Reading Time, Scrolling and Interaction: Exploring Implicit Sources of User Preferences for Relevance Feedback During Interactive Information Retrieval

Diane Kelly and Nicholas J. Belkin
School of Communication, Information and Library Studies
Rutgers, The State University of New Jersey
New Brunswick, NJ 08901 USA
Email: dianekel@scils.rutgers.edu, nick@belkin.rutgers.edu

1. INTRODUCTION

Implicit feedback techniques appear to be attractive candidates to improve retrieval performance through relevance feedback without requiring more effort on the part of the user. Implicit feedback techniques gather data indirectly from the user by monitoring behaviors of the user during searching. If information about a document's relevance to a user's query can be gathered passively rather than actively, then users can experience the benefits of relevance feedback without having to expend any additional effort. Implicit feedback techniques have been primarily investigated in information filtering and recommendation systems [2, 3, 5, 7].

Behaviors most extensively investigated as sources for implicit feedback have been reading, saving and printing. For instance, [5] found that the major factor influencing the amount of time a user spends with a news article is the user's preference for that article. Specifically, [5] found that there is a strong tendency for users to spend a greater length of time reading those articles rated as interesting, as opposed to those rated as not interesting. This finding has been replicated by others in similar environments [4]. Other behaviors that have been explored include printing, saving, scrolling and bookmarking [6].

The work reported here is an explicit test of the work of [5] in an IR context other than information filtering. Three sources of implicit feedback are of particular interest: reading time per document, scrolling and interaction. The specific hypotheses for this study are, accordingly:

H1: Users will spend more time reading those documents that they find relevant.

H2: Users will scroll more often within those documents that they find relevant.

H2: Users will interact more with those documents that they find relevant.

2. METHOD

A secondary analysis of data was employed for the method for this study. The data for this study was extracted from the trace files generated during the Rutgers' TREC-8 Interactive Searching Study [1], which implemented relevance feedback techniques in two experimental IR systems.

2.1 Participants

A total of 36 volunteers, recruited from the Rutgers community, participated in the original project. Data from only the first 6 subjects are included in this report. Each subject conducted six searches. A total of 561 documents were opened by these 6 subjects. The instructions to the subjects were that for each search, they should find and save documents which identified the different instances or aspects of the specified topic.

2.2 Procedures

Several pieces of data were extracted from the trace files and analyzed. These data included time spent reading a document, scrolling and amount of interaction with a document. Time spent reading a document began when the user clicked on a document title to display its full text. The end time for reading a document was indicated by the user executing another action that signals he or she was finished reading the document, such as saving the document, displaying the text of another document, scrolling through the title summary window, running a new search or exiting the system. Scrolling was measured by the number of times the user clicked on the scroll bar. Interaction with a document encompassed the following activities: clicking on Show Next Keyword, Show Best Passage, Show Next Best Passage or Show Previous Passage buttons.

The relationship between time, scrolling and interaction and the user's relevance judgements was considered. For this study, saving a document was considered to be a positive relevance judgement and not saving a document was considered as a negative relevance judgement. The instructions for the experiment state that users should save those documents that identify positive instances of a particular information problem. Assuming that users want to perform well, construing a user's relevance judgement as either saving a document or not saving a document seems reasonable.

3. RESULTS

A total of 561 documents were viewed by the 6 subjects. Within the 561 documents, a small number of documents ($< 1\%$) were displayed multiple times per subject. Of these 561 documents, 240 (43%) were identified by users as relevant and 321 (57%) were identified as non-relevant.

On average, subjects spent 26.49 seconds with each document. The amount of time that users spent with relevant and non-relevant documents was similar (Relevant: $\underline{M}$=27.62, $\underline{SD}$=25.99; Non-relevant: $\underline{M}$=25.63, $\underline{SD}$=23.65). This difference was not significant, $\underline{t}$ (558)=.94, ns.

The frequency distributions of relevant and non-relevant documents and time spent viewing the document are displayed in Figure 1 and Figure 2, respectively. As can be observed from the histogram, users most often spent approximately 10 seconds viewing those documents that they eventually identified as relevant and also those that they eventually did not mark relevant. Indeed, the distribution curves are nearly identical. These numbers

provide quite a different, and more accurate, description of the data than the means reported in the proceeding paragraph. This is most likely the result of high standard deviations for each group.
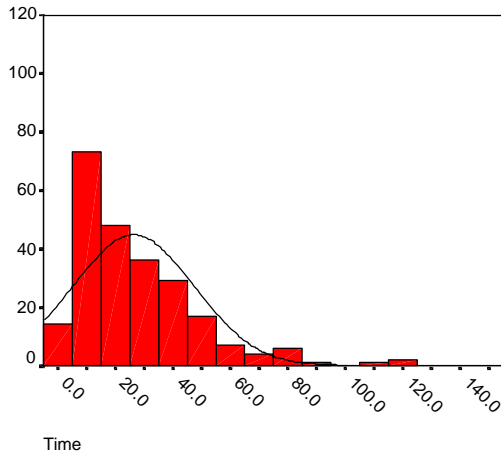

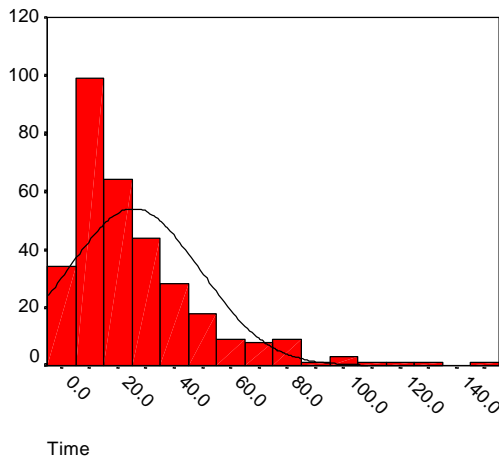Figure 1. Frequency Distributions of Time Spent Reading Relevant Documents.

Figure 2. Frequency Distributions of Time Spent Reading Non-relevant Documents.

Subjects scrolled an average of 4.28 times per document. There was no significant difference between the amount of scrolling users did in relevant documents ($\underline{M}$=4.41, $\underline{SD}$=12.70) and non-relevant documents ($\underline{M}$=4.20, $\underline{SD}$=9.92), $\underline{t}$ (558) = .23, ns.

Interaction was measured by the number of the times users clicked on the Show Next Keyword, Show Best Passage, Show Next and Show Prev buttons. Overall, subjects interacted with documents very little ($\underline{M}$=.20). There was no significant difference between the amount of interaction that users engaged with relevant documents ($\underline{M}$=.19, $\underline{SD}$=.48) and non-relevant documents ($\underline{M}$=.21, $\underline{SD}$=.56), $\underline{t}$ (558) = -.52, ns.

4. CONCLUSIONS

Previously, it had been found that users spend more time reading those documents that they find relevant than those that they do not find relevant [5]. The primary goal of the current study was to see if this finding could be replicated in another IR context. For the current study, overall, the length of time that a user spends viewing a document was not significantly related to the user's subsequent relevance judgement.

The results suggest that the original theory motivating the study may be limited in its scope. It appears that things like tasks, document collection and searching environment may affect the generalizability of the original findings more than was initially anticipated or specified. The non-significant results in the present study may be a result of the characteristics of the document collection, the searching environment and/or the experimental protocol. The users in [5] were only required to read incoming articles and assign scores to them. The users in the present study were engaged in a more complex task, where they were required to construct queries, evaluate, save and label document all within a specific time period. This may explain why viewing time for both relevant and non-relevant documents was so low. Users may have felt compelled to perform as quickly as possible because of the current experimental protocol. Task may have also affected the results. The tasks in [5] were those that naturally interested the user, since presumably the users originally subscribed to the news group because they found the general topic of discussion relevant and/or interesting. In the current study, the tasks were artificial and unfamiliar to the users. A relevant document could most likely be better distinguished in the former case.

At present, we are working to analyze the remainder of the data from the TREC-8 study. We are also working to identify and investigate other potentially useful sources of implicit relevance feedback in a more traditional IR context.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Belkin, N. J., Cool, C., Head, J., Jeng, J., Kelly, D., Lin, S., Park, S. Y., Savage-Knepshield, P., & Sikora, C. (2000). Relevance feedback versus local context analysis as term-suggestion devices: Rutgers' TREC-8 interactive track experience. In D. Harman, & E. Voorhees (Eds.), TREC-8, Proceedings of the Eighth Text Retrieval Conference. Washington, D. C.: NIST, 565-574.

[2] Hill, W. C., Hollan, J. D., Wroblewski, D., & McCandless, T. (1992). Edit wear and read wear. Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, 3-9.

[3] Kamba, T., Sakagami, H., & Koseki, Y. (1997). ANATAGONOMY: A personalized newspaper on the World Wide Web. International Journal of Human-Computer Studies, 46, 789-803.

[4] Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., & Riedl, J. (1997). GroupLens: Applying collaborative filtering to usenet news. Communications of the ACM, 40(7), 77-87.

[5] Morita, M., & Shinoda, Y. (1996). Information filtering based on user behavior analysis and best match text retrieval. Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 272-281.

[6] Oard, D. W., & Kim, J. (1998). Implicit feedback for recommender systems. Proceedings of the AAAI Workshop on Recommender Systems. http://www.glue.umd.edu/~oard/research.html.

[7] Seo, Y. W., & Zhang, B. T. (2000). Learning user's preferences by analyzing wed-browsing behaviors. Proceedings of the ACM 4th International Conference on Autonomous Agents, 381-387.