Using Learned Browsing Behavior Models to Recommend Relevant Web Pages

Tingshao Zhu, Russ Greiner
Dept of Computing Science
University of Alberta
Edmonton, Alberta T6G 2E8
{ tszhu, greiner }@cs.ualberta.ca

Gerald Häubl
School of Business
University of Alberta
Edmonton, Alberta T6G 2R6
Gerald.Haeubl@ualberta.ca

Kevin Jewell, Bob Price
Dept of Computing Science
University of Alberta
Edmonton, Alberta T6G 2E8
{ kjewell, price }@cs.ualberta.ca

Abstract

We introduce our research on learning browsing behavior models for inferring a user's information need (corresponding to a set of words) based on the actions he has taken during his current web session. This information is then used to find relevant pages, from essentially anywhere on the web. The models, learned from over one hundred users during a five-week user study, are session-specific but independent of both the user and website. Our empirical results suggest that these models can identify and satisfy the current information needs of users, even if they browse previously unseen pages containing unfamiliar words.

Keywords: machine learning, web mining

1 Introduction

While the World Wide Web contains a vast quantity of information, it is often difficult for web users to find the relevant information they are seeking. While modern search engines have now made search over billions of web pages an everyday occurrence, this approach still requires the user to explicitly initiate a search and to formulate a specific search query. The alternative approach, which involves surfing the web by following appropriate links, is also challenging as users cannot always determine which links are most likely to lead to the relevant information. Our goal is a client-side Web recommender system that simply observes the user's browsing behavior, then uses this information to suggest relevant pages from anywhere on the Web, without requiring the user to provide any additional input.

Historically, information filtering approaches avoid the need for explicit queries, but they assume the user has stable long-term interests that can be used to pick out relevant material from an ongoing, dynamically changing flow of content [Lewis and Knowles, 1997]. In this case, a user's interests can be determined by statistical inference over extensive historical samples of the content that the user preferred. Alternatively, information assistants are designed to respond effectively to transient short-term information needs that vary from task to task using content stored in large stable repositories. In contrast, we present a method to determine the user's current information need without either explicit queries nor a large sample of past content.

2 Model

For the purposes of our present work, we focus on users using the Web to obtain some information. Our goal is to automatically provide the user with additional relevant content beyond what would be immediately revealed by either links on their current page, or results from a search query they might initiate. We assume that this activity can be separated into distinct sessions in which the user is pursuing a single specific information need.

As the user browses the Web, he visits pages and takes actions on these pages, such as following a link, backing up to a previous page, or bookmarking a page. We view the user's web browsing actions as implicit judgments on the content of the pages (with respect to his current information need). Presently, we assume that the user is looking for text-based resources. In this case, the content of the pages is represented by words. We theorize that the actions the user takes while browsing provide evidence about the degree to which words on visited pages represent the user's current information need.

This intuition has appeared in previous work in special cases: for instance, the idea that words appearing in the anchors of hyperlinks followed by the user are probably more representative of user's needs than words that do not. Indeed, the Letizia [Lieberman, 1995] agent and the recommender created by Watson [Budzik and Hammond, 1999] use observations of user behavior to help them locate pages, but they are both based on a limited set of hand-coded heuristic rules. Alternatively, correlation-based recommenders [Agrawal and Srikant, 1995] tend to be tied to a specific site. Moreover, such systems typically suggest that the current user go to pages that other similar users have visited, without knowing whether previous users actually found the page helpful. Our "behavior-based inference" differs from content-based systems [Billsus and Pazzani, 1999], as we do not require any previous experience with the actual words in the hyperlink anchor; moreover, our models are explicitly trained to identify pages that satisfy any particular information need.

For example in our models, following any hyperlink provides evidence for the relevance of those words appearing in the hyperlink anchor, independent of what words are actually present in the hyperlink. Consequently the technique is not restricted to a subset of indexed sites. Indeed, we attempt to predict the words that best capture the user's information need with a combination of basic text-based features (e.g., the overall frequency of the word in the user's current session, etc.), as well as a generalized set of behavior-based features (e.g., the presence of the word in anchor text, the use of the word on a page from which the user backed up, and the order in which a list of links using the word were selected, etc.).

Since the precise significance of a word appearing in a followed hypertext anchor, or being in some font, etc., is difficult to state a priori, we use machine learning methods to learn the weight of each of our features, towards identifying which word will be relevant. This is based on data collected from a pool of calibration subjects. Once the weights have been calculated, we can apply our user-independent models to new users without retraining. In fact, since the models are trained on behaviors (e.g., the value of appearing in a followed hyperlink, etc.) *instead* of the information viewed (i.e., "the user is interested in 'baseball' "), the models can also be applied to new domains without retraining, and can respond to new user needs as they arise during browsing.

To apply machine learning to this domain, we need a training set. The input to the machine learning process consists of the various text-based and behavior-based *features* associated with each word in the current context [WebIC, 2005], plus a target label describing whether or not the word is representative of the user's interests. We have developed three methods to produce this label using different sources of information, each based on the subjects in a calibration pool.

IC-wordModel: Each subject hand-labels pages that contain desired information content as IC-pages (otherwise the pages are considered ¬IC-pages by default). We then assume that all non-stopwords appearing on an "IC-page" are representative of the user's current information need, whereas words that do not are unrepresentative.

IC-RelevantModel: Each subject explicitly chooses words from the current browsing session that they felt best represented their current information need. All other words are assumed to be unrepresentative.

IC-QueryModel: Each subject hand-labels pages that contain desired information content as IC-pages (otherwise the pages are considered ¬IC-pages by default). But while the IC-word model viewed *all* words in the IC-page as important, this model is more selective, as it includes only the subset of non-stopwords that enable a search engine (e.g., Google) to find that page. In a separate sub-project, we estimate a search-engine-specific function that, given a page (here an IC-page, p) and a list of words W, estimates whether giving W as a query to the search engine will produce p. Using this function, we label only the k highest-scoring words from the session as representative.

Given a set of browsing sessions, we can form a matrix where each row corresponds to a word that appears, and each column to one of the browsing features. We then label each row using one of the models described above, and run a standard learning algorithm (e.g., C4.5) to construct 3 different classifiers. Each classifier predicts whether a word is representative of the current information need from the text-based and behavior-based features extracted from the user's current browsing session.

3 LILAC Study

The goals of the LILAC (Learn from the Internet: Log, Annotation, Content) study are to evaluate the browsing behavior models presented above (Section 2), and to gather data for future research, from people working on their day-to-day tasks.

3.1 Overview of WebIC

*W*ebIC (Figure 1) is a client-side, Internet Explorer-based web browser and recommender system. It uses information from the current browsing session to recommend a page, from anywhere on the Web, that it predicts the user will find useful.

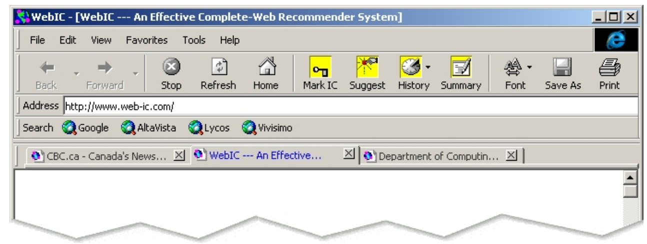

Figure 1: *W*ebIC — An Effective Complete-Web Recommender System

To make a prediction, *W*ebIC first computes the browsing features for all stemmed non-stopwords that appear in any page of the current browsing session. It then determines which of these words to submit as a query to a search engine, using one of the models learned previously. *W*ebIC recommends the top page returned from the query, as the page it considers most likely to satisfy the user's current information need.

We modified *W*ebIC for the LILAC study. Here, whenever the user requests a recommendation by clicking the "Suggest" button, *W*ebIC will select one of its models randomly to generate a recommendation page. As one of the goals of the LILAC study is to evaluate our various models, this specialized version of *W*ebIC will therefore ask the user to evaluate this proposed page.

In addition, we also instructed the participants to click "MarkIC" whenever they found an IC-page. After marking an IC-page, *W*ebIC will recommend an alternative web page as before (excluding the IC-page), as if the user had clicked "Suggest" here. Once again, this specialized version of *W*ebIC will then ask the user to evaluate this recommended page.

As part of this evaluation, the user is asked to "Tell us what you feel about the suggested page", to indicate whether the information provided on the suggested page was relevant to his/her search task. There are two categories of relevance evaluations: *related* and *not related at all*.

In addition, the user was also asked to select informative "Descriptive Keywords" from a short list of words that *W*ebIC predicted as relevant. The information collected here will be used to train the IC-Relevant model described above.

3.2 Experiment Design

To participate in the LILAC study, the subjects were required to install *W*ebIC on their own computer, and then use it when browsing their own choice of non-personal English language web pages. They were told to use another browsing engine when dealing with private issues, such as banking, e-mail, or perhaps chat-rooms, etc.

*W*ebIC kept track of all the interactions, including a record of the pages the user visited, as well as the evaluation data for the pages that *W*ebIC recommends.

LILAC considered four models: the three described in Section 2 — IC-word, IC-Relevant, and IC-Query— and "Followed Hyperlink Word" (FHW), which is used as a baseline. FHW collects the words found in the anchor text of the followed hyperlinks in the page sequence. As such, there is no training involved in this model. This is similar to "Inferring User Need by Information Scent (IUNIS)" model [Chi *et al.*, 2001]. In all models, words are stemmed and stopwords are removed prior to prediction.

We used the data collected during the study to weekly retrain each of our IC-models. That is, the users initially used the IC-word$_0$, IC-Relevant$_0$ and IC-Query$_0$ models, which were based on a model obtained prior to the study. On the 2nd week, they used the IC-word$_1$, IC-Relevant$_1$ and IC-Query$_1$ models, based on the training data obtained from week 1, as well as the prior model. We repeated this process throughout the study.

3.3 Overall Results

A total of 104 subjects participated in the five-week LILAC study, from both Canada (97) and the US (7); 47% are female and 53% are male, over a range of ages (everyone was at least 18, and the majority were between 21-25). The subjects visited 93,443 Web pages, clicked "MarkIC" 2977 times and asked for recommendations by clicking the "Suggest" button 2531 times. As these two conditions are significantly different, we analyzed the evaluation results for "Suggest" and "MarkIC" separately.

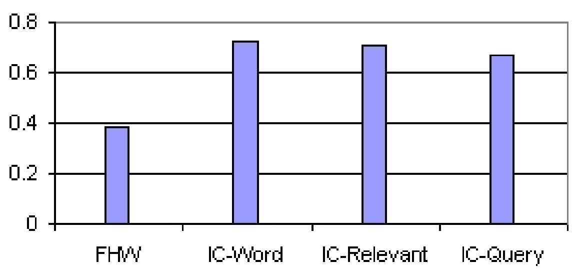

Figure 2: How often the User rated a Recommended Page as "Related", after "Suggest"

Figure 2 indicates how often the user considered the "Suggest"ed page to be "Related". We see that each of our 3 IC-models works much better than the baseline model — each was over 65%, versus the 38% for FHW.

Figure 3 shows the evaluation results for the recommended pages after the user clicked "MarkIC". We again observe that our IC-models work better than FHW. The scores for IC-word and IC-Relevant remained at around 70%, roughly the same values they had for the "Suggest" case, while the IC-Query model increased from 66% to 74%. We also observed that FHW increased by almost 10%. As an explanation, we speculate the following: If the subject is able to find an IC-page, then the links followed in the current session appear to provide a very strong hint of what constitutes the relevant information content; and FHW benefits significantly from this hint.

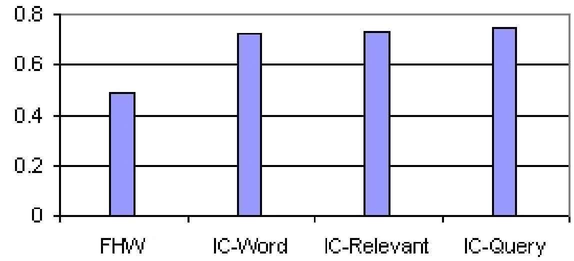

Figure 3: How often the User rated a Recommended Page as "Related", after "MarkIC"

4 Conclusion and Future Work

We have demonstrated a practical and extensible framework for a behavior-based web recommender system. The basic framework covered here was able to find relevant pages 70% of the time and consistently outperformed a common baseline method (FHW). Our method can easily be extended to include additional features (e.g., timing, multiple search engines), and to employ personalized training models for custom feature weights. In addition, we anticipate our results, achieved using C4.5, could be improved by more sophisticated learning techniques. Extensive results from our user study suggest that behavior-based recommendation is a promising approach for automatically finding the pages, from anywhere on the Web that address the user's current information need, without requiring the user to provide any explicit input.

References

[Agrawal and Srikant, 1995] R. Agrawal and R. Srikant. Mining sequential patterns. In *Int'l Conference on Data Engineering (ICDE)*, 1995.

[Billsus and Pazzani, 1999] D. Billsus and M. Pazzani. A hybrid user model for news story classification. In *User Modeling*, 1999.

[Budzik and Hammond, 1999] J. Budzik and K. Hammond. Watson: Anticipating and contextualizing information needs. In *American Society for Information Science*, 1999.

[Chi *et al.*, 2001] 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.

[Lewis and Knowles, 1997] D. Lewis and K. Knowles. Threading electronic mail: A preliminary study. *Information Processing and Management*, 33(2), 1997.

[Lieberman, 1995] H. Lieberman. Letizia: An agent that assists web browsing. In *IJCAI*, 1995.

[WebIC, 2005] 2005. http://www.web-ic.com.