Behavior-based Recommender Systems for Web Content

Tingshao Zhu, Russ Greiner
Dept. of Computing Science
University of Alberta
Canada T6G 2E8

Gerald Häubl
School of Business
University of Alberta
Canada T6G 2R6

Bob Price, Kevin Jewell
Dept. of Computing Science
University of Alberta
Canada T6G 2E8

ABSTRACT

In this paper, we demonstrate the implementation of an effective complete-web recommender system (i.e., *WebICLite*) using browsing behavior models to predict relevant Web pages. Behavior-based models use fine-grained information about the actions a user takes while browsing the web and the exact sequence of pages they follow to proactively provide responsive session-specific site-independent recommendations. The current paper also briefly presents browsing behavior-based models, and summarizes initial results from a large-scale field trial. The study suggests that the positive laboratory results for the original model transfer to real users browsing arbitrary web pages for day-to-day tasks.

Keywords

Browsing Behavior Model, Machine Learning, Information Search, Web Content Recommendations

1 Introduction

While the World Wide Web contains a vast quantity of information, it is often time consuming and difficult for web users to find the information they are seeking on the Web. Typically users will employ a search engine to find information. In order to benefit from these search engines, however, users must have intuitions about what keywords they should use to effectively discriminate the information they are seeking from the information they don't want from among the billions of Web pages that search engines typically index.

A number of researchers have proposed web-recommender systems that attempt to learn a user's information needs from observations of their past web-browsing behaviors. These recommenders use advanced information retrieval techniques to locate web resources that satisfy the user's needs. In this way, the user receives the information they need without having to reason about the best query to retrieve the information.

Zukerman [16] distinguishes two main classes of web-recommenders: *Content-based* systems use samples of past user behavior to learn what types of content appeal to a user. The system then recommends pages with similar content. *Collaborative filtering* systems uses samples of past user behavior to learn how the current user is similar to other users. The system then recommends pages to the user that have also been selected by similar users. Both Content-based and collaborative filtering systems have been well studied in the literature and their strengths and weaknesses are well understood. Content-based systems need a large sample of past user selections to establish user interests whereas collaborative filtering systems only work well when there is a sufficient pool of similar users. Both types of systems suffer from objective measures of validity as we cannot know if the user's past choices or the choices made by other similar users were really satisfactory and both types of systems tends to be site specific due to their need for information about the user's past browsing behavior.

In previous work [14], we have pointed out two opportunities for extending current web-recommender systems. First, we observed that a user's needs can change dramatically as the user plays different roles in life and works on various tasks and subtasks. A sensible recommender systems should recognize the differences between current interests and long term interests and makes its recommendations based on the user's current needs. We think of the searches for each distinct information need as occurring in distinct "sessions" and we call this concept *session specific recommendation*.

Second, we observed that recommenders can also help users by bringing relevant material to their attention even though they may not have thought to ask for it. We call this *proactive recommendation*.

We then proposed that we can use passive observations of the user's fine-grained web-browsing actions and the specific sequences of web-pages they were applied to to learn more about user's interests than is possible with static analysis of a bag of web pages visited by the user and with less input from the user than systems that require the user to label specific pages with their judgments. Since our analysis is based on the user's current dynamic actions it can be made session

Workshop: Beyond Personalization 2005
IUI'05, January 9, 2005, San Diego, California, USA
http://www.cs.umn.edu/Research/GroupLens/beyond2005

specific. Since we can obtain this information unobtrusively, we can gather user's interests and make proactive recommendations. We say that we use the user's browsing behaviors to make proactive session-specific recommendations.

Our focus on the extraction of the user's information needs instead of the indexing of material has an additional benefit: namely, the approach is compatible with many existing methods for indexing material within a content-based framework. In particular, our system can turn inferences about user information needs into queries for standard search engines. This potentially allows us to recommend any web-resource indexed by major search engines to our users.

In this paper we report on a newly developed Web recommender system — *WebICLite*. Like most recommendation systems, *WebICLite* watches a user as s/he navigates through a sequence of pages, and suggests pages that (it predicts) will provide the relevant information. *WebICLite* differs from most other web recommendation system in several respects. First, while many recommendation systems are server-side and hence specific to a single web site [10, 1, 13], our client-side system is not so specific, and so can point users to pages *anywhere on the Web*. Secondly, *WebICLite* can predict the user's information need *dynamically*, based on the current context — that is, the current session. (This is based on patterns found over the "browsing properties" of the words appearing in the session; see Section 3.) The third difference deals with the goal of the recommendation system: our goal is to recommend only useful pages; i.e., pages that are relevant to the user's task. These "Information Content" pages (aka IC-pages) are just the pages the user needs to solve his/her current task, and not the overhead pages required to reach them. This differs from systems that instead attempt to lead the user to pages that similar users have visited earlier (independent of whether those "familiar" pages in fact contain the answer to the user's current quest). Finally, *WebICLite* is "passive", in that it can recommend pages relevant to the user's *current* information need without requiring the user to do any additional work — e.g., the user does not need to answer intrusive questions, etc.

Section 2 discusses related work. Section 3 then describe a simple procedure for training our models, and the results of a user study (i.e., LILAC) that demonstrates the ability of our model to predict pages useful to the current user, from anywhere on the Web. Section 4 describes an implementation of our ideas in the form of a stand-alone web browser, *WebICLite*, that runs on the user's computer, and provides on-line recommendations, to pages anywhere on the Web, not just on the user's current Web site. Finally, Section 5 concludes with a summary of our key contributions and insights.

2 Related Work

Many groups have built various types of systems that make recommendations to users. One can get a sense of the breadth of work in this area from the table below which summarizes a number of common approaches and representatives systems within these approaches:

- COB: Co-occurrence Based — e.g., Association Rule [2], Sequential Pattern [3], etc.
- CF: Collaborative Filtering [12]
- CB: Content-Based [5, 8, 4]
- HBM: Heuristic-Based Model [11, 9, 6]
- IC-Models: IC-based Models; see Section 3.

We find it useful to compare these systems on a number of parameters:

- All systems require a model of the user's interests, but some learn the model and some do not.
- Some systems require a training phase in which users distinguish content they desire from content they do not.
- Systems vary in the extent to which they can use information learned from specific users (individual) and groups of users (Group or Population).
- Systems vary according to how they validate the recommendations they make. Some use indirect information contained in correlations whereas others use explicit direct judgments of content.
- Some systems take the sequence of pages into account, and some do not.

A table comparing our representative systems on these dimensions appears in Table 1. Due to space restrictions, we have had to abbreviate our discussion of recommender systems, but we invite the reader to consult the references to follow up on the details of these approaches.

3 Session Specifc Information Needs Model

Like other researchers, we have chosen to conceptualize web browsing as a search for content satisfying a specific, well-defined "information need". Like many systems, we observe choices made by users while browsing. In our model, however, we are interested in the user's session specific information need and we use the user's individual fine-grained browsing actions and the exact sequence of pages visited by the user to find out what it is. First, we give some general background on our browsing-behavior based approach to recommendation and then we briefly describe several specific algorithms.

3.1 Browsing Behavior Models

Consider the example suggested by Figure 1. Imagine the user, needing information about marine animals, sees a page with links on "Dolphins" and "Whales". Clicking on the "Dolphins" link takes the user to a page about the NFL Football team. As this page does not contain the information that this user is seeking (at least, not at this time), the user "backs up". As the word "Football" appears on the previous page but not on the current page, this "backing-up" behavior suggests that "Football" might not be relevant to his/her search

Dimensions	COB	CF	CB	HBM	Our Models
Specific Site/Domain	Yes	Yes	Yes	No	No
Model Acquisition	Learning	Learning	Learning	Hand-coded	Learning
Annotation Required (training)	No	No	Yes	No	Yes
Annotation Required (performance)	No	No	No	No	No
Reference Set	Population	Group	Individual	None	Individual/Group/Population
Recommendation Validity	Indirect	N/A	Direct	Direct	Direct
Using Sequential Information	No	No	No	Yes	Yes

Table 1: Techniques for Recommender System

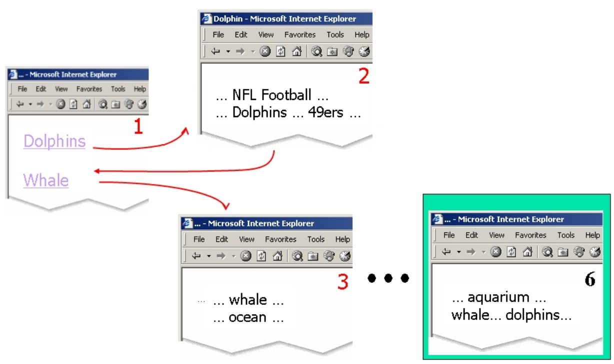

Figure 1: Browsing Behavior of Words, within Session

task. The user then tries the "Whale" pointer, which links to a page whose title includes "whale", and which includes whales, oceans, and other marine terms in its content. The user then follows a link on that page, anchored with "whale" to another with similar content, and so forth, until terminating on a page about a local aquarium. This behavior suggests that words such as "Whale" and "Ocean" are relevant, but "football" is not.

The above observation suggest that the user's current information need can be identified from the pages the user visits and the actions that the user applies to the pages. The content of the page is communicated by the roles that words play on the page. We make the simplifying assumption that we can represent the information need of the session by a set of significant words from the session. We further assume that the significance of words can be judged independently for each word from the roles played by instances of the word on pages throughout the sequence (e.g., appearing as plain text, highlighted in a title, appearing in followed hyperlink, etc.). To capture the reaction of the Web user, we recorded a number of browsing features from the page sequence. Our papers [14, 15] provide a complete list of these "browsing features".

To obtain the user model for Web recommendation, we first collected a set of annotated web logs (where the user has indicated which pages are IC-pages), from which our learning algorithm learned to characterize the IC-page associated with the pages in any partial subsession [14, 15].

3.2 Specific Models

Within the framework described above, we have implemented a number of specific algorithms for identifying information-need-revealing patterns. These patterns can be used to form queries to a search engine which does the actual retrieval of recommended pages. All of the algorithms require data labeled by human subjects.

IC-word: is our original algorithm. It requires subjects to explicitly identify pages with useful content. In this algorithm, we attempt to predict if a word that occurs in pages during the user's browsing sequence will appear on the information content page identified by the user. This prediction is done for each word independently and is based on only the browsing features of words in the sequence; that is, their pattern of occurrences within pages in the sequence and the actions applied to the pages they appear on. Any word with the same features will get the same score.

IC-Relevant: is a new algorithm developed for the current study. It requires subjects to explicitly indicate words that were relevant to their need. In this algorithm, we attempt to predict if a word that occurs in the pages of the user's browsing sequence will be in a set of words the user explicitly marks as relevant keywords.

IC-Query: is our most sophisticated algorithm and was also newly developed for this study. It is based on the observation that many of the words occurring on IC-pages are general (e.g., "the", "page", etc.) and therefore not particularly relevant to the page content. In particular, few of these words would help locate this page, in a search engine. We also observe that the words used in a search query are not independent. The specific combination of words and the order they appear are significant. The goal of the IC-Query algorithm is to find the 4-word search query that would most likely return the IC-pages identified by the user. Empirical investigations has revealed that 4-word queries are quite effective. The precise details of training the IC-Querymodel are complex and will be the subject of a future paper.

3.3 Experimental Design of the LILAC Study

Earlier laboratory studies revealed significant potential for behavior based methods [14, 15] and were an important motivation for the current work. The current study, a large field experiment code-named "LILAC" (Learn from the Internet:

How does your page compare to the recommended page?

- Fully answered my question
- Relevant, but does not answer my question fully
- Interesting, but not so relevant
- Remotely related, but still in left field
- Not related at all

Figure 2: Evaluation dialog options

Log, Annotation, Content), was intended to gather training data for creating future recommendation models and to evaluate the quality of recommendation models on a wide sample of users working on realistic, unconstrained tasks, seeking information from arbitrary sites on the Web.

LILAC was scheduled to last 5 weeks and involved 104 participants who installed a modified version of the internet-explorer web-browser, *WebIC*, on their home or business computers. Users were encouraged to disable tracking during personal or confidential browsing and were given the option of declining to submit web logs. User's were paid an honorarium for both their time and the number of sessions they generated.

The experimental design had 4 cases based on the model used to provide recommendations to the user. The four models were Followed Hyperlink Word (FHW) [1], a sensible recommendation strategy taken from the literature, our original behavior based model IC-word and our two newer methods IC-Relevant and IC-Query described above. Each of these models produced a set of words which we then sent to the Google™ search engine. Our recommendation consisted of the first results page returned by Google.

The joint goals of obtaining training data and evaluating models lead to a slightly complex experimental protocol: Subjects were asked to browse normally, but explicitly mark information content pages by pressing a button on the browser tool bar. This step provides us with a complete sample of the user's behavior: a browsing sequence and the resulting information content page. This data can be used for training future models.

At this point, the recommender generates a web-page recommendation using a randomly chosen recommendation model. The subject is then asked to compare the usefulness of the page they marked as having satisfying information content to the page generated by the recommender model (See Figure 2). This second step allows us to assess the ability of the chosen recommender model relative to the user's own standard of quality and to assess the value added to the user's existing information search efforts.

In the case where subjects could not find a page that addresses his/her information need, subjects were instructed to click on the "Suggest" button. *WebIC* presents a recommended page for review. Subjects were then asked to provide an absolute subjective rating of the usefulness of the suggested page with respect to his/her current information needs. Users were also asked to rate subsets of words appearing in the sessions according to how relevant they felt they were to their current needs.

[1]FHW is based on a simple rule: rank words according to how often they appear in the anchor text of followed hyperlinks. This is essentially the core of the model used in "Inferring User Need by Information Scent" (IUNIS) model [7].

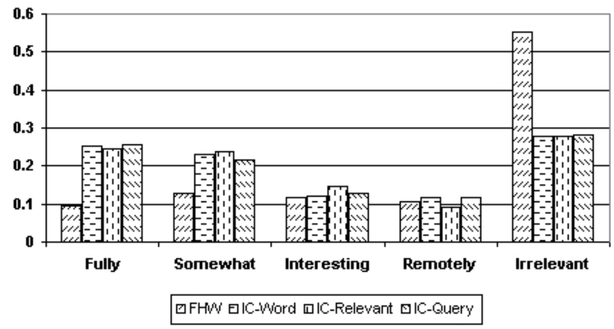

Figure 3: Overall Results of the LILAC Study

Data obtained from earlier weeks in the study were used to train improved behavior models for the current week of the study.

The 104 subjects visited 93,443 web pages, marked 2977 pages as IC-pages and asked for recommendations by clicking the "Suggest" button 2531 times over the course of the 5-week LILAC study. Summary statistics for the comparative ranking used for IC-pages appear in Figure 3. The bars show the relative percentage of each of the evaluation responses for each model. The best models would be expected to have more "Fully" answered ratings and fewer "Irrelevant" ratings.

As suggested by this figure, and confirmed by statistical tests (shown in "http://www.web-ic.com/lilac/results.html"), each of the different IC-models perform better than the baseline model (FHW). This result supports our basic assumption that we are able to provide useful recommendations by integrating the user's browsing behaviors into the prediction. Further analysis of the results will appear in a future paper.

4 *WebICLite*— An Effective Complete-Web Recommender System

The *WebIC* system that we used in the LILAC study has evolved into the *WebICLite* recommendation system, whose interface appears in Figure 4. *WebICLite* is also a client-side, *Internet Explorer*-based multi-tab web browser, that observes the user's browsing behavior, extracts the browsing properties of the words encountered, and then uses those browsing properties to predict the user's current information need, which it then uses to suggest (hopefully) useful pages from anywhere on the Web, without any explicit input from the user. It first gathers browsing properties for essentially all of the words that appear in any of the observed pages in the current session, then uses a model of user browsing patterns, obtained from previously annotated web logs, to gen-

Figure 4: *WebICLite* — An Effective Complete-Web Recommender System

erates an appropriate query to a search engine (here Google), which produces a candidate page to present to the user.

4.1 Hybrid Recommender models

Data from the LILAC study suggests that people tend to surf the Web by following some general browsing session patterns. One example of a search pattern looks like:

Query a search engine (Q)
Obtain a search results page, P
Open one URL from P
Return to P
Obtain another URL from P
Return to P
…

This pattern suggests that this query to this search engine is not producing the relevant page.

We found that some models work better for certain browsing patterns, as determined by general characteristics of the current session. For example, our evidence shows that IC-Relevant works better than any other models for the above pattern. *WebICLite* therefore includes a set of rules to choose the model that works best for the current browsing session.

4.2 Ongoing Evaluation

In the current implementation, the user can click "Suggest" to ask *WebICLite* to propose a Web page, anytime s/he needs assistance.

In order to collect the relevance feedback, which our system can use to improve its performance, *WebICLite* will then ask the user to evaluate the suggested page, using the interface shown in Figure 5. Here, the user is asked to "Tell us what you feel about the suggested page", to indicate whether the information provided on the page suggested by *WebICLite* was relevant for his/her search task, just as the user did in LILAC. Note that this is optional, if the user could provide such feedback, we can train his/her personalized models.

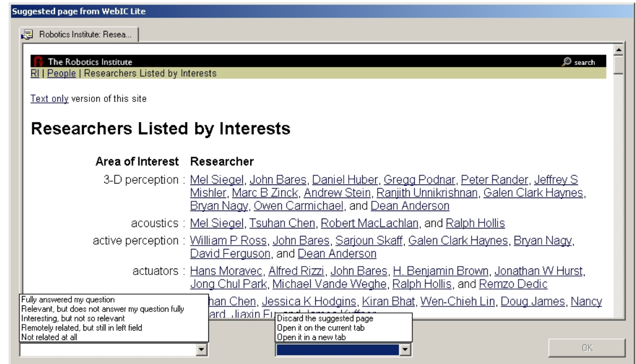

Figure 5: The Evaluation Interface of *WebICLite*

WebICLite also provides the options that allow the user to keep surfing from the suggested page P, by asking the user if s/he wants to *Discard* P, *Open* P *on the current tab*, or *Open* P *in a new tab*.

4.3 Learning from Evaluation

In order to train our models in LILAC, the study participants must actively label IC-pages while browsing the Web; of course, this is inconvenient for the user, and unrealistic in a production version of the product. To solve this data collection problem, we propose to passively train a model based on previous evaluation results. Recall that every time a user requests a recommendation, we generate a search query using one of the models to return a page to the user, which s/he is then asked to evaluate. If we assume that the search engine (e.g., Google) remains relatively consistent (i.e., in terms of the quality of pages returned) over time, we can infer the evaluation of the search query from the actual evaluation of the recommended page. Thus we can label each query as one of the evaluation outcomes. We can then attempt to learn a "Fully"-page classifier by considering (as positive examples) only the queries that are evaluated as "Fully", and the rest as negative.

In the LILAC study, the IC-Query models were trained directly based on the pages marked IC-pages. In the last week, we changed the experimental protocol to train a model based on all queries that resulted in a "Fully" evaluation in the previous weeks. Figure 6 presents the results of these two models, trained by the pages marked IC-page, vs the retrospective one.

As suggested here, both approaches produce similar performance. This result is significant as it will allow us to continuously refine the model without requiring user input to label IC-pages while browsing the Internet. Importantly, this alternate training method will make the use of *WebICLite* more realistic in real world situations.

5 Conclusion

In this paper we introduced two new behavior-based models. While we are still in the process of analyzing our results, we

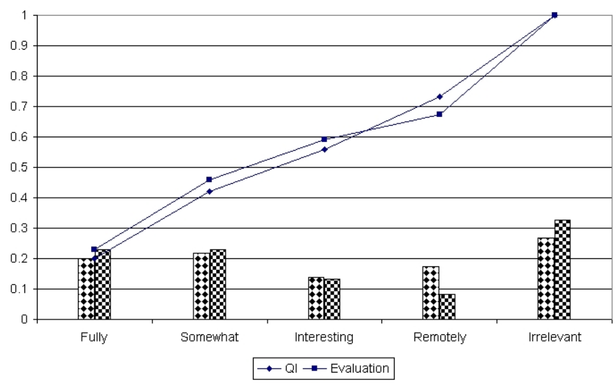

Figure 6: Alternate Training on Evaluation

have some evidence that these new models outperform both the control model (FHW) and our existing model. Our current study has shown that the positive potential of behavior-based recommendation models seen in our laboratory studies can be transferred to real users browsing arbitrary web pages during day-to-day tasks. In the couse of this study, we have collected a large amount of high quality data and expect to train significantly better models in the near future. While still preliminary, we believe our results support the conclusion that behavior-based models have a unique ability to provide responsive session-specific recommendations independent of any particular site and that these models have a promising range of useful future extensions.

Acknowledgement

The authors gratefully acknowledges the generous support from Canada's Natural Science and Engineering Research Council, the Alberta Ingenuity Centre for Machine Learning (http://www.aicml.ca), and the Social Sciences and Humanities Research Council of Canada Initiative on the New Economy Research Alliances Program (SSHRC 538-2002-1013).

REFERENCES

1. F. Abbattista, M. Degemmis, N. Fanizzi, O. Licchelli, P. Lops, G. Semeraro, and F. Zambetta. Learning user profiles for content-based filtering in e-commerce. In *Proceedings AI*AI Workshop su Apprendimento Automatico: Metodi e Applicazioni*, Sienna, Italy, 2002.

2. R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *Proc. of the 20th Int'l Conference on Very Large Databases (VLDB'94)*, Santiago, Chile, Sep 1994.

3. R. Agrawal and R. Srikant. Mining sequential patterns. In *Proc. of the Int'l Conference on Data Engineering (ICDE)*, Taipei, Taiwan, Mar 1995.

4. C. Anderson and E. Horvitz. Web montage: A dynamic personalized start page. In *Proceedings of the 11th World Wide Web Conference (WWW 2002)*, Hawaii, USA, 2002.

5. D. Billsus and M. Pazzani. A hybrid user model for news story classification. In *Proceedings of the Seventh International Conference on User Modeling (UM '99)*, Banff, Canada, 1999.

6. J. Budzik and K. Hammond. Watson: Anticipating and contextualizing information needs. In *Proceedings of 62nd Annual Meeting of the American Society for Information Science*, Medford, NJ, 1999.

7. E. Chi, P. Pirolli, K. Chen, and J. Pitkow. Using information scent to model user information needs and actions on the web. In *ACM CHI 2001 Conference on Human Factors in Computing Systems*, pages 490–497, Seattle WA, 2001.

8. A. Jennings and H. Higuchi. A user model neural network for a personal news service. *User Modeling and User-Adapted Interaction*, 3(1):1–25, 1993.

9. H. Lieberman. Letizia: An agent that assists web browsing. In *International Joint Conference on Artificial Intelligence*, Montreal, Canada, Aug 1995.

10. D. S. W. Ngu and X. Wu. Sitehelper:a localized agent that helps incremental exploration of the world wide web. In *6th International World Wide Web Conference*, pages 691–700, Santa Clara, CA, 1997.

11. P. Pirolli and W. Fu. Snif-act: A model of information foraging on the world wide web. In *Ninth International Conference on User Modeling*, Johnstown, PA, 2003.

12. P. Resnick, N. Iacovou, M. Suchak, P. Bergstorm, and J. Riedl. Grouplens: An open architecture for collaborative filtering of netnews. In *Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work*, pages 175–186, Chapel Hill, North Carolina, 1994. ACM.

13. A. Stefani and C. Strapparava. Personalizing access to web sites: The siteif project. In *Proceedings of the 2nd Workshop on Adaptive Hypertext and Hypermedia HYPERTEXT'98*, Pittsburgh, USA, June 1999.

14. T. Zhu, R. Greiner, and G. Häubl. An effective complete-web recommender system. In *The Twelfth International World Wide Web Conference(WWW2003)*, Budapest, HUNGARY, May 2003.

15. T. Zhu, R. Greiner, and G. Häubl. Learning a model of a web user's interests. In *The 9th International Conference on User Modeling(UM2003)*, Johnstown, USA, June 2003.

16. I. Zukerman and D. Albrecht. Predictive statistical models for user modeling. *User Modeling and User-Adapted Interaction*, 11(1-2):5–18, 2001.