Evaluation of Personalized Social Ranking Functions of Information Retrieval

Mohamed Reda Bouadjenek[1,⋆], Amyn Bennamane[2,⋆], Hakim Hacid[3,⋆], and Mokrane Bouzeghoub[1]

[1] PRiSM Laboratory, Versailles University
{mrb,mok}@prism.uvsq.fr
[2] Dell Innovation House, Dublin, Ireland
amyn_bennamane@dell.com
[3] SideTrade, Boulogne-Billancourt, France
hhacid@sidetrade.com

Abstract. There is currently a number of interesting research works performed in the area of bridging the gap between Social Networks and Information Retrieval (IR). This is mainly done by enhancing the IR process with social information. Hence, many approaches have been proposed to improve the ranking process by personalizing it using social features. In this paper, we review some of these ranking functions.

1 Introduction

The Web 2.0 has introduced a new freedom for the user in his relation with the Web through social platforms, which are commonly used as means to interact. Hence, users are more active in generating content, which is one of the most important factors for the increasingly growing quantity of data. From the research perspective, this brings important and interesting challenges for many research fields like Information Retrieval (IR), which is the focus of this paper.

IR is performed every day in an obvious way over the Web, typically under a search engine. However, finding relevant information remains challenging for end-users. In existing IR systems, queries are usually interpreted and processed using document indexes and/or ontologies, which are hidden for users. The resulting documents[1] are not necessarily relevant from an end-user perspective, in spite of the ranking. To improve the IR process and reduce the amount of irrelevant documents, there are mainly three possible improvement tracks: (i) query reformulation, (ii) improvement of the IR model, and (iii) post filtering or re-ranking of the retrieved documents. In this last track, many approaches has been proposed to improve the ranking process by personalizing it using social features. In this paper, we propose to review some of these personalized social ranking functions that rely on social annotations as source of social information. These annotations are associated to documents in social bookmarking systems. In this paper, we try to mainly answer the following questions: *What are these*

⋆ This work has been mainly done when authors was at Bell Labs France, Villarceaux.
[1] We also refer to documents as web pages or resources.

F. Daniel, P. Dolog, and Q. Li (Eds.): ICWE 2013, LNCS 7977, pp. 283–290, 2013.

functions and how do they work? What is the context where each function is more efficient? What is the best ranking function?

The main contributions of this work can be summarized as follows:

1. We propose a deep study of the state of the art in social ranking functions.
2. We propose a deep analysis of the performances of these personalized social ranking functions and a comparison with non-personalized social approaches.
3. Finally, we propose a discussion on the effectiveness, the weakness and the performance of each approach in different contexts.

The rest of this paper is organized as follows: in Section 2, we introduce the main concepts used throughout this paper. In Section 3, we review the personalized ranking functions studied. Section 4 presents the dataset we used, and the evaluation methodology. The evaluations are presented and discussed in Section 5. Finally, Section 6 concludes this paper.

2 Background

In this section, we formally define the basic concepts that we use in this paper. Then, we formally define the problem of personalized ranking.

2.1 Background and Notation

Social bookmarking systems are based on the techniques of *social tagging*. The principle is to provide the user with a mean to freely annotate resources on the Web with tags, e.g. URIs in *delicious*. These annotations can be shared with others. This unstructured approach to classification is often referred to as a *folksonomy*. A folksonomy is based on the notion of bookmark defined as follows:

Definition 1. *Let U, T, R be respectively the set of Users, Tags and Resources. A bookmark is a triplet (u,t,r) such as $u \in U, t \in T, r \in R$, which represents the fact that the user u has annotated the resource r with the tag t.*

Then, a folksonomy is formally defined as follows:

Definition 2. *Let U, T, R be respectively the set of Users, Tags and Resources. A folksonomy $\mathbb{F}(U, T, R)$ is a subset of the Cartesian product $U \times T \times R$ such that each triple $(u, t, r) \in \mathbb{F}$ is a bookmark.*

In this paper we use the notation summarized in Table 1.

2.2 Problem Definition

Let consider a folksonomy $\mathbb{F}(U, T, R)$ whose a user $u \in U$ submits a query q to a search engine. We would like to re-rank the set of resources $R_q \subseteq R$ (or documents) that match q, such that relevant resources for u are highlighted and pushed to the top for maximizing his satisfaction and personalizing the search results. The ranking follows an ordering $\tau = [r_1 \geq r_2 \geq \cdots \geq r_k]$ in which $r_k \in R$ and the ordering relation is defined by $r_i \geq r_j \Leftrightarrow Rank(r_i, u, q) \geq Rank(r_j, u, q)$, where $Rank(r, u, q)$ is a ranking function that quantify similarity between the query and the resource w.r.t the user [7].

Table 1. Paper's Notation Overview

Variable	Description
u, d, t	Respectively a user u, a document d and a tag t.
U, D, T	Respectively a set of users, documents and tags.
$\mid A \mid$	The number of element in the set A.
T_u, T_d, $T_{u,d}$	Respectively the set of tags used by u, tags used to annotate d, and tags used by u to annotate d.
D_u, D_t, $D_{u,t}$	Respectively the set of docs tagged by u, docs tagged with t, and docs tagged by u with t.
U_t, U_d, $U_{t,d}$	Respectively the set of users that use t, users that annotate d, and users that used t to annotate d.
$Cos(\boldsymbol{A}, \boldsymbol{B})$	The cosine similarity measure between two vectors.
$\overrightarrow{p_u}$	The vector of the profile of the user u, estimated by its social annotations weighted using the tf-idf.

3 Personalized Ranking Functions Based on Folksonomies

In this Section, we formally define the different personalized ranking functions studied in this paper. We each time present the ranking score of a document d for a query q issued by a user u denoted $Rank(d, q, u)$.

3.1 Profile Based Personalization (Xu08)

The approach presented by Xu et al. [9] assumes the ranking score of a document d is decided by two aspects: (i) a textual matching between q and d, and (ii) a user interest matching between u and d. Hence, following our notation in Table 1, their approach can be defined as follows:

$$Rank(d, q, u) = \gamma \times Cos(\overrightarrow{p_u}, \overrightarrow{T_d}) + (1 - \gamma) \times Sim(\overrightarrow{q}, \overrightarrow{d}) \quad (1)$$

where, γ is a weight that satisfies $0 \leq \gamma \leq 1$, and $Sim(\overrightarrow{q}, \overrightarrow{d})$ denotes the textual matching score between d and q.

3.2 Topics Based Personalization (LDA-P)

We present here a topics-based approach. This approach is based on Latent Dirichlet Allocation (LDA) [3]. LDA-P relies on the fact that the set of tags can be used to represent web pages and as input for LDA to construct a model. Then, for each document that matches a query, LDA-P computes a similarity between its topic and the topic of the user profile using the cosine measure (inferred using the previous constructed LDA model). The obtained similarity value is merged with the textual ranking score to provide a final ranking score for a document that matches a query w.r.t the query issuer as follows:

$$Rank(d, q, u) = \gamma \times Cos(\overrightarrow{u_{topic}}, \overrightarrow{d_{topic}}) + (1 - \gamma) \times Sim(\overrightarrow{q}, \overrightarrow{d}) \quad (2)$$

where, $0 \leq \gamma \leq 1$, $\overrightarrow{u_{topic}}$ and $\overrightarrow{d_{topic}}$ are respectively the vectors that model the user and the document topics based on the constructed LDA model.

3.3 Social Context Based Personalization (SoPRa)

The approach proposed by Bouadjenek et al. [4] is similar to [9]. However, the authors propose to enhance the ranking process by considering a new aspect, which is the social matching score. This approach takes into account the entire social context that surround both users and documents and is called SoPRa. Following our notation, SoPRa can be defined as follows (β is set to 0.5):

$$Rank(d,q,u) = \gamma \times Cos(\overrightarrow{p_u}, \overrightarrow{T_d}) + (1-\gamma) \times \left[\beta \times Cos(\overrightarrow{q}, \overrightarrow{T_d}) + (1-\beta) \times Sim(\overrightarrow{q}, \overrightarrow{d})\right] \quad (3)$$

3.4 Scalar Tag Frequency Based Personalization (Noll07)

The approach presented by Noll and Meinel [6] considers only a user interest matching between u and d. This approach does not make use of the user and document length normalization factors, and only uses the user tag frequency. The authors normalize all document tag frequencies to 1, since they want to give more importance to the user profile. Following the notation given in Table 1, their ranking function can be defined as follows:

$$Rank(d,q,u) = \sum_{t \in T_u \wedge t \in T_d} |D_{u,t}| \quad (4)$$

3.5 Scalar tf-if Based Personalization (tf-if)

Vallet et al. [7] proposed to improve the Noll07 approach above by including a weighting scheme based on an adaptation of the *tf-idf* as follows:

$$Rank(d,q,u) = \sum_{t \in T_u \wedge t \in T_d} (tf_u(t) \times iuf(t) \times tf_d(t) \times idf(t)) \quad (5)$$

3.6 Affinity Based Personalization

Bender et al. [1] proposed several personalized ranking functions based on relations in a folksonomy. More precisely, we study in this paper the following two ranking functions that we consider as relevant to this survey:

1. Semantic Search: This approach ranks documents by considering users that hold similar content to the query, i.e., users who used at least one of the query terms in describing their content.
2. Social Search: This approach ranks documents by considering friends of the query issuer who used at least one of the query terms for tagging.

We refer the reader to the original paper for their definition. In the next sections, we describe the evaluations we have performed on these functions.

4 Dataset and Evaluation Methodology

In this section, we describe the dataset we used and the evaluation methodology.

4.1 Dataset

To evaluate our approach, we have selected a *delicious* dataset, which is public, described and analyzed in [8]. Before the experiments, we performed four data preprocessing tasks: (1) We remove annotations that are too personal or meaningless, e.g. "toread", "Imported IE Fa-vorites", etc. (2) The list of terms undergoes a stemming by means of the Porter's algorithm in such a way to eliminate the differences between terms having the same root. (3) We downloaded all the available web pages while removing those which are no longer available using the *cURL* command line tool. (4) Finally, we removed all the non-english web pages. Table 2 gives a description of the resulted dataset:

Table 2. Details of the delicious dataset

Bookmarks	**Users**	**Tags**	**Web pages**	**Unique terms**
9 675 294	318 769	425 183	1 321 039	12 015 123

4.2 Evaluation Methodology

Making evaluations for personalized search is a challenge since relevance judgments can only be assessed by end-users [2]. This is difficult to achieve at a large scale. However, different efforts [2,5] state that the tagging behavior of a user of a folksonomy closely reflects his behavior of search on the Web. In other words, if a user tags a document d with a tag t, he will choose to access the document d if it appears in the result obtained by submitting t as query to the search engine. Thus, we can easily state that any bookmark (u, t, r) that represents a user u who tagged a document d with tag t, can be used as a test query for evaluations. The main idea of these experiments is based on the following assumption:

For a query $q = \{t\}$ issued by u with query term t, relevant documents are those tagged by u with t.

Hence, for each evaluation, we randomly select 2000 pairs (u, t), which are considered to form a personalized query set. For each corresponding pair (u, t), we remove all the bookmarks $(u, t, r) \in \mathbb{F}, \forall r \in R$ in order to not promote the resource r (or document) in the results obtained by submitting t as a query in our algorithm and the considered baselines. By removing these bookmarks, the results should not be biased in favor of documents that simply are tagged with query terms and making comparisons to the baseline uniformly. Hence, for each pair, the user u sends the query $q = \{t\}$ to the system. Then, we retrieve and rank all the documents that match this query using a specific baseline, where documents are indexed based on their textual content using the *Apache Lucene.* Finally, according to the previous assumption, we compute the Mean Average Precision (MAP) and the Mean Reciprocal Rank (MRR) over the 2000 queries.

5 Results and Discussion

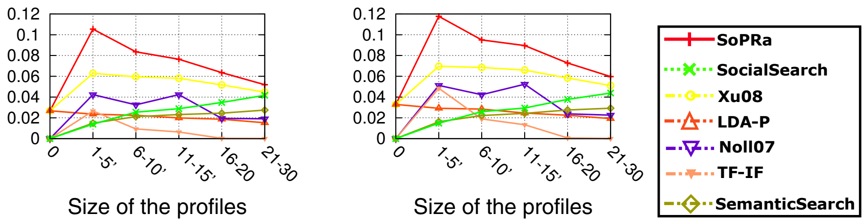
In this section, we conduct several experiments, which intend to address the following questions:

1. What is the effectiveness of these personalized ranking functions on users with different profile lengths?
2. Can these personalized ranking functions achieve good performance even if users have no bookmarks?
3. Are these personalized ranking functions efficient for large datasets?
4. What is the best personalized ranking function?

In the following, Section 5.1 addresses question 1 and 2, Section 5.2 shows the analysis of question 3, and lastly, Section 5.3 tackles question 4.

5.1 Performance on Different Users

Here, we try to study the ability of the personalized ranking approaches to achieve good performance for users that have different profile length, i.e. users that used few terms in their tagging actions. Hence, we propose to compare these approaches using the evaluation process described in Section 4.2. We select 2000 query pairs (u, t) based on the number of tags the users used in their tagging actions. The query pairs are grouped into 6 classes: “0”, 1-5”, “6-10”, “11-15”, “16-20”, and “21-30”, denoting how many tags users have used in their tagging actions, e.g. class “1-5” is composed with users who have a profile length between 1 and 5. Note that we fixed γ to 0.5 for all the approaches. The experimental results are shown in Figure 1.



(a) Mean Average Precision **(b)** Mean Reciprocal Rank

Fig. 1. Performance comparison on different queries, while fixing $\gamma = 0.5$

The results show that the performance of all the profile based approaches decrease for users with high profile length, i.e. SoPRa, Xu08, LDA-P, Noll07, tf-if. This is certainly due to the fact that these approaches fail to determine the user expectations, if he expressed his interest in different fields. However, the affinity based personalization approaches increase their performance for users with high profile length. These approaches are based on other user experiences with common tastes and affinities with the query issuer. Hence, we believe that modeling a user profile with simply his tags is not enough to generate satisfactory search results, especially for active users on social networks. We must go beyond that by considering their social relatives for ranking purpose.

Finally, we note that many personalized ranking functions are not able to provide a suitable ranking of documents for users with no tags. Currently, all

Table 3. Summary of the analysis

	General Performance[a]	Time Complexity[b]	Cold Start[c]	Adaptability[d]	Effectiveness[e]
Xu08	⋆⋆	$O(\mid \overrightarrow{p_u} \mid + \mid \overrightarrow{q} \mid)$	+	+	-
LDA-P	⋆	$O(n + \mid \overrightarrow{q} \mid)$	+	-	-
SoPRa	⋆⋆⋆	$O(\mid \overrightarrow{p_u} \mid + 2\times \mid \overrightarrow{q} \mid)$	+	+	-
Noll07	⋆	$O(\mid T_u \mid)$	-	+	-
tf-if	⋆	$O(\mid T_u \mid)$	-	+	-
SemanticSearch	⋆⋆	$O(\mid q \mid \times \mid U_t \mid)$	+	+	+
SocialSearch	⋆⋆	$O(\mid q \mid \times \mid U_t \mid \times \mid \overrightarrow{p_u} \mid)$	-	+	+

[a] The general retrieval performances. ⋆⋆⋆ : very effective; ⋆⋆ : effective; ⋆ : not effective.

[c] The cold start is a potential problem of a system to effectively handle new entities, e.g. users, items, or tags. In other words, it concerns the issue that the system cannot draw any inferences for users or items about which it hasn't information. + : can cope with cold start problem; - : cannot cope with cold start problem.

[b] The complexity is given for computing the ranking score of one document.

[d] Adaptability refers to the ability of approaches to consider new data and to quickly update their model. Considering new data is a key problem for these ranking functions since they are based on social networks, which are growing quickly. + : can easily update the model; - : cannot easily adapt the model.

[e] The effectiveness of the approaches for different profile lengths. + : effective for users with high profiles lengths; - : not effective for users with high profiles lengths.

the approaches, which are able to rank documents for users with no tags relay on the Lucene naive score for dealing with cold start problem.

5.2 Efficiency Analysis

We compare here the algorithms from the point of view of complexity. If we look at the complexity of each algorithm, we can distinguish 3 categories of algorithms, upon which we find common properties in term of computing complexity. These categories are the following: (i) Xu08, LDA-P and SoPRa algorithms, (ii) follow the Noll07 and tf-if algorithms, and (iii) the two last Affinity-based algorithms. The second category of algorithms is the most efficient with a complexity borned by the profile size of the user. Xu08-based algorithms come second in complexity, keeping the user profile size linearity and adding to it the query length. Finally, the third category is the affinity-based algorithms, which are the slowest ones, because they grow with at least the product of the profile size and the query size. This complexity analysis is summarized in Table 3.

5.3 Summary

Table 3 summarizes the personalized ranking functions studied from different point of views. This table is built upon our appreciation of the approaches.

As a conclusion, we believe that SoPRa offers the best trade off between retrieval performance, time complexity, cold start problem, and adaptability. However, the retrieval performance of this approach decreases for users with high profile length. We believe that we can tackle this issue by extending this ranking function by leveraging the social relatives of the query issuer.

6 Conclusion

This paper discusses a contribution to the area of Social Information Retrieval, which bridges the gap between traditional Information Retrieval and Social Networks. In this context, many approaches have been proposed to improve the ranking process by personalizing it using social features. We reviewed many of these personalized functions by proposing: (i) a deep study of the state of the art of ranking functions in social collaborative setting, (ii) a deep analysis of the performances of these personalized social ranking functions, and (iii) a discussion on the effectiveness, the weakness and the performance of each approaches in different contexts.

References

1. Bender, M., Crecelius, T., Kacimi, M., Michel, S., Neumann, T., Parreira, J.X., Schenkel, R., Weikum, G.: Exploiting social relations for query expansion and result ranking. In: ICDE Workshops (2008)
2. Bischoff, K., Firan, C.S., Nejdl, W., Paiu, R.: Can all tags be used for search? In: CIKM (2008)
3. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. 3, 993–1022 (2003)
4. Bouadjenek, M.R., Hacid, H., Bouzeghoub, M.: Sopra: A new social personalized ranking function for improving web search. In: SIGIR (2013)
5. Krause, B., Hotho, A., Stumme, G.: A comparison of social bookmarking with traditional search. In: Macdonald, C., Ounis, I., Plachouras, V., Ruthven, I., White, R.W. (eds.) ECIR 2008. LNCS, vol. 4956, pp. 101–113. Springer, Heidelberg (2008)
6. Noll, M.G., Meinel, C.: Web search personalization via social bookmarking and tagging. In: Aberer, K., et al. (eds.) ASWC 2007 and ISWC 2007. LNCS, vol. 4825, pp. 367–380. Springer, Heidelberg (2007)
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. Wetzker, R., Zimmermann, C., Bauckhage, C.: Analyzing social bookmarking systems: A del.icio.us cookbook. In: ECAI (2008)
9. Xu, S., Bao, S., Fei, B., Su, Z., Yu, Y.: Exploring folksonomy for personalized search. In: SIGIR (2008)