

Estimation of User Interest in Visited Web Page

Michal Holub

Institute of Informatics and Software Engineering
Faculty of Informatics and Information
Technologies, Slovak University of Technology
Ilkovičova 3, 842 16 Bratislava, Slovakia

miso.holub@gmail.com

Maria Bielikova

Institute of Informatics and Software Engineering
Faculty of Informatics and Information
Technologies, Slovak University of Technology
Ilkovičova 3, 842 16 Bratislava, Slovakia

bielik@fiit.stuba.sk

ABSTRACT

Nowadays web portals contain large amount of information that is meant for various visitors or groups of visitors. To effectively navigate within the content the website needs to “know” its users in order to provide personalized content to them. We propose a method for automatic estimation of the user’s interest in a web page he visits. This estimation is used for the recommendation of web portal pages (through presenting adaptive links) that the user might like. We conducted series of experiments in the domain of our faculty web portal to evaluate proposed approach.

Categories and Subject Descriptors

H.5.4 [Information Interfaces and Presentation (e.g., HCI)]: Hypertext/Hypermedia—*Navigation*. H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—*Relevance Feedback*.

General Terms

Algorithms, Design, Experimentation.

Keywords

Adaptive navigation support, user behavior, user interest estimation, link recommendation

1. INTRODUCTION

Web portals are being visited by various users pursuing different goals. However, most websites offer all visitors the same content. Therefore, the visitors are often presented information in which they have no interest [1]. Another problem is inappropriate navigation that confuses users. They have difficulties to decide which link from the large amount of possibilities they should follow. While navigating, some users can discover interesting information. We believe that the link to the web page with such interesting information can also concern other users with similar goals, so using social recommendation can help in this context.

Often user’s interests are determined based on the content of documents the user has read [3]. Interests are then included in the user model which can be expressed by concepts (or just keywords) extracted from these documents. If we know what topics (expressed by keywords) the user prefers, we can recommend him documents (web pages) with similar content.

There are several ways how to get implicit feedback and use it in estimation of user’s interests. In areas where links are well

annotated (like news portals where links to articles are followed by a short introduction) the event of user clicking on a link can be considered as positive interest [4]. However, in general we cannot predict user’s interest in the following web page solely on the fact that the user clicked on the link pointing to this page. Another approach is to track the actions user performs while reading a web page. Actions like printing the page or adding it to bookmarks show positive interest. On the other hand, spending very small amount of time reading it or even closing the browser while the page is being loaded show negative interest [6].

For estimation of user’s interest we propose a method of tracking his behavior when visiting a particular web page. With this data we are looking for users who behave similarly and recommend them links based on estimated interest and collaborative filtering.

2. DETERMINING USER’S INTEREST

We combine behavioral analysis for deriving user’s interest in a web page he currently visits with collaborative filtering technique as described in [5]. We use collaborative filtering for predicting user’s interest in a web page he has not yet visited. For our purpose the items are web pages and the rating of an item is an estimate of user’s interest based on his behavior.

To determine user’s interest we observe actions he makes on a web page. These include *time spent on a web page*, *number of scrolling events that occur* and *number of times he copies text into the clipboard*. Our method is based on comparison of current user’s behavior with behavior of others. We compare the values of first two actions with values from other people who visited the same page. If the value for current user is more than X % higher than the average we consider it as a sign of positive interest in the page. On the other hand, when it is more than X % lower than the average we consider it as a sign of negative interest. When the value is around average ($\pm X$ %) it is a sign of neutral interest. Experiments show that optimal value of X is in range of 20-30.

We estimate the actual value of user’s interest in each page he visits according to Figure 1, where the x axis represents time spent on a web page, the y axis represents the number of scrolls done. Symbol \uparrow means higher than average value, symbol \downarrow means lower than average value, and symbol $-$ means average value. The time spent and the number of scrolls made has the same weight in the final score. We increase this value by 0.1 when the user also copied text into clipboard; otherwise we decrease it by 0.1. Resulting aggregated interest is in the interval $\langle 0,1 \rangle$ with 0 meaning no interest and 1 meaning total interest in the visited web page. For the value of spent time we set an inactivity threshold to be 4 times the average time spent on that page by others.

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

WWW 2010, April 26–30, 2010, Raleigh, North Carolina, USA.
ACM 978-1-60558-799-8/10/04.

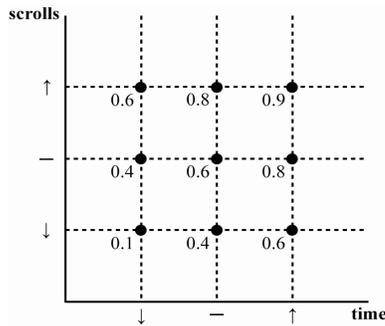


Figure 1. Estimation of user's interest in visited web page.

3. LINK RECOMMENDATION

We recommend links by predicting user's interest in yet unseen pages using collaborative filtering method. We compute the values of Pearson correlation coefficient between the user to whom we want to recommend a web page and all other users [5]:

$$S_{a,u} = \frac{\sum_{i=1}^I (r_{a,i} - r_a) \times (r_{u,i} - r_u)}{\sqrt{\sum_{i=1}^I (r_{a,i} - r_a)^2 \times \sum_{i=1}^I (r_{u,i} - r_u)^2}}$$

In the domain of web pages recommendation $r_{a,i}$ means the interest of user a in page i , r_a means average interest of user a and I is the total number of pages visited by user.

We use this value to predict user's interest in a page which he has not yet visited but which other users have. Pearson correlation coefficient says how similar two users are considering their ratings of items. In our particular case this rating is represented by behavior of users on every page they both visit. We compute the predicted value of interest like this [5]:

$$p_{a,i} = r_a + \frac{\sum_{u=1}^N (r_{u,i} - r_u) \times S_{a,u}}{\sum_{u=1}^N S_{a,u}}$$

Here $p_{a,i}$ means prediction of interest of user a in page i , $S_{a,u}$ means the similarity of users a and u (value of their Pearson correlation coefficient) and N is the number of similar users.

4. EVALUATION AND CONCLUSIONS

To evaluate proposed method we developed a software tool that supports adaptive navigation for guests of particular web portal. It enhances each original web page by adding links that represent interesting part of the portal. We experimented with web portal of our faculty (www.fiit.stuba.sk) by adding the recommendations to the right navigational menu.

We designed client-server architecture with an adaptive proxy server developed in our group in the middle. Adaptive proxy server is a platform for undisturbed involving of methods and techniques for the adaptation of the content and navigation on the web [2]. It enables developers to control the adaptation process by means of services. On the client side there is a behavior tracking script. It invokes a web service on the server side which collects data about user behavior. Another component on the server is responsible for computing user's interest and making predictions for unseen pages. It then selects the links to be recommended.

We realized a plug-in to the adaptive proxy server, which handles client requests and server responses. It can be extended to conduct

various methods of web adaptation. Our plug-in inserts the tracing script to each page together with recommended links. Many web pages on our faculty portal inform about an event. Our tool extracts dates from these pages. We use our interest estimation method to construct a personalized calendar of events. If the interest is positive, we add the event to user's calendar and show it on the page. We can also recommend events between users.

We did a series of experiments on our faculty website. In the experiments the visitors of our website were asked to express their interest in visited web page as an integer ranging from 0 to 10 with 0 meaning no interest and 10 meaning total interest. During the visit of each page interest was estimated by proposed method.

Our results indicate that the most important quantity that determines user's interest is the time spent on a webpage. The more time users had spent on a web page above average the higher they rated it. This is the same result as our method gives.

Scrolling also proved to indicate positive interest in the web page. However, the results showed that when a user does not use scrolling it does not always mean his lack of interest. Thus we should give higher weight to the value of time spent and less weight to number of scrolling events. Copying text into clipboard proved to be a clear sign of interest. Nevertheless, only few people actually used it during our experiments.

We have proposed an approach to automatic estimation of user's interest in a web page. We are able to predict his interest for unvisited pages and recommend him interesting links. In the future work we plan to use this estimation in social adaptive navigation support that employs groups of users determined by observing their paths of navigation through the whole web portal.

Acknowledgements. This work was partially supported by the Scientific Grant Agency of SR, grant VG1/0508/09 and it is partial result of the Research & Development Operational Programme for the project Smart Technologies, Systems and Services, ITMS 26240120005, co-funded by the ERDF.

5. REFERENCES

- [1] Barla, M., Tvarožek, M., and Bieliková, M. Rule-based user characteristics acquisition from logs with semantics for personalized web-based systems. *Computing and Informatics*, Vol. 28, No. 4 (2009), 399-427.
- [2] Barla, M., and Bieliková, M. „Wild web“ personalization: adaptive proxy server. In *Proc. of the 4th Workshop on Intelligent and Knowledge oriented Technologies (Herľany, Slovakia, 2009)*, 48-51 (in Slovak).
- [3] Brusilovsky, P. Methods and techniques of adaptive hypermedia. *User Modeling and User-Adapted Interaction*, Vol. 6, No. 2-3 (1996), Springer Netherlands, 87-129.
- [4] Das, A.S., et al. Google news personalization: scalable online collaborative filtering. In *Proc. of the 16th Int. Conf. on WWW (Banff, Canada, 2007)*, ACM Press, 271-280.
- [5] Sugiyama, K., Hatano, K., and Yoshikawa, M. Adaptive web search based on user profile constructed without any effort from users. In *Proc. of the 13th Int. Conf. on WWW (New York, NY, USA, 2004)*, ACM Press, 675-684.
- [6] Velayathan, G., and Yamada, S. Behavior-based web page evaluation. In *Proc. of the 15th Int. Conf. on WWW (Edinburgh, Scotland, 2006)*, ACM Press, 841-842