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

Arjumand Younus[1,2], Colm O'Riordan[1], and Gabriella Pasi[2]

[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

M. de Rijke et al. (Eds.): ECIR 2014, LNCS 8416, pp. 727–732, 2014.

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 Twitter[1] 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:

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

[1] `http://twitter.com`

where w is a word in the title and snippet of a document returned by a search system (i.e., d), W the set of all the words in the title and snippet of document d, *n(w,d)* the term frequency of w in d, and u is the user for whom we want to personalize Web search results. Here, T is used to represent the uniform mixture of the user's Twitter model as follows:

$$P(w \mid T) = \lambda_o * P(w \mid T_o) + \lambda_m * P(w \mid T_{U_m}) + \lambda_r * P(w \mid T_{U_r}) \tag{2}$$

Let T_o denote the original tweets by the user u, T_{U_m} denotes the tweets by those Twitterers whom the user u mentions (i.e., Twitterers in set U_m) and T_{U_r} denotes the tweets by those Twitterers whom the user u retweets (i.e., Twitterers in set U_r). The individual Twitter models can be estimated as:

$$P(w \mid T_o) = \frac{1}{|T_o|} \sum_{t \in T_o} P(w \mid t) \tag{3}$$

$$P(w \mid T_m) = \frac{1}{|U_m|} \sum_{u \in U_m} \frac{1}{|T_{u_m}|} \sum_{t \in T_{u_m}} P(w \mid t) \tag{4}$$

$$P(w \mid T_r) = \frac{1}{|U_r|} \sum_{u \in U_r} \frac{1}{|T_{u_r}|} \sum_{t \in T_{u_r}} P(w \mid t) \tag{5}$$

i.e., a single user's Twitter model is estimated by a mixture of his own tweets, those Twitterer's tweets whom the user mentions and those Twitterers' tweets whom the user retweets. The constituent language models for T_o, T_{U_m} and T_{U_r} are a uniform mixture of their tweets' language models employing Dirichlet prior smoothing:

$$P(w \mid t) = \frac{n(w,t) + \mu \frac{n(w, coll)}{|coll|}}{|t| + \mu}$$

where *n(w,.)* denotes the frequency of word w in (.), *coll* is short for collection which refers to all tweets by user u (in case of equation (3)), all tweets by Twitterers in set U_m (in case of equation (4)) and all tweets by Twitterers in set U_r (in case of equation (5)), and |.| is the overall length of the tweet or the collection.

Finally, after estimation of a user's Twitter model (using equations 2-5) we use equation (1) to re-rank the documents returned by a search system and hence, present personalized search results to the user u.

2.2 Similarity Measure between Users

In the previous section, we defined U_m as the set of users mentioned by u and U_r as the set of users whose tweets were retweeted by user u. We refine the definition of these sets to only include those users who have a sufficient similarity to the user

u. We present a network-based similarity measure which we use to decide whether or not to include a particular user in U_m and U_r. The underlying intuition behind the use of such a similarity measure is to exclude those Twitterers from the user's Twitter model who do not provide a strong indication of the user's preferences.

We calculate the similarity between the current user u and each user u_i occurring in either U_m or U_r based on the heuristic that the more people u_i follows in these sets, the more likely that user's interests overlap with the user u. Furthermore, we normalise this score by the total number of users, user u_i follows. We use the following formula to calculate the similarity score between user u and a user $u_i \in U_m$.

$$Sim(u, u_i) = \frac{follow(u_i) \cap U_m}{follow(u_i)}$$

where *follow(u_i)* is the set of users followed by u_i.

We also calculate similarity for all users in U_r using the same approach. Finally, we retain those Twitterers in U_m and U_r[2] whose network similarity measure is above a certain threshold.

3 Experimental Evaluations

In this section we describe our experimental evaluations that demonstrate the effectiveness of our proposed approach.

1. We wish to check whether personalization through a Twitter-based user profile improves search quality over the underlying non-personalized search engine.
2. We wish to evaluate the effect of the network based similarity of section 2.2 in an attempt to study the usefulness of incorporating microblog characteristics when building a user profile for Web search personalization.

3.1 Experimental Setup

We recruited 14 active Twitter users and used their Twitter data for the purpose of experimental evaluations. We obtained the search queries, their corresponding relevance judgements and underlying corpus (i.e., search documents' collection) from a publicly available dataset called *"CiteData"* by Harpale et al. [2]. As mentioned by Harpale et al., the dataset is useful for benchmark evaluations of personalized search performance. *CiteData* comprises 81,432 academic articles and 41 queries; we asked each user who participated in our user-study to select a subset of the queries that were similar to a search query that he/she had issued at some point. Note that since the dataset comprises academic articles we recruited Twitter users who are academics with specific, personalized information needs

[2] These are used as part of equation (4) and equation (5) for estimation of the user's Twitter model.

for academic articles. Each user was asked to select 10 queries from the 41 queries of the dataset; of these we selected the queries that had been selected by at least three users which amounted to a total of eight unique queries. We then asked each user to mark as relevant or irrelevant 20 documents per query; we obtain these 20 documents using a BM25 non-personalized search algorithm. Note that each user in our study was asked to mark 20 documents across the queries they selected from the short-listed eight queries. Finally, we calculate the Cohen's kappa across the relevance judgements for the eight short-listed queries; for the purpose of calculating Cohen's kappa we used the relevance judgements by the graduate students of Harpale et al.'s study and the relevance judgements by the users in our study. We obtain an average Cohen's kappa value of 0.86 across all queries and all users reflecting the high reliability of the *CiteData* dataset. We perform this step of measuring inter-annotator agreement via Cohen's kappa to ensure the agreement in relevance judgements between different sets of users in the two studies (i.e., the study by Harpale et al. and ours).

Table 1. Comparison of Retrieval Performance for our Proposed Personalization Model

Chosen Algo	Measures	
	MAP	$P@10$
np	0.389	0.567
p_{ns}	0.451	0.598
p_s	0.487	0.634

3.2 Experimental Results

Once we ensure reliability of the underlying dataset and relevance judgements through the method explained in section 3.1, we evaluate the performance of our proposed personalization model using the relevance judgements of the *CiteData* dataset. As evaluation metrics, we use mean average precision (MAP) and precision at top 10 documents (P@10) which respectively measure the systems overall retrieval accuracy and its performance for those documents that are most viewed. Table 1 shows the experimental results i.e. MAP and P@10 values for the non-personalized approach (denoted as np), our personalized approach without the similarity measure of section 2.2 (denoted as p_{ns}) and our personalized approach with the similarity measure of section 2.2 (denoted as p_s)[3]. The network similarity threshold was chosen following empirical analysis; for each user a threshold equivalent to half of the maximum similarity score was found sufficient to gather a significant amount of similar users in U_m and U_r. Moreover, the parameters λ_o, λ_m, and λ_r are assigned uniform weights. The baseline non-personalized search system uses a language model approach with Dirichlet smoothing. We report the results together across the queries and judgements for all 14 users.

[3] We use student's t-test to verify the soundness of our evaluations and the results corresponding to p_{ns} and p_s are statistically significant with $p < 0.05$.

The results show clearly, the benefits of using Twitter data to personalize search results for users. The MAP and P@10 scores for the personalized results (p_{ns} and p_s) are superior to those achieved without personalization. Furthermore, we witness improved performance when only those Twitterers who show similarity to the active user are used in generating the model.

4 Conclusions and Future Work

The main conclusion is that exploiting evidence available from a person's microblog behaviour to allow personalization can improve the accuracy of a system. We adopt a language modeling approach and show that including similar users from the Twitter network provides the best performance. Future work will involve further analysis of the results and explore other similarity measures and sources of evidence from a user's microblog behaviour and network. We also aim to merge these sources of evidence with data available in the query and about the user's current task at hand.

References

1. Abel, F., Gao, Q., Houben, G.-J., Tao, K.: Semantic enrichment of twitter posts for user profile construction on the social web. In: Antoniou, G., Grobelnik, M., Simperl, E., Parsia, B., Plexousakis, D., De Leenheer, P., Pan, J. (eds.) ESWC 2011, Part II. LNCS, vol. 6644, pp. 375–389. Springer, Heidelberg (2011)
2. Harpale, A., Yang, Y., Gopal, S., He, D., Yue, Z.: Citedata: a new multifaceted dataset for evaluating personalized search performance. In: CIKM 2010, pp. 549–558 (2010)
3. Matthijs, N., Radlinski, F.: Personalizing web search using long term browsing history. In: WSDM 2011, pp. 25–34 (2011)
4. Noll, M.G., Meinel, C.: Web search personalization via social bookmarking and tagging. In: Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L.J.B., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) ASWC 2007 and ISWC 2007. LNCS, vol. 4825, pp. 367–380. Springer, Heidelberg (2007)
5. Tan, B., Shen, X., Zhai, C.: Mining long-term search history to improve search accuracy. In: KDD 2006, pp. 718–723 (2006)
6. Teevan, J., Dumais, S.T., Horvitz, E.: Potential for personalization. ACM Trans. Comput.-Hum. Interact. 17(1), 4:1–4:31 (2010)
7. Vallet, D., Cantador, I., Jose, J.M.: Personalizing web search with folksonomy-based user and document profiles. In: Gurrin, C., He, Y., Kazai, G., Kruschwitz, U., Little, S., Roelleke, T., Rüger, S., van Rijsbergen, K. (eds.) ECIR 2010. LNCS, vol. 5993, pp. 420–431. Springer, Heidelberg (2010)
8. Wang, Q., Jin, H.: Exploring online social activities for adaptive search personalization. In: CIKM 2010, pp. 999–1008 (2010)
9. Younus, A., O'Riordan, C., Pasi, G.: Predictors of users' willingness to personalize web search. In: Larsen, H.L., Martin-Bautista, M.J., Vila, M.A., Andreasen, T., Christiansen, H. (eds.) FQAS 2013. LNCS, vol. 8132, pp. 459–470. Springer, Heidelberg (2013)