Capturing User’s Interest from Human-Computer Interaction Logging

Vincenzo Deufemia, Massimiliano Giordano, Giuseppe Polese, and Genoveffa Tortora

Università di Salerno, Via Ponte don Melillo, Fisciano(SA), Italy
{deufemia,mgiordano,gpolese,tortora}@unisa.it

Abstract. Mining user’s expectations and interests has become the focus of many Internet-based application providers, such as those operating in the areas of social networks, search engines, e-commerce, and so forth. This is often accomplished by means of explicit feedbacks requested to end-users, which might yield distorted results due to the intrusive nature of this kind of approach. Thus, it would be desirable using implicit feedbacks, provide that they faithfully reflect user’s habits and expectations. In this paper we propose an approach to capture user’s feedbacks from their interaction actions while processing a document, with particular emphasis on web documents. To this end, we propose a new model to interpret mouse cursor actions, such as scrolling, movement, text selection, while reading web documents, aiming to infer a relevance value indicating how the user found the document useful for his/her purposes. We have implemented the proposed model through light-weight components, which can be easily installed within major web browsers as a plug-in. The components log mouse cursor actions that we have used as experimental data in order to validate the proposed model. The experimental results show that the proposed model is able to predict user feedbacks with an acceptable level of accuracy.

1 Introduction

Nowadays, the goal of many internet application providers is to collect feedbacks from their users. Thus many modern internet applications frequently inquire their users asking them to provide feedbacks on the quality of provided services, or their trends and expectations to help providers meet their needs.

However, this has often resulted in a continuous spoiling of users daily work, which can potentially affect the accuracy by which they provide the requested feedbacks. This is particularly true in social networks or in search engines. Thus, many researchers have started devising methods for capturing users feedbacks in a non intrusive way. For instance, several years ago many internet search engines have started analysing the linked structure of the web in order to implicitly derive an index of the usefulness of web documents. Page-rank is the most famous method using link structure analysis [1]. The idea behind Page-rank algorithm is to exploit the macro-scale link structure among pages in order to implicitly derive an index of the usefulness of web documents. Page-rank is the most famous method using link structure analysis [1]. The idea behind Page-rank algorithm is to exploit the macro-scale link structure among pages in order to implicitly derive an index of the usefulness of web documents. Page-rank is the most famous method using link structure analysis [1]. The idea behind Page-rank algorithm is to exploit the macro-scale link structure among pages in order to implicitly derive an index of the usefulness of web documents. Page-rank is the most famous method using link structure analysis [1]. The idea behind Page-rank algorithm is to exploit the macro-scale link structure among pages in order to implicitly derive an index of the usefulness of web documents. Page-rank is the most famous method using link structure analysis [1]. The idea behind Page-rank algorithm is to exploit the macro-scale link structure among pages in order to implicitly derive an index of the usefulness of web documents.

For example, they are vulnerable to spamming, and links may have several meanings or purposes.

With the advent of Web 2.0, social bookmarking systems have started calculating the popularity of a Web document as the total number of times it has been bookmarked, which is interpreted as the number of users voting for the page. In the context of web search engines, explicit ranking systems are more dynamic than Page-rank, and often ensure shorter time for pages to reach their popularity peaks [3]. However, due to the insufficient amount of bookmarked pages, it is not advisable relying on explicit ranking systems alone. Furthermore, explicit ranking is subjective, since users need to explicit vote a web content to rate it, and not all the web users are keen on voting each site they visit. Thus, despite the rapid growth in the number of bookmarked pages, the combination of link structure-based and social bookmarking-based page ranking measures seems to be currently an optimal strategy.

Alternatively, methods that are able to implicitly capture user interests are potentially more useful, since there is no noise in the ranking process introduced by subjective evaluations [4, 5]. Thus, we have started exploiting methods for logging user interaction actions in order to derive an implicit index expressing the web page usefulness with respect to user interests. In particular, we propose a new model to interpret mouse cursor actions, such as scrolling, movement, text selection, while reading web documents, aiming to infer a relevance value indicating how the user found the document useful for his/her purposes [6, 7].

We have embedded the proposed model in a ranking system for the web. In particular, we have implemented the YAR (Yet Another Ranker) system, which re-ranks the web pages retrieved by a search engine based on the relevance values computed from the interaction actions of previous visitors. YAR has been implemented by means of lightweight components, which can be easily installed within major web browsers as a plug-in (we used it experimentally with Google, but any other search engine could be easily adapted). The implemented components capture mouse cursor actions without spoiling user browsing activities, which enabled us to easily collect experimental data to validate the proposed model. The experimental results demonstrate that the proposed model is able to predict user feedbacks with an acceptable level of accuracy.

The paper is organized as follows. Section 2 describes the metrics for deriving the web page relevance from mouse tracking logging data. An implementation of the proposed metrics in the context of ranking systems is presented in Section 3. Section 4 presents an experimental evaluation with analysis of the results. A comparison with related work is described in Section 5. Finally, conclusions and future work are discussed in Section 6.

2 The Metrics for Web Page Relevance

In order to compute the web page relevance value we consider several metrics. The application of all these metrics will be used to produce a value between 1 and 5, as usually done in social bookmarking systems. In particular, we have defined the following metrics: