



An iterative method for personalized results adaptation in cross-language search

Dong Zhou^{a,*}, Wenyu Zhao^a, Xuan Wu^a, Séamus Lawless^b, Jianxun Liu^a

^aSchool of Computer Science and Engineering & Key Laboratory of Knowledge Processing and Networked Manufacturing, Hunan University of Science and Technology, Xiangtan, Hunan 411201, China

^bADAPT Centre, Knowledge and Data Engineering Group, School of Computer Science and Statistics, Trinity College Dublin, Dublin 2, Ireland

ARTICLE INFO

Article history:

Received 16 September 2016

Revised 13 October 2017

Accepted 22 November 2017

Available online 22 November 2017

Keywords:

Personalized cross-language information retrieval
Results adaptation
User model
Iterative method

ABSTRACT

On today's Web, people often desire to not only retrieve results which are of relevance to their query, but for those results to be of particular relevance to them as an individual. In most personalized search systems, the scores obtained from different rankers are linearly combined to provide the personalized ranked list. Moreover, when compared to the personalization research in monolingual web search, relatively few studies extend to the cross-language domain. In this paper we investigate the personalized results adaptation problem in the context of cross-language web search. The main contribution of this research is a novel iterative ranking method based on document associations obtained from an initial ranker. The method assumes that results retrieved by non-personalized rankers and personalized rankers mutually reinforce each other, rather than being used in linear combination. The method is applied in a personalized cross-language search scenario on a semi-automatically constructed test collection and a real-world dataset. The experimental results suggest that the proposed personalized result adaptation method can produce better results than previous approaches for cross-language web search. The results also prove that the semi-automatically constructed test collection can be used as an alternative dataset for evaluation in the absence of available real-world datasets.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

In recent years personalized search systems have been extensively studied in the literature [30,32]. Such systems do not retrieve results that are solely relevant to the query, but rather, results that are also relevant to the individual user's interests, preferences, needs etc. These systems generally pass through three main steps in order to provide their personalized service [17]: (1) information gathering, using tools to collect information about users and their usage history; (2) information representation of user context, which is usually stored as the user model [33]; and (3) personalization implementation, which usually takes the form of different approaches to adapt a user's query or results [12,16]. All of these steps try to minimize user effort and maximize user satisfaction.

The results adaptation approach is commonly used for personalization in many monolingual search systems. Adaptation of result lists can be performed by result scoring, result re-ranking or result filtering. There are three groups of techniques

* Corresponding author.

E-mail address: dongzhou@hnust.edu.cn (D. Zhou).

where the adaptation factors related to particular users can be obtained. The first group of techniques uses lexical matching between the documents and terms stored in the user model [33]. In this way, the selected terms from the user model can be viewed as the user's perspective of the query, in contrast to the initial ambiguous query that was issued by the user. However, lexical matching may ignore the semantic information, so researchers have attempted to map the documents and user models onto external knowledge bases. Among this second group of techniques, ODP categories or WordNet [13,28] and folksonomies [6,40,42–44,47] are the most popular choices of external knowledge. Semantic information can also be obtained by using latent semantic models. This third group of techniques use models, such as matrix factorization and Latent Dirichlet Allocation (LDA) [4], to calculate the latent topics and then map the user models and documents onto these topics to perform personalization [5,9].

In the literature [5,7,9,10,40,42–44], the adaptive factors that are used to score the documents according to the users' needs are linearly combined with information retrieval factors (i.e. the query term matching scores) in the scoring function. This combination is achieved using additional parameters to control the degree of personalization. However, the linearly combined scores may miss information when used to perform results adaptation. First, the associations between documents are not considered when personalizing the search results. The documents related to an initial query are not isolated; they are connected and can provide valuable information when calculating the adaptation scores. Second, the relationships between the scores acquired by non-personalized rankers and personalized rankers are not fully explored. Obviously, one set of scores may influence the other and vice versa. In addition, if one decides to add more adaptation scores into the final ranking scores, it is inevitable that the combination parameters will also grow linearly. Tuning such parameters is not easy [1,24,39]. Finally, when compared to the personalization research in monolingual web search, relatively fewer studies extend to the cross-language domain. There is only a single study using the lexical matching approach to calculate adaptation scores for cross-language search [16]. Clearly further investigations are required.

In this paper we address the personalized search problem through result adaptation techniques. We propose a novel iterative method based on document associations obtained from an initial ranker. The intuition behind this method is: *similar documents are likely to have similar scores with respect to both non-personalized and personalized factors*. The method assumes that results retrieved by non-personalized rankers and personalized rankers mutually reinforce each other rather than being used in linear combination. Updating one set of scores will iteratively propagate to other sets of scores via pairwise document relationships. To exploit these relationships, we adjust multiple scores using a function which regularizes the smoothness of document associations over a connected graph. The method also goes beyond bag-of-words based models (i.e. lexical matching). It considers latent topics derived from documents, queries and user models. In our method, we can simultaneously process multiple adaptation scores without complex parameter tuning procedures.

In order to evaluate the proposed method, we apply it in the scenario of personalized cross-language search by using two different test collections. One collection is semi-automatically constructed, motivated by an automatic methodology to evaluate personalized search systems [38] and a method to conduct known-item searches [2]. Another collection is constructed from a real click-through log. The motivation for the experimental procedure is that there are no readily available test collections for cross-language search which contain user-oriented information. Empirical evaluation performed with the two test collections attests to the effectiveness of our method. Specifically, retrieval performance is substantially better than that attained by using the linear combination of different adaptation features for cross-language personalization. Furthermore, the experimental results also prove that the semi-automatically constructed test collection can be used as an alternative basis for evaluation in the absence of real world datasets.

The contributions of this paper can be summarized as follows:

- i. *We tackle the challenge of personalized results adaptation in a novel way by assuming results retrieved by non-personalized rankers and personalized rankers mutually reinforce each other rather than being linearly combined.*
- ii. *We propose a novel iterative method based on document associations obtained from an initial ranker that can simultaneously process multiple adaptation scores.*
- iii. *We apply the proposed method in the context of personalized cross-language search by suggesting an evaluation methodology to help lower the high barrier to the evaluation of such systems. The approach aims to ensure repeatable and controlled experiments between different personalized strategies, ensuring comparable measures and generalizable conclusions about them.*

The rest of this paper is organized as follows. Related work is summarized in Section 2. Section 3 presents details of the proposed personalized results adaptation technique. Section 4 demonstrates the procedures used in the calculation of adaptation scores. In Section 5, a report on a series of experiments performed to evaluate the personalization strategies is provided. Also the procedure for building test collections for personalized cross-language search and other experimental settings are outlined. Finally, Section 6 concludes the paper and outlines areas of future research.

2. Related work

There is a significant volume of published research in the area of monolingual personalized search systems [30,32]. One common approach is result adaptation. Adaptation of result lists can be performed by result scoring, result re-ranking, or result filtering. Result re-ranking takes place after an initial set of documents have been retrieved by the system, where an additional ranking round is performed to re-order documents based on certain adaptation aspects (e.g. displaying certain documents at higher ranks in the result list based on the user's interests) [9,16,33]. Result filtering can be considered as a

Download English Version:

<https://daneshyari.com/en/article/6856846>

Download Persian Version:

<https://daneshyari.com/article/6856846>

[Daneshyari.com](https://daneshyari.com)