or behavior information.

However, it has been shown that user goals and experience vary widely and have substantial effect on user behavior [25]. Furthermore, some queries have substantial variation in intent and some do not, hence behavior can help distinguish user intent in such ambiguous cases, as we attempt to do in this paper. Recently, eye tracking has started to emerge as a useful technology for understanding some of the mechanisms behind user behavior (e.g., [17], [12]). Reference [16] attempts to identify some of the same attention indicators by using mouse movements, which has been shown to correlate to gaze position [23].

Most of the previous research on predicting ad clickthrough rate focuses on learning from the content of displayed ads (e.g., [10], [22]), but does not take the search context and client-side, individual user behavior into account. Reference [3] considered the result (and ad) relative position and presentation features to improve clickthrough estimation, within a single page. Reference [13] addressed the detection of commercial intent in the aggregate using page-level context modeling. In the context of search advertising, empirical studies of search effectiveness are reported in [15]. Another dimension of work somewhat similar in approach to ours considers query chains and browsing behavior to infer document relevance (e.g., [21], [1], [5], [20]), and for query suggestion [9] and contextual query classification [8].

However, in contrast to previous work, we address the complementary aspect of the ad clickthrough prediction problem - that is, inferring whether the searcher is in an advertising receptive state or not. Also, we explore combining search context and fine-grained user interactions, such as mouse movements, on the session-level.

VIII. CONCLUSIONS

We introduced a new model for predicting ad clickthrough on search ads that maintains session-level state for the user across multiple searches in the session. In particular, we hypothesize the existence of a "hidden state" which corresponds to advertising receptiveness of a user, which may increase or decrease the likelihood for ad clickthrough within a search session and is orthogonal to traditional ad quality and relevance metrics. To capture this effect, we introduce our contextualized interaction modeling system, **CxI**, that uses the conditional random fields as an underlying machine learning method. Our system naturally allows incorporation of sophisticated and fine-grained user interaction features such as mouse movements, scroll events, and other novel client-side interactions. We show that **CxI** substantially outperforms previously introduced methods that only capture the context of query chains (**QC/QCLK**), or model the page-level context independently of the session (**PC**).

While our evaluation is performed over a (relatively) large dataset collected over thousands of real search sessions for hundreds of users, we acknowledge some limitations of our study. Specifically, our user population is relatively homogeneous (college and graduate students, and faculty and staff), and substantially more training data may be required to achieve this performance for the general population. Another limitation is lack of conversion data: ad clickthrough is just one evaluation metric, and may not be predictive of the ultimate intent of the searcher (e.g., a searcher may click on an ad out of curiosity). In the current implementation we are not yet modeling personal user history, which is the topic of our future work.

In summary, we have shown that by modeling interactions in context (of the search session or the search result page) we can substantially improve ad clickthrough prediction accuracy. 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.

References

[1] E. Agichtein, E. Brill, and S. Dumais. Improving web search ranking by incorporating user behavior information. In *Proc. of SIGIR*, 2006.

[2] E. Agichtein, E. Brill, S. Dumais, and R. Ragno. Learning user interaction models for predicting web search result preferences. In *Proc. of SIGIR*, 2006.

[3] H. Becker, C. Meek, and D. Chickering. Modeling Contextual Factors of Click Rates. In *Proc. of AAAI*, 2007.

[4] N. J. Belkin. User modeling in information retrieval. *Tutorial at UM97*, 1997.

[5] M. Bilenko and R. White. Mining the search trails of the surfing crowds: Identifying relevant websites from user activity. In *Proc. WWW*, 2008.

[6] A. Broder. A taxonomy of web search. *SIGIR Forum*, 2002.

[7] A. Z. Broder, M. Ciaramita, M. Fontoura, E. Gabrilovich, V. Josifovski, D. Metzler, V. Murdock, and V. Plachouras. To swing or not to swing: Learning when (not) to advertise. In *Proc. of CIKM*, 2008.

[8] H. Cao, D. Jiang, J. Pei, E. Chen, and H. Li. Towards context-aware search by learning a very large variable length hidden markov model from search logs. In *WWW '09: Proceedings of the 18th international conference on World wide web*, pages 191–200, 2009.

[9] H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li. Context-aware query suggestion by mining click-through and session data. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*, 2008.

[10] M. Ciaramita, V. Murdock, and V. Plachouras. Online learning from click data for sponsored search. In *Proc. of WWW*, 2008.

[11] N. Craswell and M. Szummer. Random walks on the click graph. In *Proc. SIGIR*, 2007.

[12] E. Cutrell and Z. Guan. What are you looking for?: an eye-tracking study of information usage in web search. In *Proc. of CHI*, 2007.

[13] 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.

[14] D. Downey, S. T. Dumais, and E. Horvitz. Models of searching and browsing: Languages, studies, and application. In *Proc. of IJCAI*, 2007.

[15] A. Ghose and S. Yang. An empirical analysis of sponsored search performance in search engine advertising. In *Proc. of WSDM*, 2008.

[16] Q. Guo and E. Agichtein. Exploring client-side instrumentation for personalized search intent inference. In *Proc. of ITWP*, 2008.

[17] T. Joachims, L. Granka, B. Pan, H. Hembrooke, F. Radlinski, and G. Gay. Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. *ACM Trans. Inf. Syst.*, 25(2), 2007.

[18] U. Lee, Z. Liu, and J. Cho. Automatic identification of user goals in web search. In *Proc. of WWW*, 2005.

[19] J. G. Phillips and T. J. Triggs. Characteristics of cursor trajectories controlled by the computer mouse. *Ergonomics*, 2001.

[20] 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.

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

[22] M. Richardson, E. Dominowska, and R. Ragno. Predicting clicks: estimating the click-through rate for new ads. In *Proc. of WWW*, 2007.

[23] K. Rodden, X. Fu, A. Aula, and I. Spiro. Eye-mouse coordination patterns on web search results pages. In *Proc. of CHI*, 2008.

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

[25] R. W. White and S. M. Drucker. Investigating behavioral variability in web search. In *Proc. of WWW*, 2007.