

Using Web Browser Interactions to Predict Task

Melanie Kellar and Carolyn Watters
Faculty of Computer Science, Dalhousie University
Halifax, Nova Scotia, Canada
{kellar, watters}@cs.dal.ca

ABSTRACT

The automatic identification of a user's task has the potential to improve information filtering systems that rely on implicit measures of interest and whose effectiveness may be dependant upon the task at hand. Knowledge of a user's current task type would allow information filtering systems to apply the most useful measures of user interest. We recently conducted a field study in which we logged all participants' interactions with their web browsers and asked participants to categorize their web usage according to a high-level task schema. Using the data collected during this study, we have conducted a preliminary exploration of the usefulness of logged web browser interactions to predict users' tasks. The results of this initial analysis suggest that individual models of users' web browser interactions may be useful in predicting task type.

Categories and Subject Descriptors

I.5.1 [Pattern Recognition]: Models - *Structural*. H.5.4. [Information Interfaces and Presentation]: Hypertext/Hypermedia - *Navigation*.

General Terms: Algorithms, Experimentation, Human Factors.

Keywords: Web, task, field study, information filtering, task prediction, decision tree.

1. INTRODUCTION

The automatic categorization of a user's web activity could serve useful in developing more intuitive and effective information filtering systems. Many information filtering systems rely on implicit measures to infer user interest, such as dwell time, mouse and keyboard activity, and document interactions (e.g., copy, cut, save, print) [7]. However, we do not yet have a solid understanding of the role task plays in the effectiveness of implicit measures for information filtering. While the evidence is somewhat incongruous, there is evidence to suggest that the effectiveness of an implicit measure may be dependant upon the task at hand. For example, previous research has found through laboratory experiments that the usefulness of dwell time was influenced by the type of task [5; 2]. However, a later field study found that task did not significantly impact the usefulness of dwell time as an implicit measure of interest [4]. We believe that once it becomes clear which implicit measures are most effective for a task type, the ability to automatically infer a user's task would allow us to apply the most appropriate measures of user

interest, thereby improving the effectiveness of information filtering systems.

While previous research has examined different aspects of automatic task prediction, such as the identification of goals [6], active tasks [8], and focused task types [1], we are interested in identifying task types from a high-level perspective across all user tasks. In a recently conducted field study [3], we observed that participants displayed significant differences in how they interacted with their web browser across different information seeking tasks and online transactions. We are looking to capitalize on these differences by automatically categorizing users' tasks according to a high-level schema. In this paper, we report on preliminary work using traces of users' web browser interactions, collected in a realistic setting, to automatically identify high-level tasks.

2. METHODOLOGY

The data was collected during a one week field study conducted with 21 participants. The goal of the study was to gain an understanding of the characteristics of information seeking tasks on the Web and how the features of current web browsers are being used to complete these tasks. Over the course of the study, participants were asked to use a custom web browser that logged all of their interactions with the web browser, including URLs visited, navigation mechanisms used (e.g., back button, bookmarks, clicked links, auto-complete), and use of browser functions (e.g., copy, paste, find).

Participants were also asked to provide a categorization and brief textual description for all their web usage over the course of the study. This could be done either in real time using a task toolbar embedded in the custom web browser or at the end of each day using a task diary. Participants categorized their web usage according to the following schema: *Fact Finding* (looking for specific information, such as an email address or a sports score), *Information Gathering* (collecting information, often from multiple sources to make a decision, write a report, etc.), *Browsing* (serendipitous web navigation for the purpose of entertainment or to see what is new), and *Transactions* (online actions such as email or banking). We also included the category *Other* for tasks which did not fit within the given categories.

3. DATA

Overall, participants recorded 1192 tasks involving 13,498 web pages over the week long study. Each task was associated with a number of data features logged by the web browser during the field study. The features included in our model are: task duration (calculated by summing the web page dwell time across the task), number of pages loaded, number of windows opened, type of web browser navigation mechanism used to initiate the task (e.g., bookmarks, clicked link, typed-in URL), use of browser functions (e.g., cut, copy, paste, save), and use of Google.

4. RESULTS

4.1 Task Prediction

We observed strong differences in how participants interacted with their web browsers during different tasks and were interested in whether logged interactions could be used to predict the category of task, from a high-level perspective. Using decision trees, we constructed several preliminary models using the Weka C4.5 decision tree package. All models were trained using 10-fold cross validation. Only a small number of tasks were labelled as Other (1.7% - 20/1192); therefore, we did not include this data in the task prediction.

An initial decision tree was built using 1172 tasks from all 21 participants. This aggregate model correctly classified 53.8% of the task instances. The results of this classification suggest that an aggregate model is not appropriate and we suspect it is because user behaviour on the Web is simply too individual. For instance, we observed that the use of web browser navigation mechanisms was highly dependant upon the individual [3]. We also observed that there was a high degree of variability across the participants for measures such as dwell time, number of pages viewed, and number of windows loaded.

In response to these results, we constructed individual models for each participant who had logged more than 30 tasks (17/21 participants). The percentage of positive classification is shown in Figure 1 for each participant. The number of correctly classified tasks by participant ranged from a high of 94.3% to a low of 43.6%. We expect that more accurate models could be produced with a larger data set, as well as more sophisticated feature selection and machine learning techniques.

4.2 Task Initiation

There are several aspects of a user's task that an automatic task categorization system must detect, such as task initiation, task type, task switching, and task termination. We observed a strong identifier of new task initiation through the use of a particular set of web browser navigation mechanisms. This consists of New Task Session (NTS) navigation mechanisms [3], which were typically employed by participants either when initiating a new task session or when changing navigation strategies within a session. The NTS navigation mechanisms include the auto-complete function, bookmarks, the Google toolbar, the home

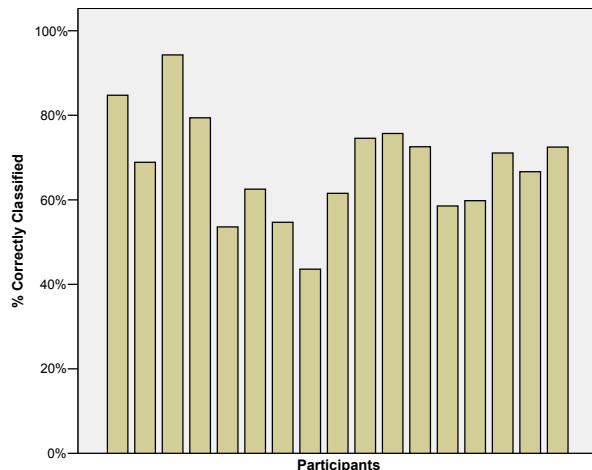


Figure 1. The percentage of correctly classified tasks, by participant

button, selecting a URL from the drop-down address bar, and typed-in URLs. We observed 1314 instances of NTS navigation mechanisms across all participants and tasks. Of those instances, 70.3% (924/1314) occurred when participants were initiating a new task. This suggests that the use of the web browser navigation mechanisms may play a useful role in detecting the beginning of new task sessions.

5. CONCLUSIONS

We have presented initial work in automatically classifying web-based tasks based on user interactions with their web browser. The data used in our predictive models represents a realistic picture of how users interact on the Web and was collected over the course of a week long field study. Moreover, the task categorization represents all high-level tasks in which users engage on the Web, not simply a subset of randomly chosen tasks. The results of this work suggest that more accurate task prediction is possible when individual models are used, in comparison with aggregate models of user behaviour.

As part of our future work, we would like to further explore machine learning techniques to improve the accuracy of the task predictions. As well, we would like to incorporate a categorization of the web pages visited, potentially according to web page genre.

6. ACKNOWLEDGMENTS

This research was funded by the Natural Science and Engineering Research Council of Canada (NSERC).

7. REFERENCES

- [1] Iqbal, S. T. and Bailey, B. P. (2004). Using Eye Gaze Patterns to Identify User Tasks. In Proceedings of the Grace Hopper Celebration of Women in Computing.
- [2] Kellar, M., Watters, C., Duffy, J. and Shepherd, M. (2004). Effect of Task on Time Spent Reading as an Implicit Measure of Interest. In Proceedings of ASIS&T 2004 Annual Meeting, Providence, RI, 168-175.
- [3] Kellar, M., Watters, C. and Shepherd, M. (2006). The Impact of Task on the Usage of Web Browser Navigation Mechanisms. In Proceedings of Graphics Interface 2006, Quebec City, Canada, To Appear.
- [4] Kelly, D. and Belkin, N. (2004). Display Time as Implicit Feedback: Understanding Task Effects. In Proceedings of SIGIR 2004, Sheffield, UK, 377-384.
- [5] Kelly, D. and Belkin, N. J. (2001). Reading Time, Scrolling and Interaction: Exploring Implicit Sources of User Preference for Relevance Feedback. In Proceedings of SIGIR 2001, New Orleans, LA, 408-409.
- [6] Lee, U., Liu, Z. and Cho, J. (2005). Automatic Identification of User Goals in Web Search. In Proceedings of WWW 2005, Chiba, Japan, 391-400.
- [7] Oard, D. W. and Kim, J. (2001). Modeling Information Content Using Observable Behavior. In Proceedings of ASIS&T 2001 Annual Meeting, Washington, DC.
- [8] Oliver, N., Smith, G., Thakkar, C. and Surendran, A. C. (2006). Swish: Semantic Analysis of Window Titles and Switching History. In Proceedings of IUI '06, Sydney, Australia, 194-201.