Study of the Usefulness of Known and New Implicit Indicators and Their Optimal Combination for Accurate Inference of Users Interests

Bracha Shapira
Ben-Gurion University
Department of Information Systems Engineering
Beer-Sheva, Israel
972-8-6477551
bshapira@bgu.ac.il

Meirav Taieb-Maimon
Ben-Gurion University
Department of Information Systems Engineering
Beer-Sheva, Israel
972-8-6479335
meiravta@bgu.ac.il

Anny Moskowitz
Ben-Gurion University
Department of Information Systems Engineering
Beer-Sheva, Israel
972-8-6477003
Ya_Anny@yahoo.com

ABSTRACT
Explicit relevance feedback involves explicit ratings of documents or terms by users and disrupts their browsing and searching. The alternative non-disruptive method is implicit feedback inferring users' needs and interests by monitoring their regular interaction with the system. Some implicit indicators of interest, such as reading time, have been investigated in previous studies and were found indicative to the relevance of documents but not sufficiently accurate [1,2,3,4]. In this paper we present and examine several new relative implicit feedback indicators, and study the effect of combining several implicit indicators. The paper describes a large-scale user study on which users' searches were observed by a specially developed browser that recorded their behavior (implicit indicators) as well as their explicit ratings. We analyzed the relationship between implicit indicators and explicit ratings and found that a certain combination of implicit indicators achieved higher correlation with the explicit ratings than any of the individual indicators. We have also found that the relative indicators are more indicative to the level of interest of a user item than the non-relative indicators.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval– *Relevance feedback.*

General Terms
Measurement, Experimentation, Human Factors.

Keywords
Implicit and explicit relevance feedback, User studies

1. INTRODUCTION

The current research has three main goals:

1) Re-examine the effectiveness of known implicit indicators.
2) Suggest and examine new relative indicators.
3) Investigate the effect of combining implicit indicators on the accuracy of the prediction of users' needs.

SAC'06, April, 23-27, 2006, Dijon, France.

Since former studies revealed that not all implicit indicators are equally useful [1,2,3,4], it might well be that some will only be useful when combined or related to others. In addition, we suggest relative indicators, i.e., indicators normalized by other factors. Our intuition was that the interest of users in a page might be better predicted when the indicators are normalized by properties of the data or by other implicit indicators. For example, reading time normalized by the length of the page being read might indicate better the interest of the user assuming that longer pages naturally require a longer time to read. The new suggested (single and relative) indicators are:

1) Mouse movement relative to reading time; - relative indicator
2) Scrolling time relative to reading time; - relative indicator
3) Reading time normalized by page size; - relative indicator
4) Number of links visited on a page; - single indicator
5) Number of links visited on a page relative to the number of existing links on the page; - relative indicator
6) Level of interaction on a page. Interaction includes one of the following actions: Print, bookmark, copy&paste, mouse movement, scroll, visit links. The maximal interaction level is six, if a user performed all actions, and the minimal is zero, if no actions were performed. The assumption is that the more interaction observed, the relevancy of the page for the user is higher (single indicator).

In addition, we have re-examined the following known single indicators:
Examination: Reading time, mouse movement, and scrolling.
Retention: bookmark, print, copy& paste.

2. EXPERIMENT
An experiment was conducted to examine the effect of the single, relative and combination of implicit indicators compared to explicit ratings. A specially developed browser was installed for four months in an Internet station (a PC dedicated for Internet searches) in a large high-tech firm. 25 computer engineers, workers of the company, were asked to use the browser when searching the Web for their professional needs. Their behavior on every page they accessed was recorded by the browser, as well as their explicit relevancy rating for the page. The rating ranged from 0-5 according to the following key:
0- No opinion; 1- Not interesting; 2- Somewhat interesting;
3- Helps navigation; 4. Interesting; 5- Very interesting
The following users' behaviors (implicit indicators) were gathered by the browser for each user on each access to a page:

URL, user ID, current time, time spent (seconds), number of links, number of visited links, Total text size (bytes), total time of mouse movement (milliseconds), and total scroll time (milliseconds). In addition the following three Boolean indicators (y/n) were collected: Print – indication of any print operations, Bookmark – indication of addition to the bookmark list, Copy/paste – indication of any copy/paste operations.

The users were also asked to explicitly rate the relevancy of each page they visited (on a 0-5 scale as detailed above).

1758 pages were collected from which 198 pages were removed since the participants had no opinion about their relevancy. Six pages were identified as outliers and removed since their reading time was higher than 500 seconds. 1554 pages were analyzed.

2.1 Evaluation Methods and Measures

For every implicit indicator we examined whether it correlates with explicit ratings. Our prediction was that there should be significant differences in the values of implicit indicators among the explicit rating groups (1-5). Once our prediction was verified, for each indicator, we used descriptive statistics to study the trend. The common parametrical statistical test, ANOVA, was not used since one of its assumptions – the equality of variances - was violated. We therefore applied the Kruskal-Wallis test for each implicit indicator using the group medians[1]. For each implicit interest indicator we performed a Kruskal-Wallis test on the implicit indicator as the dependent variable and the explicit rating as the independent variable. Only for those variables where the Kruskal-Wallis test rejected the null hypothesis (that the median values are the same), we performed a post-hoc multiple comparison using the Dunnet test [5] to evaluate the degree of difference (variation) between the groups. We later use the results of this test to rank the implicit indicators by their effectiveness.

The effectiveness of combining implicit indicators in predicting users' interests was tested using stepwise linear regression, a technique for screening candidate variables to obtain a regression model containing the "best" subset of regressor variables [5]. We were interested in choosing the set of "best" implicit indicators to be included in the regression model.

3. RESULTS

We generated box-plots graphs for the significant implicit indicators to show the relation between implicit indicators and explicit ratings. Due to space limitations we only present (figure 1) results for the "Mouse moving normalized by reading time" indicator, which was found to be the best indicator. The upper dashed line on the graph shows the average implicit indicator for each rating group. The boxes show the range of values from the bottom quartile (25%) to the top quartile (75%) and the lines (in the boxes) show the medians of the implicit indicator. For this indicator, after the null hypothesis was rejected by the Kruskal-Wallis test ($p<0.001$), the Dunnet post-hoc test showed significant differences between 80% of the groups. The Dunnet test identified four different levels of rating (1,2,3-4,5) i.e., this indicator distinguishes between 4 levels of interest.

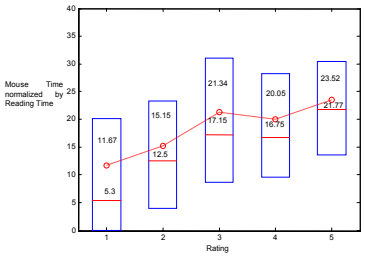

Figure 1 – Mouse Moving Time Normalized by Reading Time

Following is a summary of our findings in this research**:**

1) For the single non-relative indicators we showed positive correlation to explicit indicators. However, these indicators are not effective in distinguishing the level of interest in a page.

The number of visited links and the level of interaction in a page were found to be binary indicators, i.e., they distinguish between interesting and not interesting pages. Printing and bookmarking were found indicative only for interesting pages, but users do not perform them very often.

2) For the relative indicators we found the relative mouse movement to be the best single indicator, more accurate than reading time or mouse movement. It is able to distinguish between levels of interest and is the recommended indicator if no combination of indicators is applied.

The other relative indicators were not significantly better than the corresponding non-relative indicators. We believe that these findings point to the potential of the relative indicators to improve the accuracy of implicit indicators.

Another finding is the ability of the new indicator "number of visited links relative to the number of links on a page" to accurately indicate a navigational page.

3) Combination of indicators. The best correlation with explicit indicators was obtained when some indicators were combined. The optimal selection of the indicators to be included in the combination was performed by applying stepwise regression.

4. REFERENCES

[1] Claypool , M. Brown, D., Le, P., and Waseda, M. Inferring user interest . In Proceedings of the 6th Intl. Conference on Intelligent User Interfaces (IUI 01), USA, 33-40. 2001.

[2] Kelly, D., and Teevan, J. Implicit feedback for inferring user preferences, SIGIR Forum (37),2, 18-28. 2003

[3] Oard, D.W., and Kim., J., Implicit feedback for recommender systems. In proceedings of the AAAI Workshop on Recommender Systems. 1998

[4] Oard, D.W., and Kim, J. Modeling information content using observable behavior. In proceeding of the 64th Annual meeting of the American Society for Information Science and Technology, USA, 38-45. 2001.

[5] Montgomery, D.C. Design and Analysis of Experiments. Wiley & Sons, Inc. (ed). 2001.

[1] We did not use means that are known to be affected by outliers, since most indicators had a heavy-tailed distribution (i.e., outliers).