

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 investigate factors in user's browsing behavior to automatically evaluate web pages that the user shows interest in. To evaluate web pages automatically, we developed a client-side logging/analyzing tool: the GINIS Framework. This work focuses primarily on client-side user behavior using a customized web browser and AJAX technologies. First, GINIS unobtrusively gathers logs of user behavior through the user's natural interaction with the web browser. Then it analyses the logs and extracts effective rules to evaluate web pages using C4.5 machine learning system. Eventually, GINIS becomes able to automatically evaluate web pages using these learned rules.

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

Web-human interaction, browser interface, navigation.

1. INTRODUCTION

Over the last decade, the Web has grown massively in size, popularity, business application, and number of users. Accompanying this remarkable growth many proposals have been made for tools and systems to assist with everyday user browsing and searching. Most previous research working towards web personalization had relied upon overt methods of asking users for their answers in order to construct user profiles. 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 sites interest the user most without asking the user to interact explicitly with the system. This is laborious; it requires time and inclination, and users often forget to notify the system. A less intrusive approach to the detect user interest is required.

With this in mind, this study proposes a new method of automatically discovering and judging user interest based on user browsing behavior. User behavior is defined here as the habitual actions performed by users when browsing and searching, such as clicking links, bookmarking, printing or selecting text, and so on.

This study focuses on highly accurate logging at the client-side. Also, two issues emerged with regard to acquiring correct user browsing behavior. The first was that overt collection of data for experimental purposes was not desirable; in short, if users were conscious of an experimental environment, this would interfere

Copyright is held by the author/owner(s).

WWW 2007, May 8-12, 2007, Banff, Alberta, Canada.

ACM 978-1-59593-654-7/07/0005.

with their usual browsing habits. The other issue is that it was necessary for the experimental instrument for data collection to be identical to the browser regularly used by the users. Starting from these considerations, we undertook construction of the GINIS Framework, to be identical with Internet Explorer 6.0. We used the GINIS Framework to conduct experiments with test participants. This paper presents those findings.

2. RESEARCH OVERVIEW

After carefully studying the daily browsing behaviors of users, we chose around 70 navigation actions and 40 user behaviors and built a monitor for the browser. We use the term *navigation action* here to describe the individual "components" of user behavior. We use the term *user behavior* to describe the result achieved by performing these navigation actions. For example, the navigation actions "Hit Backspace Key," "Click Back Arrow on Menu," or "Use Back Button on Mouse" all constitute the same "Move Back" user behavior. During the experimental stage, we gathered this information at the navigation action level, after which we pre-processed the log and construed user behavior. The logging attributes were designed based on the work presented by Catledge and Pitkow [1], with some additions.

The GINIS Framework consists of 4 main modules: a client interface to detect and log user behavior (the Browser), a database to store the user log information (the Logger), a learning engine (the Learner) and a prediction engine (the Predictor) [2].

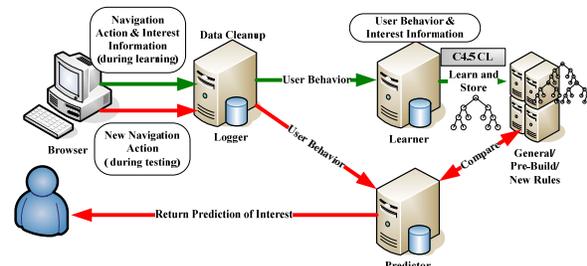


Figure 1. An Overview of GINIS Framework

3. EXPERIMENTS

3.1 Overview

We conducted systematic experiments using the GINIS Framework to gather data in order to discover new rules linking user behavior and interest. C4.5 was used as the classification algorithm [3]. Verification was performed using the 10-fold validation method. We designed the experiment to be performed within an open environment in order to expose and clarify unintentional user behavior while browsing. The users were requested to evaluate "like," "dislike," or "unknown" on the web pages they viewed by a questionnaire dialog box pop up or by clicking on the menu button.

3.2 Data Gathering and Pre-processing

10 unpaid volunteers (6 male and 4 female) were recruited to participate in this study. They ranged in age from 21 years old to 38 years old (mean age: 29.1 years). All of them had between 4 and 12 years of web browsing history (mean history: 8.3 years). Almost all the participants used the internet for more than 6 hours per day (mean usage time: 10 hours per day), mainly either during work hours or for pleasure and entertainment. The participants used the GINIS browser for a period of 16 to 31 days (mean usage duration: 22 days). A total of 2856 pages were evaluated, and 1997 pages were tagged as “*of interest*” while 859 pages were tagged as “*not of interest*” by the participants. GINIS was built with very high logging functionality to enable future use of the data. The original log file comprises over 460,000 lines, including 195,000 action data. In the preliminary data processing stage, we performed a data clean-up to 65%. Around 65,000 lines were left after this process. Most of the removed data were mouse movement logging (mouse locus and axis) and text copied/pasted to/from the clipboard by the user.

4. RESULTS

Classification learning was performed to 2856 cases of training data. Error rate was set at 25%. 10-fold cross validation was performed and 2027 pages (70.97%) were correctly classified, and 829 pages (29.03%) were incorrectly classified at this point. Furthermore, 13 rules were obtained. 8 of these were found to be rules governing pages “*not of interest*”, and 5 of these were rules governing pages “*of interest*.”

As a result of validating the data, we found that our evaluations contained many instances of inconsistency, for example, same behaviors but different evaluation. Following on from this, we performed further clean-up of the data by using a method of adopting those cases where inconsistent evaluations were more frequent, given the occurrence of the same behavior only, and an equal number of random cases. As a result, 2249 cases were left, and within these 1885 were tagged as “*of interest*” and 364 were tagged as “*not of interest*” by the participants. Of these evaluations, 2005 pages (89.15%) were correctly classified, and 244 pages (10.85%) were incorrectly classified. This represented a significant improvement in classification capability.

Furthermore, 26 rules were output in total. 5 of these were found to be rules governing “*of interest*”, and 21 of these were rules governing “*not of interest*.” The default class was set to “*of interest*”. We found that most of the rules generated were not governed by “Stay Time: the time spent on a web page”, which conflicts with the findings of most previous studies [4].

5. DISCUSSION AND FUTURE WORK

5.1 Data Cleaning

Two things that affected our clean-up result were search engine result pages and the use of online dictionaries. Most of the form input takes part here over at these pages. The participants in this study tended to search for a particular word and then evaluate the answer of the query.

5.2 Classification Rules

One early objective of this work was to locate common features of web browsing shared by all users. After all user behaviors were gathered, the removal of inconsistent evaluations yielded a tremendous improvement in classification learning accuracy.

From above experiment result, the mean accuracy was 91.4% (SD=4.80), precision was 0.915, and recall was 0.959 for the rules “*of interest*,” which was a very satisfying result for this classification learning. For the classification “*not of interest*”, mean accuracy for the 21 rules generated was 74.8% (SD=9.32), with precision of 0.719 and recall of 0.541. Over the experiment, only 364 pages were classified as “*not of interest*”, which is around 18.2 % of the total data. Since the experiment was not performed in a controlled environment, it was difficult to collect an equal number of “*of interest*” and “*not of interest*” web pages. Most rules were found not to be governed by the time a user spends on a web page.

5.3 Present Limitations and Future Work

At present, our research was limited to searching for patterns of user behavior from all the behaviors available during browsing. We did not make detailed examination of the order of occurrence and situation of occurrence of this behavior (whether mouse or keyboard was used, etc.) of the behaviors themselves. One more factor limiting this research at the moment is that the log was collected in detail, but the log cleaning process was performed semi-automatically. This is because navigation actions tend to occur simultaneously, often making it difficult to determine which navigation action is part of which behavior. The scroll and mouse action is the most frequently performed action on a page; we assume that more detailed study of this scroll and mouse action will result in a better understanding of action/behavior conversions. In addition, a great number of mouse locus were logged during this experiment, and these still have to be studied carefully, as there is a strong possibility of finding new behaviors as well as valuable new research ideas.

6. CONCLUSION

We have presented here a method to automatically detect and log user behavior at the client-side by creating a client-side browser (the GINIS Framework). We used this framework to conduct an open environment field study (rather than a controlled laboratory study). Past research has tended to focus on the time spent on a web page as the main factor in evaluating implicit user interest. Our experiment confirms that the time spent on a page is not the only main factor; scroll action, form action, searching text, copying text etc. should also be taken into consideration as factors in evaluating user interest.

7. REFERENCES

- [1] Catledge, L.D, Pitkow, J.E.: Characterizing Browsing Strategies in the World-Wide Web, *Computer Networks and ISDN Systems*, 27(6), 1995.
- [2] Velayathan, G, Yamada, S.: Behavior Based Web Page Evaluation, *Proceedings of International Conference on Web Intelligence-Workshop*, 2006.
- [3] Quilan, R. : C4.5: Programs for Machine Learning, Morgan Kaufmann, San Mateo, CA, 1993.
- [4] Morita, M., Shinoda, Y.: Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval, In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 1994)*, pp. 272-281, 1994.