Overview of the social information's usage in Information Retrieval and Recommendation systems

Abir Gorrab, Ferihane Kboubi and Henda Ben Ghezala

RIADI Laboratory-ENSI
University of Manouba
Manouba 2010, Tunisia

Abir.Gorrab@riadi.rnu.tn

ABSTRACT. Web 2.0 and social networks represent a huge and rewarding source of information. Our work revolves around the issue of access and identification of social information and their use in building a user profile enriched with a social dimension, and operating in a process of personalization and recommendation. We present several approaches of Social IR (Information Retrieval), distinguished by the type of social information integrated; as well as social recommendation approaches. We also expose a study of the modeling techniques of the user profile's social dimension, followed by a discussion and proposed directions.

RÉSUMÉ. Le web 2.0 et les réseaux sociaux représentent une source d'information énorme et enrichissante. Notre travail s'articule autour de la problématique d'accès et d'identification des informations sociales et leur exploitation dans la construction d'un profil utilisateur enrichi d'une dimension sociale, et son exploitation dans un processus de personnalisation et de recommandation. Nous présentons différentes approches de RI (Recherche d'Information) Sociale, distinguées par le type d'informations sociales intégrées; ainsi que des approches de recommandation sociale. Nous exposons également une étude des techniques de modélisation de la dimension sociale du profil utilisateur, suivie par une discussion et des directions envisagées.

KEYWORDS: social information retrieval, social recommendation, social networks, user profile

MOTS-CLÉS : recherche d'information sociale, recommandation sociale, réseaux sociaux, profil utilisateur.

1. Introduction

With the apparition of the social web and the explosion of social networks, users become able not only to consume, but also to product informational content. As a matter of fact, the huge number of web users and time spent daily on internet motivated researchers in IR and encouraged them to benefit from this content as an enlightening source of information. Besides, social networks and collaborative sites (such as Facebook, LinkedIn, Google+, Twitter, YouTube, delicious, CiteUlike, etc) are the most common and popular source of interactive content. In this paper, we focus on the impact of social information integration in an IR process and a recommendation system.

This paper is organized as follows: in section 2, we discuss the main approaches used in Social IR. While section 3 is devoted to the social recommendation, and section 4 is reserved to social dimension in user profile modeling, section 5 is dedicated to discussion and future directions. Finally, we conclude in section 6.

2. Social IR approaches

Social IR approaches are various. They are mostly based on social information identification and integration in a search process. In fact, there are several types of social information specific to each social network, such as folksonomies, tags, social relations (friends, co-authors, followers), comments, tweets, conversations, hashtags, like mentions, shares and many others. Proposed approaches widely use many social information, which can be integrated in different levels in IR process: user profile construction, query expansion and result ponderation. In this section, we present different social IR approaches that can be classified in three categories, according to social information used.

2.1. Approach based on annotations

Social annotations are a valuable informational source that enhances social IR by including user's area of interest. Bouhini et al. [2] propose a user profile generation approach from folksonomies. As a matter of fact, this work combines queries with user profile based on terms frequency. Actually, it presents two Social IR models inspired from BM25 model: BM25S Score Comb and BM25S Freq Comb, which combines query and user profile using respectively scores and terms frequency. Bao et al. [17] calculate similarity between web query and social annotations. They propose two algorithms that enhance web IR: SSR (Social Sim Rank) which computes similarity degree and SPR (Social Page Rank) that estimates web pages popularity. PengLi et al

[15] propose a TR-LDA model of annotations categorization. They introduce representation and ponderation methods of annotation categories. In this respect, they study the effect of annotations' incorporation in IR process.

2.2. Approach based on social relations

At first sight, users may be linked by different relationships that are specific to each social network, e.g., friend relationships, followers and co-authors. Works based on this approach usually use this informational content generated by relations, in a way that combines a social and a thematic score. In this context, Ben Jabeur et al. [9] investigate on a social model based on Bayesian network, incorporating two social relevance factors: User social importance, evaluated by a PageRank score; and the number of temporal neighbors. Moreover, Amer et al. [14] propose a probabilistic model of conversation indexation in twitter. This model incorporates social relations to measure users' influence, activity and expertise.

2.3. Approach based on social signals

Social signals like comments, shares and like mentions are being more explored in social IR works, due to the significant information they bring. Chelaru et al [16] study the impact of these social signals in video search on YouTube, by combining social information such as comments, like and dislike mentions, with basic search criteria (similarity between the query and video title). Hence, this unification enhances the performance of videos extraction process. Furthermore, Badache et al. [5] describe a language model exploiting temporal characteristics of social signals (number of like mentions, shares and comments) to estimate resources relevance and sort search results. Moreover, Ramesh et al. [13] examine the personalized social IR process and suggest an algorithm of user profile construction using pages liked on Facebook, through different user's accounts. This social content personalizes search results.

2.4. Comparative study

For more information about Social IR approaches, we led a comparative study of different categories. For each work we considered the following six points as a comparative criteria: (1) the social network used for the experimentation, (2) the techniques used in the presented models, (3) the metrics used for the evaluation, (4) if there is a combination of information, (5) if there is a combination of social networks and (6) if the work considered the temporal aspect. Table 1 summarizes the results of our study.

Social information		Social network	Techniques	Evaluation	Combining Information	Combining social networks	Temporal aspect
Annotati-ons	[2]		BM25and derivatives	MAP, P[0,1]	-	-	-
	[17]	Del.ic.ious	SSR, SPR	MAP, nDCG	-	-	-
	[15]	Del.ic.ious	TF_IDF, Inference algorithm	-	-	-	-
Social relations	[9]	Twitter, Citulike	PageRank, language model, ImpG:social score, TF-IDF	MAP, recall	√	√	√
	[14]	Twitter	BM25, language model, PageRank	Leave One Out approach, MAP	√	-	-
Social signals	[16]	Youtube	TF_IDF, Lucene, SentiWord-Net	nDCG	√	-	-
	[5]	Facebook, Twitter, LinkedIn, Del.ic.ious, Google+	Language model	MAP, nDCG, Recall, Precision	√	√	√
	[7]	Facebook	Clustering data TF-IDF	Performance measure	√	-	-

Table1. Comparative table of Social IR approaches categorized by social information types

In table1, we present some works related to the three approaches described in this section, based on annotations, social relations and social signals. Characteristics studied are the combination of many social information or social networks and the consideration of temporal aspect. These features enhance IR processes and improve their performances. In fact, many networks are used and many techniques are conducted, but

temporal aspect and the combination of different networks represent the greatest motivation for researchers.

3. Social recommendation

Social recommendation is a set of methods that try to suggest items or entities that seem to be interesting to the user, using his social information [12]. In fact, there are two main recommendation techniques. The first one is a content based approach which is based on recommending items similar to those the user has chosen in the past. The second one is a collaborative filtering approach; this approach recommends items to the user based on the choice of other people, who seem to have similar preferences. Moreover, Hybrid recommendation is a technique that comprises both content-based and collaborative filtering approaches, so as to provide the user with better recommendations. Additionally, many researchers have explored social information to improve recommender systems. Notably, Hafsi et al. [11] exploit user-generated content (rating and review) in books recommendation system. Their work measures books reputation and popularity concepts and tests three approaches: book tags and reviews indexation, themes interrogation and users similarity calculation. Unlike in [6], authors have proposed a content-based approach that compares users profiles' information in order to determine similarities between them and recommend friendship relations. On the other hand, Wang et al. [18] investigate on tag based social recommendation by calculating tags similarities and connecting users that are likely to have similar tastes and preferences. In the same context, Hannon et al. [7] propose an hybrid recommendation system using content and collaborative-based approaches that recommends users to follow in the social network Twitter, by analyzing their profiles.

4. Social dimension in user profile modeling

User profile modeling is an essential task in Personalized IR. This entity brings and organizes the information necessary to define the user and describe his interests.

Following the emergence of social networks, Social IR has widely evolved. Thus, the social dimension of the user profile has become an essential component in social personalization systems. A lot of works were directed towards the construction of a social profile based on annotations [8] [4], given the importance of the data they generate. Others have focused on the analysis of egocentric social network, they are interested in friendship relationships in social networks [3] [10]. This information produces relevant content for collaboration within social IR systems. It solves the cold start problem, or lack of user's activity on social networks. The temporal aspect is also

reflected in some works [10], which differentiates between recent and old social activities, to estimate their importance. Other social signals have also been integrated into the social dimension of user profile such as comments and shares. Once the profiles are built, some authors have thought of building virtual communities of users, based on similarity degree between the profiles. These communities are considered as a dimension in the profile. They are very rewarding and provide additional relevant information. In [1], Dridi et al model a user profile based on annotations and exploit it to detect communities based on annotations similarities. For community detection, Katz index is used. It calculates the similarity taking into account the direct and indirect links in a graph.

5. Discussion and future directions

In this section, we discuss different aspects related to research in Social IR topic. In fact, IR classical approaches do not take into account the user's social content provided by his interactions and social relations. Moreover, most of the existing approaches in Social IR use either social signals, tags or relational information. Some works started leveraging different types of information. Also, combining social content from many social networks and matching different user's social profiles improve the collection of relevant information that better describe the user and enhance his affluence.

The construction of a data collection relative to SIR systems is basically a major challenge. For this issue, we led a technical study of a set of social networks API that are likely to be the most known. networks don't present yet API for developers, like ResearchGate. Some social In the extraction process, the majority of social networks use the OAuth 2.0 for connection and authentication authority, like Twitter, Youtube, Google+, LinkedIn, and Foursquare. Delicious and CiteUlike require basic http authentication, while some other networks need API keys for authentication (Last.fm, Flickr). REST is the common API used to have access to resources, and the result is always a JSON or XML file. Actually, this study is our way to construct a data collection suitable for Social IR.

Temporality is a fundamental issue and the most central aspect in social content. This factor is being investigated in several works [10][6][11]but still presents new contribution areas. Temporal aspect supports the eventual and permanent evolution of users' tastes, preferences and behaviors. Indeed, information appreciated by users now may not remain the same after a moment. Besides, trend events attract users' attention at a specific moment and are no more important after a while. Thus, Social IR systems should be adapted to this evolution. The same as for Social IR systems, the freshness of the information is essential in recommendation systems. So, to enhance recommendation quality, temporal factor should be considered.

Social approaches present certainly some limitations. A big challenge is to map user's accounts across social networks [19], and to predict missing social information, by combining for example social information and the social graph [20], in order to have an enriched social user profile. We will make a deep study in works dealing with these challenges.

6. Conclusion

In this paper, a review of different aspects of Social IR is proposed. We presented a classification of Social IR approaches into three main categories, based on social information used. We also posed a study of Social recommendation systems. Then, we referred to user profile models proposed in Social IR studies, and specially the social dimension. In this respect, works included in this review reflect how deep the impact of social content in IR and recommendation process is. Furthermore, we discuss different aspects of Social IR. As for coming studies, we start the process of user profile construction, based on temporal social signals.

7. References

[1] A. Dridi and M. Kacimi. "Information Retrieval Framework based on Social Document Profile". CAiSE(Forum/DoctoralConsortium), 2014.

[2] C. Bouhini, M. Géry and C. Largeron. "Integrating user's profile in the query model for Social Information Retrieval". Eighth International Conference on Research Challenges in Information Science (RCIS) ,1-2. IEEE, 2014.

[3] D. Tchuente, M.F. Canut, N. Jessel, A. Péninou and F. Sèdes. "A community-based algorithm for deriving users' profiles from egocentrics networks: experiment on Facebook and DBLP". Social Network Analysis and Mining, 3(3), 667-683, 2013.

[4] H. Xie, X. Li, T. Wang, L. Chen, K. Li, F. L. Wang, Y. Cai, Q. Li and H. Min. "Personalized Search for Social Media via Dominating Verbal Context." Neurocomputing, 172, 27-37, 2016.

[5] I. Badache and M. Boughanem. "Document Priors Based On Time-Sensitive Social Signals." ECIR, 617-622. Springer,2015.

[6] J. Chen, W. Geyer, C. Dugan, M. Muller and I. Guy. "Make new friends, but keep the old: recommending people on social networking sites". InProceedings of the SIGCHI Conference on Human Factors in Computing Systems, 201-210, ACM, 2009.

[7] J. Hannon, M. Bennett and B. Smyth. "Recommending twitter users to follow using content and collaborative filtering approaches". In Proceedings of the fourth ACM conference on Recommender systems, 199-206, ACM, 2010.

[8] K. Han, J. Park and M. Y. Yi."Adaptive and multiple interest-aware user profiles for personalized search in folksonomy: A simple but effective graph-based profiling model". In 2015 International Conference on Big Data and Smart Computing (BIGCOMP), 225-231, IEEE, 2015.

[9] L. Ben Jabeur, L.T. Lechani and M.Boughanem. "Intrégration des facteurs temps et autorité sociale dans un modèle bayésien de recherche de tweets". CORIA, 301-316, 2012.

[10] M.F Canut, S. On-at, A. Péninou and F. Sèdes. "Time-aware Egocentric network-based User Profiling". In 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 569-572, IEEE, 2015.

[11] M. Hafsi, M. Géry and M. Beigbeder. "LaHC at INEX 2014: Social Book Search Track". Working Notes for CLEF 2014 Conference, 2014.

[12] M. R. Bouadjenek, H. Hacid and M. Bouzeghoub. "Social networks and information retrieval, how are they converging? A survey, a taxonomy and an analysis of social information retrieval approaches and platforms". Information Systems, 56, 1-18, 2016.

[13] N. Ramesh and J. Andrews. "Personalized Search Engine using Social Networking Activity". Indian Journal of Science and Technology, 8(4), 301, 2015.

[14] N. Ould Amer, P. Mulhem and M. Géry. "Recherche de conversations dans les réseaux sociaux: modélisation et expérimentations sur Twitter". Conférence en Recherche d'Infomations et Applications-12th French Information Retrieval Conference, 2015.

[15] P. Li, B. Wang, W. Jin, J. Nie, S; Zhiwei, Z. and B; He. "Exploring categorization property of social annotations for information retrieval". In Proceedings of the 20th ACM international conference on Information and knowledge management, 557-562, ACM, 2011.

[16] S. Chelaru, C. Orellana-Rodriguez and I. Sengor Altingovde. "Can social features help learning to rank youtube videos?". International Conference on Web Information Systems Engineering, 552-566, Springer Berlin Heidelberg, 2012.

[17] S. Bao, X. Wu, B. Fei , G. Xue , Z. Su and Y. Yu . "Optimizing web search using social annotations". In Proceedings of the 16th international conference on World Wide Web, 501-510, ACM, 2007.

[18] X. Wang, H. Liu and W. Fan. "Connecting users with similar interests via tag network inference". In Proceedings of the 20th ACM international conference on Information and knowledge management, 1019-1024, ACM, 2011.

[19] E. Raad, R. Chbeir and A. Dipanda."User profile matching in social networks". In 13th International Conference on, 297-304, IEEE, 2010.

[20] A. Mislove, B. Viswanath, P. Krishna Gummadi and P. Druschel "You are who you know: inferring user profiles in online social networks". In Proceedings of the Third ACM International Conference on Web Search and Web Data Mining, 251-260, 2010.