Axiomatic Term-Based Personalized Query Expansion Using Bookmarking System

Philippe Mulhem, Nawal Ould Amer, Mathias Géry

Axiomatic term-based Personalized Query Expansion using Bookmarking System

Philippe Mulhem[1,2], Nawal Ould Amer[1,2,3], and Mathias Géry[3]

1. Univ. Grenoble Alpes, LIG, F-38000 Grenoble, France - 2. CNRS, LIG, F-38000 Grenoble, France - 3. Univ Lyon, UJM-Saint-Etienne, CNRS, Institut d Optique Graduate School, Laboratoire Hubert Curien UMR 5516, F-42023, SAINT-ETIENNE, France
{nawal.ould-amer, philippe.mulhem}@imag.fr,
mathias.gery@univ-st-etienne.fr

Abstract. This paper tackles the problem of pinpointing relevant information in a social network for Personalized Information Retrieval (PIR). We start from the premise that user profiles must be filtered so that they outperform non profile based queries. The formal *Profile Query Expansion Constraint* is then defined. We fix a specific integration of profile and a probabilistic matching framework that fits into the constraint defined. Experiments are conducted on the Bibsonomy corpus. Our findings show that even simple profile adaptation using query is effective for Personalized Information Retrieval.

Keywords: social network, probabilistic retrieval, profile selection, axiomatic IR

1 Introduction

Personalized Information Retrieval (IR) systems aim at returning personalized results. Personalizing IR relies on modeling *User's profiles* (interests, behavior, history, etc.). Such profile may be used for *query expansion*, or for *re-ranking*. The query expansion-based integration keeps the benefit for all the experimental and theoretical results from the IR domain. A new field of IR has emerged with [7]: the axiomatic characterization of IR models. Such works define the expected behaviors of systems using "axioms".

This paper first defines an axiom (i.e. a heuristic constraint) that is supposed to be validated by a personalized IR system using a social bookmarking system, and second evaluates the impact of the constraint on the IR system. Section 2 presents related works. The section 3 defines the proposed axiom, called PQEC (Personalized Query Expansion Constraint). Section 4 focuses on the personalized frameworks proposed, before presenting several query expansions in section 5. The experiments on the Bibsonomy corpus are presented and discussed in Section 6, before concluding.

2 State of the art

Personalized IR may consider user's model (called *user's profile*) based on user's query logs [6], posts [11], tags and bookmarking [1]. Several works improve personalized document ranking by using both the user's information and other social information.

Such search function, for bookmarking systems, is based on user's tag profiles which are derived from their bookmarks [3, 12]. [1] selects terms related to the user query terms. Similarly, [4] defines a query expansion that exploits relationships between users, documents, and tags. [3, 12] considers both the matching score between a query and the social annotations of the document, and the matching between the user's profile and the document. Other works personalize a user search using other users from the social network. For example, selecting users that have an explicit [11, 9] or implicit [3, 12] relationships with the query issuer. [11] proposes a collaborative personalized search model based on topic models to disambiguate the query. [3] integrates other users from the social network that have annotated the document.

These approaches use the whole user profile, decreasing the effectiveness of the search. Query expansions tackle this problem by selecting the terms to extend user query. Our proposal benefits of both query expansion-like approaches [1, 4] and social retrieval [3], and we defend the idea that social networks are beneficial to personalized retrieval by: (i) adapting the user profile using social neighbors that are constrained by the query, and (ii) selecting a part of a user profile adapted to a query.

Our approach also focuses on defining axioms (heuristics), i.e. expected behaviors of personalized IR systems. Such axioms serve as a basis to a) explain the role of the different elements that are used by an IR system, b) compare approaches from the theoretical basis and c) propose new approaches based on these axioms. For instance, Fang, Tao and Zhai defined in 2004 [7] the first steps of this field of IR, with constraints related to the roles of term frequency, inverse document frequency, and document length. Many works followed, like heuristics for semantic models for IR [8], or for Pseudo Relevance Feedback [5]. To the best of our knowledge, no axiomatization work did focus on personalization of IR.

3 Profile Query Expansion Constraint

We propose here: a) to show that, in social bookmarking networks, integrating a part of a user activity (i.e. his bookmarks) may help to personalize results, and b) to define a first axiomatic expression that respects the findings of a).

3.1 Empirical study

We studied a set of 200 users from the Bibsonomy corpus, according to the evaluation framework described in section 6.1. We compute that, when a query is generated for a given user using a term from his profile, 100% of the relevant documents are tagged by at least one other term of the user profile. This empirical result enforces the fact that at least a part of a user profile is relevant to be used when processing personalized IR.

We study then the topics of queries. We generated a Latent Dirichlet Allocation model [2] for the whole set of users of the corpus (see part 6.1), with the number of topics chosen to be 100. Using a threshold of 0.1 when assigning a topic to a user, we find out that 77% of the users have more than one center of interest. If we assume that a query deals with one topic, as in [11], it is then clear that we have to filter terms of the profile to expand the query. All these elements reinforce our initial idea that focusing on an *adequate* subset of the user profile may help to focus on relevant documents.

3.2 Notations

Here are the notation used in the remaining of the paper. G: The tagging social network; G is a graph: $G =<< D, U, W >, R >$. D : the set of documents $d \in D$. U: the set of users of the network, with $u \in U$. W: the set of tags (words) assigned by users to documents. R: the tags assigned by the users to the documents ($R \subset D \times U \times W$). $c(w, d)$: the count of word w in document d. $RSV(d, q)$: the Retrieval Status Value of a document d for a query q. $Profile(u)$: the profile of a user u by all the tags he used. $Profile(u) = \{w | w \in W, d \in D, R_u(d, u, w)\}$. R_u: term-term relationship for user u. $(w, w') \in R_u$ means that w and w' are related for the user u. $Profile(u, q)$: the profile of a user u filtered for the query q. $R_{u-local}$: term-term relationship for user u based on u's tagging. $RSV(d, q, u)$: the RSV of a document d for a query q and for u. $u_{sn} \subset U$: the social neighborhood of u. $R_{u-social}$: term-term relationship for u considering u_{sn}.

3.3 Profile Query Expansion Constraint (PQEC)

This constraint assumes that the integration of "adequate" terms (related to the query, and satisfying the term-term relationship R_u) from a user profile is needed:

Profile Query Expansion Constraint (PQEC): Assume that a query $q = \{w\}$, a document d from a corpus C so that $c(w, d) > 0$, and a user u with a profile $Profile(u)$. If $\exists w' \in Profile(u)$ so that $R_u(w, w')$, then for any $d' \in D$ so that $c(w, d') \neq 0$ and $c(w', d') = 0$ then $RSV(d, q_u) \geq RSV(d', q_u)$, with $q_u = q \cup \{w'\}$.

This constraint heavily relies on the personalized term-term relationship R_u that obviously influences the overall results: if R_u does not link properly terms according to the user u, then ensuring the constraint will impact negatively the quality of the system. In the following, we will focus on social inputs to define the R_u relationship. Our concern differs from semantic term constraints of [8], as we consider that the data that we have about the user is of primary importance.

4 Personalized Information Retrieval

4.1 Classical Framework

Our proposal computes a Retrieval Status Value (RSV) of a document d for a query q submitted by a user u as: $RSV(d, q, u) \propto RSV(d, q_u)$ with q_u the expanded query using terms coming from u's profile: $q_u = q \cup \{w' | w' \in W, \exists w \in q; R_u(w, w')\}$. Each document d (tagged using a social tagging system) contains 2 facets: the actual content of the document, noted σd, and the user's tags that describe d, noted τd. We combine linearly these facets in the expression $RSV(d, q)$, as in [3], using probabilities $P(q|\sigma d)$ and $P(q|\tau d)$ that rely on the classical IR language models with Dirichlet priors:

$$RSV(d, q) = \lambda . P(q|\sigma d) + (1 - \lambda) . P(q|\tau d) \quad (1)$$

4.2 Adapted Framework to ensure PQEC

A simple way to modify the classical framework to ensure PQEC is to split the retrieval in four steps:

1. Evaluate $RSV(d, q)$, i.e. without personalization, leading to a results list L_{init} of couples $< doc, rsv >$. Assign the larger score for the documents of L_{init} to $Topscore_{init}$;
2. Evaluate $RSV(d, q_u \setminus q)$, i.e. the RSV of d for the expanded query q_u without the initial query, leading to a results list L_{exp} of couples $< doc, rsv >$;
3. Fuse L_{init} and L_{exp} respecting: a) for any d in L_{init} and L_{exp}, the final RSV of d is the sum of its scores in both lists and of $Topscore_{init}$; b) for any d in L_{init} and not in L_{exp}, its final RSV is its score in L_{init};
4. Rank the result according to the final scores of documents

The documents that match the profile expansion are ranked before the documents that match only the query in the result list, thus PQEC is ensured.

5 Personalized Query Expansion Terms

We propose several variations of u_{sn}, the social neighborhood of a user u, depending on adaptations of the social neighborhood of u and of the profiles of the users in u's neighborhood, according to the query q. We define several $R_u(w, w')$ (cf. 3.3) to assess the usefulness of the constraint. Our personalized query expansion may also use others users u' in the social network: we will study in section 6 the impact of PQEC on several categories of neighbors u', and on the filtering of u's profiles added to the query. In the following, $Profile(u, q)$ is the personalized profile of u asking q. We have: $Profile(u, q) = \{w' | w' \in W, \exists w \in q; R_u(w, w')\}$. We define two relationships, namely $R_{u-local}$ and $R_{u-social}$, that depict two personalized profiles of u.

5.1 Local tagging expansion using $R_{u-local}$

Assuming a query asked by a user u. The first simple element proposed is to add terms from u's profile to the query, relying only on u's tagging behavior. We select the tags from u's profile that were used jointly with a query term q by u to tag one document. The idea is then to be able to expand the query with terms that are related to one query term according to u. More formally, the relation R_u is then expressed as its variant $R_{u-local}$: $R_{u-local}(w, w') \Leftrightarrow \exists d \in D, R(d, u, w) \wedge R(d, u, w')$. Such approach cannot be used when a user u does not: (i) use a query term in his profile, and (ii) tag documents with multiple tags, and this is mandatory for this local expansion. We propose then a second way to support a query expansion.

5.2 Social tagging neighborhoods using $R_{u-social}$

When considering other users than u for the social tagging network, we need to define which users are considered experts to support the query expansion. Such neighbors set is noted u_{sn}. These experts are chosen according to their familiarity with the query,

and/or their similarity with the user u. In section 6.2, we define several neighborhoods. We consider here a simple definition of the profiles of $u' \in u_{sn}$. These profiles are built the same way as the profile of u using $R_{u'-local}$, i.e. they are filtered to keep the terms of $Profile(u')$ that co-occur with at least one query term in one document tagged by u'. Finally the profile of u is computed using the following expression of $R_{u-social}$: $R_{u-social}(w, w') \Leftrightarrow \exists u' \in u_{sn}, R_{u'-local}(w, w')$

6 Experimental evaluation

6.1 Bibsonomy dataset and evaluation protocol

We consider here explicit annotations of documents provided by a user from a tagging social network, namely Bibsonomy[1], which is a social tagging network dedicated for users to share their documents (using text tags) with other users of the network. It contains tags assigned by identified users to scientific articles (DOI) and Web pages (URL). From the full original corpus, we considered only the Web pages that still exist in September 2015, leading to a set of 308'906 documents. 241'706 document d are tagged by 4'911 users u, with 1,5 million occurrences of 59'886 unique tags w. On average, each user used 263 tags and each document has 6 tags.

On this dataset, we use the evaluation protocol of [3], which selects randomly one user u, and one random tag t used by u, as a query. All the tagging made by u using t on documents are then removed from the dataset. Then, the documents d initially tagged by u are marked relevant. We created 200 single term queries using this protocol.

Classical measures evaluate the quality of the retrieval: MAP, $P@5$, $P@10$. Two other measures detail the configurations studied: a) $PQEC@10$ measures the level of validation of PQEC on the top-10 results the the frameworks: it is the ratio of the top-10 documents that do not contain a user u tag and that are ranked before a document that is tagged by a tag used by u. A strict validation of PQEC (i.e. $PQEC@10 = 1.0$) is expected to lead to better results; and b) the $Prof_{overlap}$ values that describes the amount of overlap between the extended query and the user u profile. Such value is in $[0, 1]$. All statistical significance tests are paired bilateral Student t-tests.

6.2 Tested configurations

All the experiments are based on language models with Dirichlet priors using the default parameters of Terrier 4.0 [10] (english stoplist, Porter stemmer, $\mu = 2500$). Similarly to [3], we fix $\lambda = 0.5$ for the documents matching in equation (1). We tested four groups of configurations: baselines (without query expansion, or with the full user profile expansion), very dense, dense and sparse neighborhoods. They simulate different topologies of users networks.

Baselines - Our approach is compared to two baselines: (1) general profile retrieval, where the user profile is represented by all his tags in $Profile(u, q) = Profile(u)$, and (2) a non-personalized retrieval, where the initial query only is used.

[1] http://www.bibsonomy.org

The results are presented in Table 1 (runs a, b and c). Using the full user profile (runs a and b) clearly outperforms the run c without any profile. The MAP differences between a and b are not significant (p=0.101), but they are significant between runs a and c (p=7.95E-09), as well as between b and c (p=1.32E-10). Moreover, the adaptation described in 4.2 outperfoms the classical framework. We notice also that the run b has a relatively low value for $PQEC@10$: most of the time in the classical framework the constraint PQEC does not hold.

Very dense neighborhoods - Here, all the user set U is used as a neighborhood, so $u_{sn} = U$. We also study the fact that we filter, or not: a) the users from u_{sn} according to the fact that they are related to the query (i.e. they tagged one document with one query term). When filtering these users, we obtain an average of 152 neighbors for u; b) the profiles of the users u' from u_{sn}. When they are not filtered we use the $Profile(u', q) = Profile(u')$, when they are filtered the used profiles for u' are filtered according to 5.2.

These results are presented in Table 1 (runs d to i). The runs f and h (resp. g and i) have exactly the same values for MAP, $P@5$ and $P@10$, because the filtered u_{sn} already generates the full user profile (as $Prof_{overlap} = 1.0$). Here again the adapted frameworks outperform their respective classical ones. The filtered profiles from the neighbors outperform the unfiltered ones: choosing the "right" terms of the neighbors profiles has a positive impact. The best results are obtained with an average of 30% terms of the user's profile, which fits wells to the fact that users have more than 2 topics on average (as seen in subsection 3.1). Here again we do not conclude that there are statistically significant differences between MAP of adapted d or classical e runs (p=0.099), we notice however that adapted filtered run d has significant MAP differences (with p<0.001) with all the other runs in Table 1, where the classical filtered run e has no significant difference between runs f (and h) with a p value of 0.304.

Dense neighborhoods - We consider a relatively dense subset of U for u_{sn}. The social neighborhood of u is composed of users u' that share at least one tag with u' profile: $\{u'|Profile(u') \cap Profile(u) \neq \emptyset\}$. Here, each user has on average 872 neighbors. We filter these users according to the fact that they are related to the query or not; and we investigate the impact of filtering or not the profiles of these users. The filtering of users according to the query gives neighborhoods of 40 users in average.

The results are presented in Table 1 (runs k to p). Again, the adapted frameworks outperform the classical ones. We notice that the best results for MAP and $P@10$ are obtained when the profiles of the neighbors are not filtered (run m), with queries expansions containing 68% of the user's profile, on average. For the runs o and p, increasing the overlap with the user's profile does not help, except for the $P@5$ value, slightly higher for the run o than the run m. For the best dense neighbors run of Table 1, m, the difference in MAP is not statistically significant with its classical counterpart n (p=0.177), neither with the two unfiltered runs o (p=0.115) and p (with p=0.065). This is explained by some instability of the neighbors selected.

Sparse neighborhoods - The last set of configurations studied mimics sparse neighborhoods for a user u (inspired from [3]): the social network of u is composed of users u' that tagged at least one document that u tagged, whatever the tags are: $\{u'|\exists w \in W, \exists d \in D, R(u', d, w) \wedge R(u, d, W)\}$. Compared to other neighborhoods, the neigh-

Table 1. Retrieval performances for all the runs.

Run	framework	u_{sn}	$Profile(u', q)$	PQEC@10	$Prof_{overlap}$	MAP	P@5	P@10
a	adapted	$\emptyset$	$Profile(u)$	1.0	1.0	**0.4950**	**0.1860**	**0.1260**
b	classical	$\emptyset$	$Profile(u)$	0.0521	”	0.4639	0.1570	0.0970
c	classical	$\emptyset$	$\emptyset$	/	0.0	0.2934	0.1010	0.0585
d	adapted	filtered	filtered	1.0	0.3086	**0.5528**	**0.2060**	**0.1285**
e	classical	filtered	filtered	0.0646	”	0.5205	0.1790	0.1095
f	adapted	filtered	unfiltered	1.0	1.0	0.4950	0.1860	0.1260
g	classical	filtered	unfiltered	0.0521	”	0.4639	0.1570	0.0970
h	adapted	unfiltered	unfiltered	1.0	1.0	0.4950	0.1860	0.1260
i	classical	unfiltered	unfiltered	0.0521	”	0.4639	0.1570	0.0970
k	adapted	filtered	filtered	1.0	0.2508	0.4015	0.1590	0.0925
l	classical	filtered	filtered	0.0608	”	0.3946	0.1380	0.0810
m	adapted	filtered	unfiltered	1.0	0.6770	**0.4779**	0.1770	**0.1195**
n	classical	filtered	unfiltered	0.0410	”	0.4497	0.1590	0.0925
o	adapted	unfiltered	unfiltered	1.0	0.8695	0.4413	**0.1820**	0.1065
p	classical	unfiltered	unfiltered	0.0224	”	0.4269	0.1560	0.0880
q	adapted	filtered	filtered	1.0	0.2286	0.3923	0.1500	0.0930
r	classical	filtered	filtered	0.1020	”	0.3799	0.1310	0.0760
s	adapted	filtered	unfiltered	1.0	0.6300	0.3559	0.1480	0.1030
t	classical	filtered	unfiltered	0.0757	”	0.3708	0.1330	0.0795
v	adapted	unfiltered	unfiltered	1.0	0.8150	**0.3960**	**0.1680**	**0.1015**
w	classical	unfiltered	unfiltered	0.0804	”	0.3755	0.1350	0.0790

bors are here expected to be more similar to u, because they focused on the same document. There are 56 neighbors, on average. Moreover the filtering of users according to the query (as described before) leads to sets of 10 neighbors on average.

The results in Table 1 (runs q to w) show that, unlike the others neighborhoods, the best configuration is obtained by unfiltered neighbors and profiles, and the adapted framework. Another difference with previous results is that one adapted run, namely s, underperforms its respective classical run. This is due to an inadequate filtering of the neighbors, and then, applying subsequently the adapted framework degrades the quality of the results. The differences in MAP are small, this explains why we could not find statistically significant differences for the MAP values between these runs.

6.3 Discussion

The first point that we get from these experiments is that our frameworks that validate the PQEC constraint are consistently better than the classical framework (though without statistically significant differences taken one against one, but the repetitive outperformance is clear). Our adapted framework is very simple and should certainly be extended to tackles more clearly queries with multiple terms, but the current proposal already shows its interest on the quality of the results. The second point is that filtering the neighbors according to the query, using a very large set of potential users (i.e., very dense neighborhoods) seems to lead to better results than filtering *a priori* users (i.e.

dense or sparse neighborhoods). Processing very dense neighbors necessitates, for each query, to process the whole set of users. However, if users' profiles are represented as documents in a classical IR system, retrieving users that match a query is fast.

7 Conclusion

This paper proposes a probabilistic framework that exploits the profile of a user u asking a query q, in order to improve the search results. The profile is filtered regarding the query. We investigated two parameters that help in selecting the relevant parts of u's profile: one that exploits the query to select a useful subset of social neighbors of u, and one that uses sub-profiles of neighbors of u according to q. The main conclusion drawn from our experiments on the Bibsonomy corpus is that adapting the set of all users to the query and filtering u's profile according to the query improves the results. Short term extensions of this work will study the use of real friendship relations as social neighbors. Other future works will focus on users u with empty profiles that do not benefit from the proposed profile adaptations. We will then explore how social neighbors may be used to consider terms that do not belong to the initial profile of u.

Acknowledgements. This work is partly supported by the Guimuteic project funded by Fonds Européen de Développement Régional (FEDER) of région Auvergne Rhône-Alpes projects and by the ReSPIr project of the région Auvergne Rhône-Alpes.

References

1. Biancalana, C., Gasparetti, F., Micarelli, A., Sansonetti, G.: Social semantic query expansion. ACM Trans. Intell. Syst. Technol. 4(4), 60:1–60:43 (2013)
2. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. JMLR 3, 993–1022 (2003)
3. Bouadjenek, M.R., Hacid, H., Bouzeghoub, M.: Sopra: a new social personalized ranking function for improving web search. In: SIGIR conference. pp. 861–864 (2013)
4. Bouadjenek, M.R., Hacid, H., Bouzeghoub, M., Daigremont, J.: Personalized social query expansion using social bookmarking systems. In: SIGIR Conference. pp. 1113–1114 (2011)
5. Clinchant, S., Gaussier, E.: Information-based models for ad hoc ir. In: SIGIR Conference. pp. 234–241 (2010)
6. Dou, Z., Song, R., Wen, J.R.: A large-scale evaluation and analysis of personalized search strategies. In: Conference on World Wide Web. pp. 581–590 (2007)
7. Fang, H., Tao, T., Zhai, C.: A formal study of information retrieval heuristics. In: SIGIR Conference. pp. 49–56 (2004)
8. Fang, H., Zhai, C.: Semantic term matching in axiomatic approaches to information retrieval. In: SIGIR Conference. pp. 115–122 (2006)
9. Khodaei, A., Sohangir, S., Shahabi, C.: Recommendation and Search in Social Networks, chap. Personalization of Web Search Using Social Signals, pp. 139–163. Springer (2015)
10. Ounis, I., Amati, G., Plachouras, V., He, B., Macdonald, C., Lioma, C.: Terrier: A High Performance and Scalable Information Retrieval Platform. In: SIGIR Workshop on Open Source Information Retrieval (2006)
11. Vosecky, J., Leung, K.W.T., Ng, W.: Collaborative personalized twitter search with topic-language models. In: SIGIR conference. pp. 53–62 (2014)
12. Xu, S., Bao, S., Fei, B., Su, Z., Yu, Y.: Exploring folksonomy for personalized search. In: SIGIR Conference. pp. 155–162 (2008)