Talking the Talk vs. Walking the Walk: Salience of Information Needs in Querying vs. Browsing

Mikhail Bilenko, Ryen W. White, Matthew Richardson
Microsoft Research
One Microsoft Way
Redmond, WA 98052, USA
{mbilenko,ryenw,mattri}@microsoft.com

G. Craig Murray
College of Information Studies
University of Maryland
College Park, MD 20742, USA
gcraigm@umd.edu

ABSTRACT

Traditional information retrieval models assume that users express their information needs via text queries (*i.e.*, their "talk"). In this poster, we consider Web browsing behavior outside of interactions with retrieval systems (*i.e.*, users' "walk") as an alternative source of signal describing users' information needs, and compare it to the query-expressed information needs on a large dataset. Our findings demonstrate that information needs expressed in different behavior modalities are largely non-overlapping, and that past behavior in each modality is the most accurate predictor of future behavior in that modality. Results also show that browsing data provides a stronger source of signal than search queries due to its greater volume, which explains previous work that has found implicit behavioral data to be a valuable source of information for user modeling and personalization.

Keywords

Predictive user modeling, user profiles, personalization.

1. INTRODUCTION

Modeling user interests is a central problem for personalization and information filtering applications that attempt to meet distinct needs of individual users, which include recommender systems, and personalized search and advertising. Such applications typically rely on processing past user behavior to obtain a refined representation of user interests suitable for predicting future behavior. Most often, past user behavior is assumed to be of the same modality (type) as future behavior, *e.g.*, consumption data and ratings of domain-specific items such as movies for recommenders [7], or past search-related behavior for personalized search [3].

It is well understood in the Information Retrieval (IR) community that interaction with search systems does not completely capture all aspects of users' underlying information needs [5]. Relevance feedback (RF) [4] has been proposed as a methodology for improving retrieval performance by capturing users' implicit or explicit relevance judgments through interactions that follow a query. Although RF is typically applied within the local context of a single search session, persistent user profiles gathered from implicit data have shown promise in a number of information access environments [2].

Behavior other than interactions with retrieval systems has been previously employed for improving personalized search [6], including desktop and Web browsing data. Prior study of the *relative* value of different behavior modalities was limited to evaluating their utility for personalizing search. In this work, we consider the general problem of predicting users' future interests based on multiple modalities. We specifically focus on users' querying behavior and their browsing behavior, and evaluate their relative informativeness for predicting users' future information needs. User interest profiles can be constructed based on a history of their querying, browsing, or a combination of the two. This poster presents a study of the relative predictiveness of user profiles created using these two modes of information need expression, with the aim of determining whether querying, browsing, or their combination produce more accurate user profiles. We evaluate the performance of profiles derived from these sources based on how well they predict future behavior of a large sample of Web users.

SIGIR'08, July 20–24, 2008, Singapore.
ACM 978-1-60558-164-4/08/07.

2. MODELING INFORMATION NEEDS

2.1 Representing Information Needs

A key issue in comparing the information content across multiple behavior sources is establishing a single representation of information need that is equally capable of capturing signal from each source. In this work, we employ a representation based on *advertising keywords*, which are short phrases ("keywords") bid on by a large number of advertisers on Microsoft AdCenter, a large advertising platform. The resulting representation is very rich: it is based on several million keywords, and yields either an exact or a subsequence match for the majority of queries submitted to major search engines. This representation is arguably more semantically meaningful than simple token-based vocabularies and more naturalistic than automatically or manually constructed query categorization hierarchies.

2.2 Behavior-based Information Needs

Unlike query behavior, browsing behavior must be mapped to a keyword-based representation indirectly. To do this, we use search engine click logs. When a user issues a query to a search engine, and then clicks on a result, they are implicitly associating their query with that page. Through logs of search queries and subsequent result clicks, individual pages and domains (websites) can be represented by sets of keywords. Such a representation has been previously considered in an IR context from a formal graph-theoretic perspective [1]. In this work, we build keyword-based user profiles from browsing behavior by including the top-frequency keywords for each domain the user visited (*i.e.*, the most frequent searches leading to the domain).

3. EXPERIMENTS

3.1 Methodology and Datasets

Since the core task for any user modeling system is predicting future behavior, we evaluate the informativeness of different sources of behavioral signal based on their predictive value. In other words, we aggregate the past behavior in the two modalities considered (search queries and browsing behavior) over a given time period, and evaluate the predictiveness of the resulting aggregated user profile with respect to behavior occurring in a sub-

sequent period. We use standard IR measures, precision and recall, to evaluate the profiles. Precision is the fraction of all predicted keywords that are observed in the future data, and recall is the fraction of all future keywords predicted by the past-based profiles. We report values micro-averaged by user.

We conduct our study on a dataset that includes Web querying and browsing behavior of over 1.2 million consenting users in November 2007 obtained via a browser toolbar plugin. A complete behavior log of every user is split into "past" and "future" halves on November 15. Separate profiles are then created for both past and future by mapping the querying and browsing behavior to advertising keywords as described in Section 2. On average, each query-based user profile contains 21.2 keywords, while each browsing-based profile contains 137.4 keywords (based on 15 days of behavioral data).

3.2 Results and Discussion

First, we compare the average precision and recall for profiles based on the two different modalities (queries and browsing) as well as their combination for the task of predicting future behavior either in the same or different modality. Table 1 below presents these results. Given the large sample sizes, all observed differences in the means were statistically significant ($p \leq 0.0001$ with two-tailed t-tests based on 10-fold subsamples). Note that the relatively low precision and recall values are explained by the fact that the metrics only account for exact matches: for example, if a keyword *canon* was predicted for a user, and the query [*best canon digital slr*] was observed in future behavior which maps to keywords *canon*, *slr*, *canon slr, digital slr*, and *canon digital slr*, a recall of 0.25 would be recorded. If the example was reversed temporally, precision would suffer correspondingly.

Table 1. Prediction precision/recall for different modalities.

Based on \ Predicted for	Querying	Browsing	Querying+Browsing
Querying	0.128/0.114	0.188/0.018	0.223/0.020
Browsing	0.019/0.165	0.325/0.295	0.324/0.289
Querying+Browsing	0.022/0.195	0.317/0.296	0.320/0.291

The results in Table 1 demonstrate that information intents expressed in users' querying and browsing behavior are largely independent: the highest precision is obtained for any modality based on previous behavior in the same modality, while predictions based on a different modality are significantly less accurate. At the same time, results show that the signal contained in browsing behavior is more salient across modalities than that contained in querying behavior: recall of future querying behavior is higher when based on past browsing than on past querying. Combining the two behavior modalities leads to modest recall improvements over using either modality in isolation; recall for predicting either querying or browsing behavior based on their combination is slightly higher than that based on either behavior in isolation. This demonstrates that the information intents expressed through those different behaviors are primarily orthogonal.

Our results raise the question of whether browsing behavior may appear more salient only because it is more plentiful, and hence leads to the creation of larger profiles. We conducted further experiments where every user profile constructed based on past data was constrained to a fixed number of keywords, with keywords selected via ranking based on combining frequency and recency of keyword occurrences. Figure 1 illustrates recall obtained from these experiments.

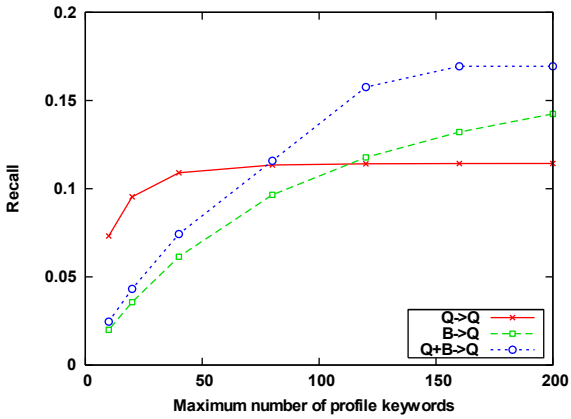

Figure 1. Querying recall for different profile sizes.

These results demonstrate that a larger volume of browsing data is required for predicting future querying behavior more effectively than prediction based on querying behavior alone (as evidenced by the $B \rightarrow Q$ curve being dominated by the $Q \rightarrow Q$ curve up to 120 keywords; the curves eventually cross due to smaller volumes of querying data). At the same time, access to both behavior modalities provides sustained improvements for all profile sizes, again demonstrating that information needs expressed in different modalities are largely independent for all amounts of available data. Finally, we note that the $B+Q \rightarrow Q$ curve is dominated by the $Q \rightarrow Q$ curve for smaller profiles because of the simplistic profile construction procedure we used. Machine learning methods would allow combining the two data sources for more accurate profiles than those obtained from each source alone.

4. CONCLUSIONS

We have described a large-scale study that demonstrates that behavior modalities differ significantly in their ability to predict users' future information needs. Furthermore, the study reveals that information intents expressed through different modalities are largely orthogonal. In particular, users' querying behavior (their "talk") is a more limited source of predictive signal than their browsing behavior (their "walk"). These findings have profound implications for user modeling and personalization applications, encouraging focus on approaches that can leverage users' browsing behavior as a source of information.

5. REFERENCES

[1] Craswell, N. and Szummer, M. (2007). Random walks on the click graph. *Proc. SIGIR-07*, 239-246.

[2] Gauch, S., Speretta, M., Chandramouli, A., and Micarelli, A. (2007). User profiles for personalized information access. In *The Adaptive Web,* 54-89, Springer.

[3] Qiu, F. and Cho, J. (2006). Automatic identification of user interest for personalized search. *Proc. WWW-06*, 727-736.

[4] Salton, G. and Buckley, C. (1990). Improving Retrieval Performance by Relevance Feedback. *J. of the American Society for Information Science*, 41(4), 288-297.

[5] Taylor, R.(1968). Question negotiation and information seeking in libraries. *College and Research Libraries*, 29,178-194.

[6] Teevan, J., Dumais, S. T., and Horvitz, E. (2005). Personalizing search via automated analysis of interests and activities. *Proc. SIGIR-05*, 449-456.

[7] Zhang, Y. and Koren, J. (2007). Efficient bayesian hierarchical user modeling for recommendation system. *Proc. SIGIR-07*, 47-54.