Behavior-Based Web Page Evaluation

Ganesan Velayathan
The Graduate University for Advanced Studies
National Institute of Informatics (NII)
2-1-2 Hitotsubashi
Chiyoda, 101-8430 Tokyo, Japan
ganesan@grad.nii.ac.jp

Seiji Yamada
National Institute of Informatics (NII)
2-1-2 Hitotsubashi
Chiyoda, 101-8430 Tokyo, Japan
seiji@nii.ac.jp

ABSTRACT

This paper describes our efforts to factor in a user's browsing behavior to automatically evaluate web pages that the user shows interest in, based on user browsing behaviors while browsing. To evaluate a webpage automatically, we have developed a client-side logging tool: the GINIS Framework. We do not focus just on clicking, scrolling, navigation, or duration of visit alone, but we propose integrating these patterns of interaction to recognize and evaluate a user's response to a given web page.

Categories and Subject Descriptors

H.1.2 **[User/Machine Systems]**: Human factors; H.3.3 [**Information Storage and Retrieval**]: Information Filtering.

General Terms

Human Factors, Design, Theory.

Keywords

Information extraction, web usage mining, automatic profiling, web browsing behavior, web-human interaction.

1. INTRODUCTION

The need to automatically build a user profile to assist in everyday user browsing tasks has been steadily increasing, due to information overload on the internet. Building a user profile that adapts to a user's daily interests is a challenging task. This is because it is hard to predict which web site most interests the user without asking the user to interact explicitly with the system.

The reason conventional profiles are not adaptive is that the user evaluation process has been manual – that is, a user must explicitly tell the system each time a webpage is relevant to the task or not. This is laborious, and users often lack the time or inclination to carry this out. A less intrusive approach to the construction of user profiles is required. How might this best be achieved?

When users use a web browser to browse the Internet, they (whether intentionally or not) interact with the browser's interface, by clicking, scrolling, book marking, printing or selecting text, etc. If we can detect and learn a user's patterns of behavior, we can use these individual browsing habits to perform automatic page evaluation, and hence preferential page selection and prediction.

We aim to build a behavioral interaction database which matches a user's actions (such as scrolling, clicking, etc) with pre-

established behavioral patterns mined from that user, thus giving some idea of the user's interest (or lack of interest) in a particular webpage. Using this method, we believe we can achieve a detailed, granular, self-constructing user profile. We believe that careful study of these patterns at the client-side will lead us to a better understanding of which pages most interest the user [1].

However, logging the user behavior at the client-side is a challenging task, for which we needed to build a high-level user client and logging tools. In this paper, we introduce the GINIS Framework, a framework developed to automatically log user browsing behavior.

2. RESEARCH OVERVIEW
2.1 Problem with Conventional Methods

Most of the profiling methods used at present consider any web pages contained in browser history as "positive training examples", which means that pages not reached or viewed by the user are regarded as "negative training examples". Expanding on this idea, [2]proposed measuring the time spent on a web page as a factor for evaluation, but this is not ideal; some users just open a web page and then navigate to another page without closing the first page. This practice is becoming more common with the use of tab-browsers, although this can be ameliorated by enforcing a limit on maximum duration. Further, researchers like [3] introduce examining navigational history to evaluate interest, but this is not always a good indication; the user might easily navigate to a page but not be interested in its contents. At present, it is difficult to discover whether a certain page is of interest or not.

2.2 Overview

Developing client side logging tools would produce a better logging environment. Logging at the client side produces richer data, although it makes necessary the difficult task of building a rich browser. However, compared to conventional web usage mining methods, like proxy side collection or server side collection[3], client side logging gives a richer and more accurate picture of the user's behavior.

3. USER BEHAVIOR
3.1 Selecting and Gathering

User's behavior during browsing can be classified into two groups: direct (implicit) behaviors and indirect (explicit) behaviors. Direct behaviors are behaviors that directly show that the user might be interested in the contents of the web page or not, e.g. printing a web page (behavior indicating interest) or closing a browser window (behavior indicating non-interest). Indirect behaviors are behaviors that, taken by themselves as single

behaviors, do not show whether the user likes or dislikes the web page, but when studied in terms of the sequence occurrence of the action, yield patterns that could predict the user's interest or non-interest.

3.2 User Behavior Examples

Table 1. Direct user behaviors

ID	Behavior	Interest	Level (high/medium /low)
(DB$_1$)	Print a page	Yes	High
(DB$_2$)	Bookmark a page	Yes	High
(DB$_3$)	Click a link	Yes	Medium
(DB$_4$)	Input a new URL	Yes	High
(DB$_5$)	Copy text to clipboard	Yes	High
(DB$_6$)	Terminate the browser	No	High
(DB$_7$)	Stop Navigation	No	Medium

In the case of direct behaviors with the browser, we might say that a single action often indicates either the user's interest or non-interest. However, the level of interest could differ from user to user.

Table 2. In-direct user behaviors

ID	Behavior	Explanation
(IB$_1$)	N(P$_1$)-N(P$_2$)…N(P$_n$)	Navigate from P$_1$ to P$_n$
(IB$_2$)	N(P$_1$)-[S(D$_n$)-S(U$_p$) -S(D$_n$)]	Navigate to P$_1$, scroll down n-times, scroll up n-times, and scroll down n-times.
(IB$_3$)	N(P$_1$)-C(L$_1$)-N(P$_2$)	Navigate to P$_1$, click a link on P$_1$, jump to P$_2$

N: Navigate, S: Scroll, D$_n$: Down, U$_p$: Up, C: Click

In the case of indirect user behaviors, one single action might not be enough to indicate interest in a particular page, but by studying the session of behavior and the action occurrence pattern, we might be able to extract indirect behaviors that can be said to indicate interest or non-interest in a page.

4. GINIS FRAMEWORK

In this section, we discuss the various modules of the GINIS Framework, which was developed specifically for web usage mining. The GINIS Framework consists of 4 main modules: a client interface to detect and log user behavior (Browser), a database to store the user log information (Raw Logger), a learning engine and a prediction engine.

4.1 Architecture

The web browser was developed based on MSIE6.0, and is capable of logging user behavior. The raw database stores raw user log data. The learning engine refines the logs from the raw database and stores it in the User Behavioral Database (UBD). Based on UBD, a decision tree is constructed [3] and stored at the server. During the testing stage, the prediction engine generates and compares the user's new behavior decision tree with the UBD

decision tree, on the basis of this comparison, predicts which pages interest the user and which do not.

Figure 1 shows the architecture of the GINIS Framework during the learning stage, and figure 2 shows the GINIS Framework testing stage.
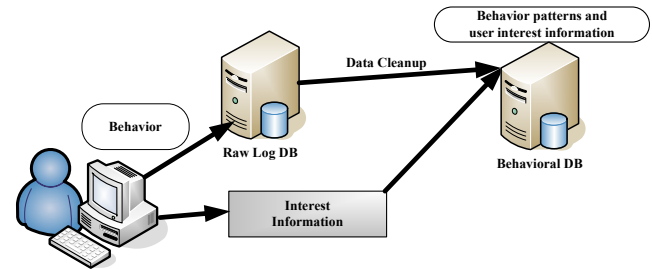
Figure 1. GINIS Learning Stage.
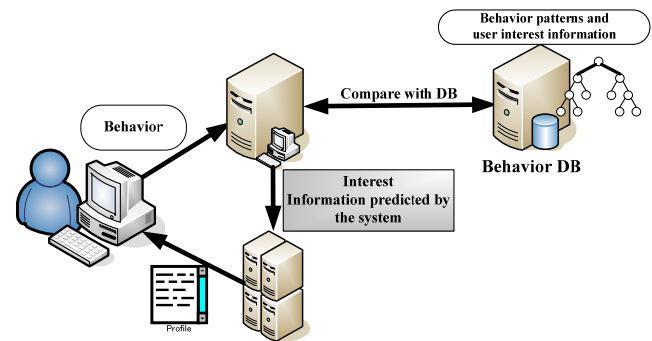
Figure 2. GINIS Testing Stage.

5. CONCLUSION

We have presented here a method to automatically detect and log user behavior at the client-side by creating a client-side browser. Whether intentionally or not, users display certain patterns of behavior when they come across a page that they like or dislike. By studying these patterns, user behavior during browsing could be made to play a major role in establishing which pages a user favors.

6. REFERENCES

[1] L. Catledge and J. Pitkow. Characterizing browsing behaviors on the World Wide Web. Computer Networks and ISDN Systems, 27(6), 1995.

[2] M. Morita and Y. Shinoda. Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval. In Proc. of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '94), pages 272-281, 1994.

[3] R. Cooley, Pang-Ning Tan, and Jaideep Srivastava. Discovery of interesting usage patterns from web data. Technical Report TR 99-022, University of Minnesota, 1999.

[4] T. Lim, W. Loh, and Y. Shih. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms, Machine Learning Journal,40:203-228, 2000.