Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2014) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

6th International conference on Intelligent Human Computer Interaction, IHCI 2014

Implicit feedback through user-system interactions for defining user models in personalized search

Gabriella Pasi

Università degli Studi di Milano Bicocca, Viale Sarca 336, Milano, Italy

Abstract

Personalized search is aimed at tailoring the search outcome to the user context. A user profile represents the user interests and preferences that can be captured either explicitly or implicitly. Since a long time it has been advocated that Information Retrieval is an interactive activity, and various types of user-system interactions are more and more considered and used to improve search. In particular implicit feedback techniques are applied to the purpose of collecting user interests and preferences via user-system interactions. In this paper a short synthesis of the implicit techniques used to define user models is presented.

Peer-review under responsibility of the Scientific Committee of IHCI 2014.

Keywords: personalized search, user modeling, implicit feedback, user-system interactions.

1. Introduction

Since a long time Information Retrieval is considered as a complex interactive activity that engages the user of a search engine in the formulation of a query, and which usually requires a user-system interaction aimed both to analyze the proposed search results and to iteratively reformulate the query to the aim of locating in the sought repository those information objects that fulfill the user's expectations. Such a complex task cannot be (and it is usually not) reduced to a single user-system interaction activated by a user query and that produces a list of items that the system has evaluated relevant to that query. The "blind" and closed behavior of the first generation of search engines, which fully relied on a query and on topical relevance, has been overcome by subsequent generations of search engines, designed on the basis of several analysis and findings. First of all, a deep study of the notion of relevance and its multi-dimensional nature has motivated a shift from topical relevance to the assessment of multiple relevance dimensions[1]. Second, search depends on several contextual variables and situations, such as the task for which the search is undertaken, the characteristics and the cultural and personal background of users, the query context, the awareness of the properties and characteristics of the information objects and their nature (which make

1877-0509
Peer-review under responsibility of the Scientific Committee of IHCI 2014.

them more or less suited to the purpose for which they have been retrieved)[2]. Last but not least, the interactive nature of the search task, as argued by Interactive Information Retrieval[3], has pointed out the importance of user-system interactions in the process of relevance assessment.

This path has brought to the evolution from systems fully unaware of the motivations and needs behind a query (the so called system-centered approach that is driven by a query as an object independent from a user) to systems that increasingly have made use of various kinds of knowledge around a query. Such knowledge can be related to the user, to the search domain, to the sought objects and to the query itself (the user-centered approach and beyond); it constitutes a crucial source of evidence of the needs behind a query, and has given rise to personalized and contextual approaches to search. This has implied a shift from a system-centered relevance assessment to a user and context centered relevance assessment.

An important implication of this transition is constituted by the attempts to also make a shift in the theoretical foundations of Information Retrieval: from the probability ranking principle to the interactive probability ranking principle[4], from IR models centered on topical relevance to models centered on the user and on the notion of context[3]. This has placed a great emphasis both on the central role of user-system interactions to capture the user interests and the search intentions, as well as on user centered evaluations.

Relevance feedback has been one of the first manifestations of the notion of context explicitly inserted in a paradigm of interactive search, where the user explicitly and proactively interacts with the system to select the retrieved information items that he/she judges relevant to her/his needs[5]. With this contextual and explicit user indication of relevance the system applies a query reformulation to better capture and express the user information needs. Since then several efforts have been made to capture the signals of the user interests and needs; in particular, this has been done by defining implicit feedback techniques, where the user feedback on possibly relevant information about his/her interests is captured by various user-system interactions, in a way that is transparent to the user[6,7,8]. The shift from explicit to implicit techniques was motivated from the need of not involving users in explicit relevance assessments; with implicit techniques the user-system interactions are instead exploited to guess/learn the users interests. Implicit feedback techniques are widely used in personalized search to the aim of collecting some elements of the context around the user and the query for improving the search outcome[9,10,11]. In the next section the task of personalized search is synthetically introduced, and in section 3 we give a very short introduction to implicit feedback techniques used to the purpose of user profiling in personalized search.

2. Personalized search

In 2002, in the report “Challenges in Information Retrieval and Language Modelling: Report of a Workshop held at the Center for Intelligent Information Retrieval”, personalized and contextual search were indicated as big challenges of IR[12]. After a decade, in 2012 a new report has appeared, where the issue of user centered approaches to information access is still outlined, and the need for methods for dealing with rich interaction sequences is emphasized[13].

In recent years, a considerable amount of research has addressed the problem of personalizing search, to the aim of taking into account the user context in the process of assessing relevance of documents to user’s queries[14,15,16]. Personalize search means injecting user preferences in the retrieval process. For example, if a vegetarian user formulates a query such as “good restaurant in Milano”, the expected results should take the user preference into account. To this aim the query evaluation should make explicit use of this information as an additional constraint (besides the query) to estimate document relevance. The quality of the search outcome strongly depends on information beyond the one expressed in a user’s query. So the effectiveness of the system strongly depends on the available quantity and quality of information about the user and its preferences. The more the user model is accurate the more the personalized answer can be effective.

In order to implement a personalized IR strategy two main activities must be undertaken. The first activity is aimed at the definition of the user model. The second activity is aimed at defining processes that, based on both the knowledge represented in the user model and the user query, are finalized to produce a user tailored search outcome; three main categories of approaches have been proposed to enhance search by exploiting the information in the user model: query modification, results re-ranking, and definition of approaches to modify/define relevance assessment[15].

In the following we focus on the former activity, and in particular on the use of implicit feedback collected by user-system interactions for defining user models in personalized search.

The definition of a user model encompasses three main phases: acquisition, formal representation and updating. The acquisition phase is aimed at the identification of the knowledge that characterizes the user and her/his interests and preferences. The phase of formal representation is aimed at the choice of a formal language to represent this knowledge; this is needed to make it possible that this information be accessed and used by the search engine. The updating phase is finalized at learning the changes of the user preferences on time.

A large amount of research has addressed the first problem; i.e. how to identify information useful to characterize the user. The user model is also called user profile, and the definition of a user profile is a crucial issue, due to the fact the quality and reliability of the user model strongly affects the effectiveness of the search personalization.

In section 3 we synthetically report on the phase related to the acquisition process.

3. User system interactions as implicit feedback for user modeling

To capture the user's interests two main techniques may be employed: explicit and implicit[7,11,15]. By the explicit approach the user is asked to be proactive and to directly communicate to the system her/his interests and preferences. This can be done through the compilation of questionnaires, by providing short textual descriptions (to specify topical preferences), and/or by providing a few documents that represent well the user interests and that will be processed by the system to automatically extract their main descriptors. However, an explicit request of information to the user implies to burden the user, and to rely on the user's willingness to specify the required information. This is generally not realistic. To overcome this problem, several techniques have been proposed in the literature to automatically capture the user's interests by implicit feedback techniques; this is done by monitoring the user's actions in the user system interaction, and by inferring from them the user's preferences and interests. The collected signals and potential preferences can be used to define predictive models.

Various kind of evidence can be collected through implicit feedback, where various kind of human system interactions are considered to the aim of implicit user modelling[6,7,8]. The proposed techniques are based on the collection and the analysis of: browser history, query history, click-trough data, desktop information, document display time, bookmarks, and several other kinds of signals[17,18,19,20,21].

The advantage of defining a user model based on techniques that consider implicit feedback from user system interactions is that several signals may be considered; the main disadvantage is that automatic processes may be error-prone, as they may introduce noise in the process of identifying the useful information. In the literature a distinction has been made between direct and indirect evidence of users' interests through interaction[3]. Actions like bookmarking or saving a document, and click-through actions can be considered as evidence of the user interest on some objects; while activities like eye tracking, scrolling and reading time are indirect indicators of users' interests. A combination of implicit feedback techniques is generally used to infer users interests, and it has been proved to be more reliable that the consideration of a single source of evidence[8,10,17].

An important issue in user modelling is the definition of long and short term profiles; the former are finalized to represent lasting user interests, and information like query history, click-through data, and desktop search activities, as well as desktop information have been fruitfully exploited to the purpose of defining long term user models[17,22,23]. Short-term profiling is very useful to capture the "immediate" user's interests, i.e. those interests that are related to activities (user-system interactions) around the current search session, which may be unrelated to a long-term profile in which usual interests are represented. Often, a user composes an initial query, views the returned documents, and if unsatisfied, she modifies the query and repeats the search process. In the literature, queries and click-through data in the short-term search history have been for example considered as an implicit feedback injected in the short term user model[18,19].

A more recent category of approaches to the definition of user profiles by implicit signals of relevance is via the analysis of users interactions in social networking services and via user generated contents in social media. Comments, blogs, ratings, social tagging activities, mentions and retweets have been usefully exploited in the profile construction[24,25,26]. The rationale behind the importance of considering user generated contents as well as the interactions among users in social networking services is that the preferences of people having similar interests to a given user may contribute to provide a good prediction of the preferences of the user[24].

Among the sources of evidence of users interests that may be collected by implicit feedback techniques, recently in the literature various approaches have appeared that consider physiological and affective evidence, such as brain wave tracing via EEG and facial expressions[28,29]. Last but not least, another important trend of research is related to the identification of the search task in contextual and personalized search; recently, task-relevant search behavior has been considered to the purpose of task-based personalization and task based groupization[30].

References

1. Saracevic T. Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: Nature and manifestations of relevance, *Journal of the American Society for Information Science and Technology* 2007; **58**:1915–1933.
2. Ingwersen P, Jarvelin K. *The Turn: Integration of Information Seeking and Retrieval in Context*. Dordrecht, The Netherlands: Springer; 2005.
3. Ruthven I. Interactive information retrieval, *Annual Review of Information Science and Technology* 2008; 42:43–91.
4. Fuhr N. A probability ranking principle for interactive information retrieval, *Information Retrieval* 2008. **11**(3):251-265
5. Salton G, Buckley C. Improving retrieval performance by relevance feedback. *Journal of the American Society for Information Science and Technology* 1990. 41(4):288-297.
6. Claypool M, Brown D, Le P, Waseda M. Inferring user interest, *IEEE Intern.Comput,* 2001*;* 32–39.
7. Kelly D, Teevan J. Implicit feedback for inferring user preference: A bibliography, *SIGIR Forum 37* 2003; 2:18–28.
8. Joachims T, Granka L, Pang B, Hembrooke H, Gay G. Accurately interpreting clickthrough data as implicit feedback, in *Proc. of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2005, pp. 154–161.
9. Speretta M, Gauch S. Personalized Search Based on User Search Histories. in *Proc. of the Inter. Conf. on Web Intelligence* 2005, 622-628.
10. Fox S, Karnawat K, Mydland M, Dumais S, White T. Evaluating implicit measures to improve Web search. *ACM Transactions on Information Systems*, 2005; 23:147–168.
11. Kelly D, Belkin N. A user modeling system for personalized interaction and tailored retrieval in interactive IR. Proc. of the American Society for Information Science and Technology, ASIST 2002; 39(1): 316–325.
12. Allan J, et al. Challenges in Information Retrieval and Language Modelling: Report of a Workshop held at the Center for Intelligent Information Retrieval University of Massachusetts Amherst, September 2002, *Sigir Forum* 2003; 34(1):31-47.
13. Allan J, Croft B, Moffat A, Sanderson M. Frontiers, challenges, and opportunities for information retrieval: Report from SWIRL 2012 the second strategic workshop on information retrieval in Lorne. SIGIR Forum, May 2012; 46(1): 2-32.
14. Pitkow J, Schutze H, Cass T, Cooley R, Turnbull D, Edmonds A, Adar E, Breuel T. Personalized search. *Communications of the ACM*, 2002; 45(9): 50–55.
15. Micarelli A, Gasparetti F, Sciarrone F, Gauch S. Personalized Search on the World Wide Web. *LNCS*, 2007; 4321: 195–230.
16. Pasi, G. Issues in Personalizing Information Retrieval. *IEEE Intelligent Informatics Bulletin* (IIB) December 2010; 3–6.
17. Teevan J, Dumais S, Horvitz E. Personalizing search via automated analysis of interests and activities. in Proc. of the *International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2005, pp. 449–456.
18. Shen X, Tan B, and Zhai C.X. Context sensitive information retrieval using implicit feedback, in *Proc. of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2005; 43-50.
19. Shen X, Tan B, and Zhai C.X. Implicit user modeling for personalized search. in *Proc. of the International Conference on Information and Knowledge Management,* CIKM 2005, pp. 824–831.
20. Agichtein E, Brill E, Dumais S, Ragno R. Learning user interaction models for predicting Web search preferences, in Proceedings of theAnnual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2006; 3–10.
21. Gauch S, Speretta M, Chandramouli A, Micarelli A. User Profiles for Personalized Information Access, The Adaptive Web 2007, pp. 54-89
22. Chirita P, Firan P, Nejdl W. Summarizing local context to personalize global Web search. in Proc. of the Annual International Conference on Information and Knowledge Management, CIKM 2006; 287–296.
23. Tan B, Shen X, and Zhai C. Mining long-term search history to improve search accuracy. In *Proc. of the 12th ACM SIGKDD international conference* (KDD 2006). 718-723.
24 Carmel D, Zwerdling N, Guy I, Ofek-Koifman S, Har'el N, Ronen I, Uziel E, Yogev S, Chernov S. Personalized social search based on the user's social network. *Proceeding of the 18th ACM conference on Information and knowledge management* 2009; 1227-1236.
25. Xu, S, Bao S, Fei B, Su Z, Yu Y. Exploring folksonomy for personalized search. *Proc. of the 31st international ACM SIGIR conference on Research and development in information retrieval*. 2008; 155-162.
26. Abel F. Contextualization, User Modeling and Personalization in the Social Web. PhD Thesis, 2011. Gottfried Wilhelm Leibniz University Hannover.
27. Younus A, O'Riordan C, Pasi G. A Language Modeling Approach to Personalized Search Based on Users' Microblog Behavior. ECIR 2014: 727-732.
28. Arapakis I, Athanasakos K, Jose JM. A comparison of general vs personalised affective models for the prediction of topical relevance. In *Proc. of the 33rd international ACM SIGIR conference on Research and development in information retrieval* (SIGIR 2010); 371-378.
29. Eugster MJA, Ruotsalo T, Spapé MS, Kosunen I, Barral O, Ravaja N, Jacucci G, Kaski S. Predicting term-relevance from brain signals. In *Proc. of the 37th international ACM SIGIR conference on Research & development in information retrieval* (SIGIR '14). 2014, 425-434.
30. White RW, Chu W, Hassan A, He X, Song Y, Wang H. Enhancing personalized search by mining and modeling task behavior. In *Proc. of the 22nd international conference on World Wide Web*, 2013; 1411-1420.