

A Language Modeling Approach to Personalized Search Based on Users' Microblog Behavior

Arjumand Younus^{1,2}, Colm O'Riordan¹, and Gabriella Pasi²

¹ Computational Intelligence Research Group, Information Technology,
National University of Ireland, Galway, Ireland

² Information Retrieval Lab, Informatics, Systems and Communication,
University of Milan Bicocca, Milan, Italy
{arjumand.younus,colm.oriordan}@nuigalway.ie, pasi@disco.unimib.it

Abstract. Personalized Web search offers a promising solution to the task of user-tailored information-seeking, and particularly in cases where the same query may represent diverse information needs. A significant component of any Web search personalization model is the means with which to model a user's interests and preferences to build what is termed as a *user profile*. This work explores the use of the Twitter microblog network as a source of *user profile* construction for Web search personalization. We propose a statistical language modeling approach taking into account various features of a user's Twitter network. The richness of the Web search personalization model leads to significant performance improvements in retrieval accuracy. Furthermore, the model is extended to include a similarity measure which further improves search engine performance.

1 Introduction and Related Work

Search engine users have diverse information needs, and it often happens that different users expect different answers to the same query [6]. In fact, given the potential of the same query to be representative of different information needs behind it, personalized Web search has emerged as a promising solution to better identify the intended information need. The usual approach to the personalization process in Web search involves incorporating user's preferences into the retrieval method of the search system thereby moving from a "one size fits all" approach to the customization of search results for people with different information interests and goals.

A significant research challenge in Web search personalization is to learn about a user's interests and preferences to build what is termed as a *user profile*. The user profile is the most essential resource within the retrieval model of a personalized search system. One of the main features that can be used to differentiate between existing solutions to Web search personalization is the source used when building the user profile. Several kinds of sources have been explored by researchers in order to build a user profile, with the most popular being search and browsing histories [3,5]. However, the use of such history data may not be

feasible given users’ privacy considerations that can limit the availability of the data. Furthermore, history data are more prone to noise as previous interactions with the search system are not necessarily reflective of user needs [8]. This paper proposes microblogs as an alternative information source to build a rich user profile.

The proliferation of Web 2.0 services has created a new form of user collaboration where users engage within a social network while at the same time generating their own content, popularly known as user-generated content. Microblogs such as Twitter¹ are an immensely popular forum for such collaboration, and we show how this forum can serve as a source of information about users’ preferences for the Web search personalization process. Earlier research efforts that aimed to exploit information from online social systems for personalized search rely mostly on social bookmarking and tagging systems [4,7]. However, Younus et al. [9] revealed in a user-survey based study a very low usage of social bookmarking sites as compared to other social networking tools.

A few works have considered Twitter as a source of user profile construction [1]; however, these works do not take into account features of a user’s microblog network. We undertake such a direction in this work and propose a statistical language modeling approach to infer a user’s profile; the proposed technique takes into account various features of a user’s Twitter network thereby providing a rich model of user preferences and interests. We evaluate the proposed methods by means of a user-study and we show that retrieval performance substantially improves when using microblog behavior as a source of information about user preferences and interests for Web search personalization.

2 Methodology

This section describes the proposed personalization model in detail. We follow a strategy in which non-personalized search results returned from a search system are re-ranked with the help of the user profile to return results that are more relevant to the user [5].

2.1 Microblog Behavior Based Language Model

We adopt a statistical language model to model various aspects of Twitter behavior. Using this model, we then define our re-ranking approach.

We incorporate the *mention* and *retweet* features of Twitter within our model with the underlying intuition that those Twitterers a particular user mentions or retweets reflect, to a large extent, the user’s own preferences and interests.

For the re-ranking step, we use a language modeling approach to compute the likelihood of generating a document d from a language model estimated from a user’s Twitter model as follows:

$$P(u)_{lm}(d/T) = \sum_{w \in W} P(w | T)^{n(w,d)} \quad (1)$$

¹ <http://twitter.com>