

Beyond Session Segmentation: Predicting Changes in Search Intent With Client-Side User Interactions

Qi Guo
Emory University
qguo3@mathcs.emory.edu

Eugene Agichtein
Emory University
eugene@mathcs.emory.edu

ABSTRACT

Effective search session segmentation—grouping queries according to common task or intent—can be useful for improving relevance, search evaluation, and query suggestion. Previous work has largely attempted to segment search sessions off-line, after the fact. In contrast, we present preliminary investigation of *predicting*, in real time, whether a user is about to switch interest—that is, whether the user is about to finish the current search and switch to another search task (or stop searching altogether). We explore an approach for this task using client-side user behavior such as clicks, scrolls, and mouse movements, contextualized by the content of the search result pages and previous searches. Our experiments over thousands of real searches show that we can identify context and user behavior patterns that indicate that a user is about to switch to a new search task. These preliminary results can be helpful for more effective query suggestion and personalization.

Categories and Subject Descriptors

H.3.3 [Information Systems]: Relevance feedback, search process

General Terms

Design, Experimentation, Evaluation

Keywords

user intent inference; search behavior modeling; search sessions.

1. INTRODUCTION AND MOTIVATION

Detecting Web search intent is challenging, as Web search queries are often ambiguous and may reflect diverse intents and information needs. Recent approaches to search intent detection and query suggestions have started to model user intent on session level. However, search session segmentation—grouping queries according to common task or intent—is challenging, and so far has been primarily attempted off-line, after the fact (e.g., in recent work by Jones et al. [3]). In contrast, a more useful task would be to identify user interest switches (search session segments) as they occur.

Before we present our methods and results, we first define what precisely we mean by a search session, and a search mission. A *search* is an interaction that leads to a result page; a *query* is a set of terms given by a search. Then a search *mission* is a sequence of consecutive searches, such that a query of a search shares at least one non-stopword with any previous query within the search mission. A mission is terminated when the query of a new search does not share any words with the previous ones. Two example search missions executed within a single session are reported in Figure 1.

Note that requiring overlap between any previous queries within a mission (not only between the first query or the adjacent previous query), supports search intent drift, naturally maintaining a single

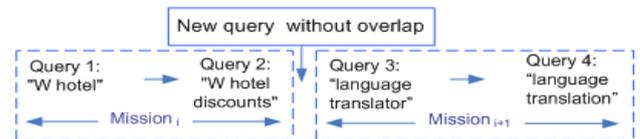


Figure 1: An example user session, consisting of two consecutive search missions.

search mission even as the users' information need evolves. We do not attempt to handle interleaved missions. To validate our mission partitioning heuristic we manually examined 100 randomly chosen sessions and agreed with the mission-based splits in all but a handful of cases (such as misspellings). Therefore, we use our definition of mission switching as a proxy for intent switching: that is, whenever a user switches to a new search mission, we consider that the next search query has a new intent.

Problem statement: Our task is, given a sequence of searches in a mission, to predict whether the user will continue on the same mission, or switch to a new one (or end searching completely).

We now introduce our techniques and the dataset used for evaluation, followed by preliminary experimental results and discussion.

2. METHODOLOGY

We represent a searcher as a non-deterministic state machine where the emissions are the observable actions, and the user intent and other private states as the *hidden states* in the model. One simple model is illustrated in Figure 2. As shown in the illustration, searcher actions such as query reformulation/query for new mission, clicks, mouse movements and button presses are observations generated by the hidden states. For example, query reformulations are more likely if the user is in a state willing to continue with the current search, and less likely if the user is about to switch to another task. Similarly, the emission probabilities of other observations would vary over different hidden states. Hence, observations including the search context and user interactions are related to the inner states of the users. If we can infer the hidden states using the observations we can in turn predict future user actions such as task switching.

To implement this formalism, we use the Conditional Random Fields graphical model. Specifically, we define a conditional probability over hidden state sequences given a particular observation sequence of searches. At training, the hidden state is assigned according to whether there are future searches within the mission. In particular, we experimented with two variants: 2-state and 3-state models. The 2-state model consists of the {*Continue*, *Switch*} states while the 3-state model contains the {*Continue*, *Switch*, *Exit*} states. At test time we predict the state sequence that maximizes the conditional probability of the sequence of observed actions. Specifically, we use the Mallet¹ library implementation of the Linear Chain CRF.

¹Available at <http://mallet.cs.umass.edu/>.

Method	Continue			Switch			Macro-averaged F1
	Precision	Recall	F1	Precision	Recall	F1	
Query URL	0.323 (-)	0.997 (-)	0.488 (-)	0.912 (-)	0.014 (-)	0.027 (-)	0.251 (-)
Content	0.363 (+12 %)	0.926 (-7 %)	0.502 (+3%)	0.760 (-17%)	0.159 (+1036 %)	0.259 (+859 %)	0.381 (+52 %)
Content+Interactions	0.382 (+18 %)	0.913 (-8 %)	0.539 (+10 %)	0.773 (-15 %)	0.207 (+1379 %)	0.327 (+1111 %)	0.486 (+73%)

Table 1: Precision, Recall, and F1 on predicting *Continue*, *Switch* for Content+Interactions Methods (2 states).

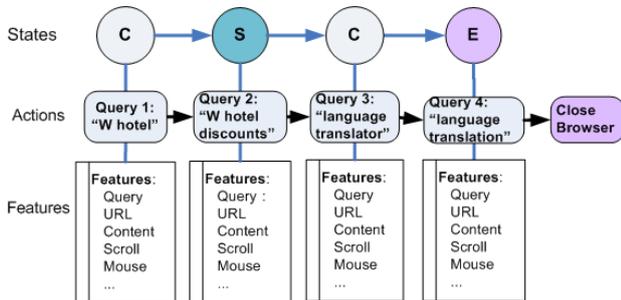


Figure 2: States (C-Continue; S-Switch; E-Exit), actions, and features model for an example user session consisting of two search missions.

We now describe our representations of search context and user interactions. All the features are computed over the *previous* searches so that the values are available before each new search to be predicted within the same mission. Due to space limitations we omit a complete description of all the features.

- **Query and clickthrough features:** we follow the previous work on session segmentation and use the query and clickthrough features, specifically, the word unigrams for the text of the query, and the URL of the clicked result. This feature sets corresponds to our baseline approach.
- **Content features:** We represent the content of the search result page (and different portions of the page such as organic results and the search ads) using word unigrams.
- **Client-side Interaction features:** As pointed out by Atterer et al. in reference [1], the detailed-tracking of user interaction can be used for usability evaluation and be used as more fine-grained implicit feedback. Our client-side features include time from page load until the first move or UI event, dwell time and mouse trajectory length, vertical and horizontal range. We also capture more precise physiological characteristics, such as *speed*, *acceleration*, *rotation* of the mouse movements, following the work of [4] and using the features similar to our previous work described in [2].

3. EXPERIMENTAL RESULTS

To collect the data, we developed a simple Firefox plugin to capture the mouse movements and other user action events on search result pages. The instrumented Web browsers were installed on 250 public-use computers in Emory University libraries. The usage was tracked only for users who have explicitly opted in to participate in our study and no identifiable user information was stored.

The data was gathered from mid-August through mid-December 2008. To ensure data consistency, we generated a longitudinal dataset of the usage for 440 users 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, 16,693 search missions and 45,212 searches.

We explore the system performance for the 2-state and 3-state settings. All results are reported for 5-fold Cross Validation, with

80% of the missions used for training, and remaining 20% used for test. Note that data was split by login session, so there are no login sessions in both training and test data.

For the 2-state setting, we experimented with different feature combinations, starting from simple query and URL features, adding page content and interactions gradually. The results are reported in Table 1. As we can see, for predicting the majority class *Continue*, the Recall values are all very high while Precision values relatively low. Incorporating content and interactions results in improvements on Precision and F1. As for the more interesting prediction for the *Switch* class, query URL baseline only covers a very small number of samples although with high Precision. Incorporating page content results in an order of magnitude increase in Recall without too much sacrifice in Precision. Finally, adding interactions on top of page content results in significant improvements on both Precision and Recall.

We now explore the variant of the 3-state model, where *Exit* is added as the third state. The results for predicting all 3 classes are reported in Table 2. Not surprisingly, the *Continue* class exhibits the highest Recall but lower Precision. The Precision of *Switch* decreases compared to that of 2-state. This probably due to the inherent similar behavioral patterns of *Switch* and *Exit*.

The results for predicting all 3 classes are reported in Table 2. Not surprisingly, the *Continue* class exhibits the highest Recall but lower Precision. The Precision of predicting the *Switch* class decreases compared to that of the 2-state configuration. This could be related to similar behavior prior to the *Switch* and *Exit* actions.

Predicted Class	Precision	Recall	F1
Continue	0.402	0.880	0.552
Switch	0.579	0.184	0.279
Exit	0.348	0.155	0.215

Table 2: Precision, Recall, and F1 on predicting *Continue*, *Switch* and *Exit* for Content+Interactions Methods (3 states)

In summary, we have introduced a new approach to predict a switch in the user search interest by modeling session-level information about the users' search context and individual behavior. Our results are encouraging and could serve as a first step towards personalized query suggestion, better behavior targeting, and more effective search evaluation.

ACKNOWLEDGMENTS: This research is partially supported by the Yahoo! Research faculty grant.

4. REFERENCES

- [1] R. Atterer, M. Wnuk, and A. Schmidt. Knowing the user's every move: user activity tracking for website usability evaluation and implicit interaction. In *Proc. of WWW*, 2006.
- [2] Q. Guo and E. Agichtein. Exploring client-side instrumentation for personalized search intent inference. In *Proc. of ITWP*, 2008.
- [3] R. Jones and K. Klinkner. Beyond the session timeout: Automatic hierarchical segmentation of search topics in query logs. In *Proc. CIKM 2008*, 2008.
- [4] J. G. Phillips and T. J. Triggs. Characteristics of cursor trajectories controlled by the computer mouse. *Ergonomics*, 2001.