Optimizing Recommendation in Collaborative E-Learning by Exploring DBpedia and Association Rules

Samia Beldjoudi[1], Hassina Seridi[2]

[1]Superior School of Industrial Technologies, Annaba, Algeria
[1, 2]Laboratory of Electronic Document Management LabGED Badji Mokhtar University, Annaba, Algeria
[1]s.beldjoudi@epst-annaba.dz [2]seridi@labged.net

Abstract. Social tagging activities allow users to add free annotations on resources to express user interests, preferences and automatically generate folksonomies. This paper demonstrates how structured content available through DBpedia can be leveraged to support recommendation of resources in folksonomies. A limitation of resources' recommendation is the content overspecialization conducting in the incapability to recommend relevant resources different from the ones that the learner already knows. To address this issue, we proposed to take advantage of the richness of the open and linked data graph of DBpedia and association rules to learn learners' behavior. The proposed approach demonstrates the efficiency of using DBpedia to enhance diversity and novelty when recommending resources to users in folksonomies. The basic idea is to iteratively explore the RDF data graph to produce novel and diverse relevant recommendations.

Keywords: Collaborative E-learning, Recommendation, DBpedia, Diversity, Novelty.

1 Introduction

Social tagging systems have achieved a great success over the web in the last years, especially in recommendations approaches. The problem of a precise recommender system is that the entire set of recommended resources may be obvious as one considers the case of a film recommendation algorithm that only returns films of the same actor. To overcome this problem, novelty and diversity should be also considered in the evaluation of a recommender system, as precision only offers an incomplete description of the system's effectiveness.

The main focus of our study is how to exploit the semantic aspect of DBpedia to enhance resource recommendation within social tagging systems. We propose a new method for analyzing learner profiles according to their tagging activities in order to

improve the recommendation of resources. The effectiveness of results depends on the resolution of social tagging drawbacks. In our process, we demonstrate how we can reduce the tags ambiguity problem by taking into account social similarities calculated on folksonomies combined with similarities between resources in DBpedia. We used also the force of Linked Open Data (LOD) to enhance resource recommendation by exploring the interlinked entities in LOD cloud. We base up on the iterative exploration of the DBpedia graph to obtain novel and diverse recommendations that should satisfy the learner and create the effect of surprise by recommending resources that the user did not expect at the beginning.
This paper is organized as follows: Section 2 is an overview of the main contributions related to our work. Section 3 is dedicated to the presentation of our approach. In section 4 we present and discuss the results of some experiments we conducted to measure the performance of our approach. Conclusion and future works are described in Section 5.

2 Related works

Social web based approaches, like folksonomies, have achieved a high level of improvement even in E-learning practice. In this section, an overview about some contributions attached to this field is proposed. [Kopeinik et al., 2017] investigated the application of two tag recommenders that are inspired by models of human memory. The authors find that displaying tags from other group members helps significantly in semantic stabilization in the group, as compared to a strategy where tags from the students' individual vocabularies are used. In [Beldjoudi et al., 2016], the authors proposed a new approach for personalizing and improving resources retrieval in collaborative learning with tackling tags ambiguity and event detection impact on resourced retrieved by ranking. In another contribution [Beldjoudi et al., 2017] proposed a method to analyze user profiles according to their tags in order to predict interesting personalized resources and recommend them. The authors proposed a new approach to reduce tag ambiguity and spelling variations in the recommendation process by increasing the weights associated to web resources according to social similarities. They base upon association rules for discovering interesting relationships among a large dataset on the web. [Karabadji et al., 2018] proposed to focus mainly on the growing of the large search space of users' profiles and to use an evolutionary multi-objective optimization-based recommendation system to pull up a group of profiles that maximizes both similarity with the active user and diversity between its members. In such manner, the recommendation system will provide high performances in terms of both accuracy and diversity. In our work we want to leverage the social and semantic web in order to enhance educational resources recommendation in collaborative e-learning.

3 Approach description

In this paper, we propose a method to analyze learner profiles according to their tags in order to predict interesting personalized resources and recommend them. We argue that the automatic sharing of resources strengthens social links among learners and we exploited this idea to reduce tag ambiguity in the recommendation process by increasing the weights associated to web resources according to social similarities. We based upon association rules that are a powerful method for discovering interesting relationships among a large dataset on the web. Our goal was to find correlations between tags, i.e. to find tags frequently appearing together, in order to extract those which are not used by one particular learner but which are often used by other users close to him in the social network.

The effectiveness of the recommendation depends on the resolution of the problems of folksonomies. In our approach we tackle the problems of tag ambiguity, diversity and novelty. To resolve the problem of tag ambiguity in recommendation, we propose to measure the similarity between learners to identify those who have similar preferences and therefore adapt the recommendation to learner profiles.

- ***First step:*** For each extracted association rule (Tags $A \rightarrow Tags\ B$) whose antecedent applies to an active learner lx, we measure the similarities between this learner and the learners of his social network who use the tags occurring in the consequent of the rule. The resources associated to these tags are recommended to the learner depending on these similarities. To measure similarity between two learners ($l1$ and $l2$), both are represented by a binary vector representing all their tags and we compute the cosines similarity between the two vectors.

- ***Second step:*** To avoid the cold-start problem which generally results from a lack of data required by the system in order to make a good recommendation, when the learner of the recommender system is not yet similar to other users, we propose to exploit semantic links between resources in DBpedia. DBpedia can be a reliable and rich source of content information that supports recommender systems to overcome problems, such as the cold-start problem and limited content analysis that restrict many of the existing systems, by building on a robust measurement of the similarities between resources using DBpedia. In this approach, we use the Linked Open Data to assess the similarity between folksonomies resources using their corresponding resources on DBpedia (i.e. we measure the similarity between the resources that would be recommended by the system, as related to a tag occurring in the consequent of an association rule, and those that are already recommended to the learner). The similarity between two resources is calculated using Jaccard index.

In another hand, when using a recommender system such as those of online stores, the results are mainly expected by the users. In this case, it is clear that the recommendation is not very helpful in the sense of the lack of diversity and novelty. To solve this dilemma in folksonomies-based collaborative learning, we propose extracting the most popular features found in the resources-based learner profile (i.e. the characteristics that interest the learner when they tag their resources) and then explore the LOD to extract resource linked with these features.

Let us consider a learner profile composed from the resources (*R1, R2, R3 and R4*), Thus the intersection between the resources' features must be calculated (*R1* ∩ *R2* ∩ *R3* ∩ *R4*), this is done because we want to extract the most popular characteristics that interest the learner when they choose tagging their resources. Then for each feature (*Pi*) in the result of intersection we will explore the LOD graph in the first level to extract other resources (*R5*) having these features or having a direct/ indirect link with these later (*R6, R7* resp).

Supposing that (*R1* ∩ *R2* ∩ *R3* ∩ *R4*) = {[domain: informatics]; [author: …]; [year: …]; [edition: …]…}. By exploring the LOD graph we find that the resource "informatics" is linked with other resources (for example: "University, Formation, Bio-Informatics…") via the predicates (*P1, P2, P3*...). In its turn the resources "University, Formation, Bio-Informatics…" are linked via other predicates (*Pj*) with other resources (for example: "Boston University…"). Therefore, it appears relevant to recommend some courses of the Boston University to the current user.

Our approach is based on the iterative exploration of the DBpedia graph, where each step depends on the result of the previous steps. In order to obtain relevant and personalized recommendations for each learner, we calculate the occurrence number of the {domain, author, year, edition…} characteristics and then we choose the ones that best reflect the learner interest to exploit them later in the exploration of the RDF graph of DBpedia.

The purpose of the graph exploration is to obtain recommendations that should not only satisfy the learner but also to have diversity and a novelty in the recommendation, to create the effect of surprise by recommending resources that the learner did not expect at the beginning. The learner evaluates the recommended resources in real time in each iteration. The process stops when none of the recommended resources has satisfied the user.

If the learner liked at least one resource among those in the proposed list, in the second iteration, we focus on these ones. Thus, we re-explore the LOD graph again starting from these items by using the query language SPARQL to return more educational resources connected with them; this technique allows us to propose a list of diverse and novel resources to ensure the surprise effect.

The real-time evaluation process as well as the exploration of the graph is iterative. At each iteration, we explore the graph based on the positive ratings assigned to the resources previously recommended. Indeed, the evaluation is an essential step to determine the new pattern of requests for the re-exploration of the graph to generate another list of recommendations. At each step, we propose to the user 10 resources, if he assigns a rating more or equal to three, we consider that he liked the recommended resource, and so we record it in his profile, otherwise we move to another resource.

After evaluating the 10 resources, the program suggests to the user to recommend after the 10 resources have been evaluated, the program suggests to recommend some more to the user. If he accepts then another list of resources is generated from his profile, otherwise, we stop and return the list of resources liked. With this method, we ensure that the recommended list of resources is diverse, where every user can obtain diverse resources even if they do not appear in the profile of his neighbors in the social network.

4 Experimental Results

In this section, experiment over a popular dataset is described and results are analyzed and discussed. The dataset exploited in our test is del.icio.us. In this experiment, we were interested in data generated from users who tagged resources about education. Thus, our database comprises 1128 tag assignments involving 95 users, 432 tags containing ambiguous tags and 314 resources.

4.1 Experimental Methodology

To evaluate the quality of a recommender system, we must demonstrate that the recommended resources are really being accepted and added by the users. Because the knowledge of this information requires asking the users of the selected databases if they appreciated the proposed set of resources, which is impossible in our case because we do not have access to this community, we have used a cross-validation where we have randomly removed some resources from the profile of each user, and we applied our approach on the remainder dataset in order to show if it can recommend the removed resources to their corresponding users or not. If it is the case, so we can conclude that our approach enables to extract the user preferences.

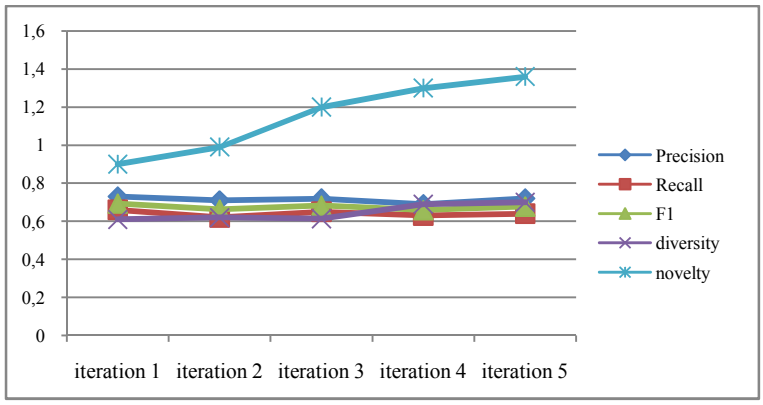

Fig. 1. Average precision, recall, F1, diversity and novelty of the recommendations

The curve presented in figure 1 show average values of precision, recall, F1, diversity and novelty measures in the five iterations. We notice that the precision achieved a good value in all iterations, this is due to the fact that the system recommends exactly the items wanted by the user i.e. those that match his profile. Sometimes the system begins to deteriorate in terms of precision but always with a value that exceeds 0.6. This decrease is quite normal since the system begins to recommend items according to different attributes (domain, year ...) which is known as diversity of recommendation. Learners sometimes accept the recommended resources and other times it was not the case. Recall and F1 measure achieved all both good values in the all iterations.

To calculate individual diversity and novelty, we used the metrics proposed in [Zhang and Hurley, 2009] and [Vargas, 2014] respectively. Figure 1 showed promising values of both diversity and novelty in the five iterations. This demonstrates the importance of DBpedia to extract more diversified and novel resources in the recommendation. It is clear that the effectiveness of recommendation depends of preserving both precision and diversity. Results demonstrate that our approach preserving both them in all iterations.

5 Conclusion

In this contribution we have exploited the strength of social aspect in folksonomies to let members in the community benefit from the educational resources tagged by other users, based on the recommendation of resources. The proposed approach is based on DBpedia, the objective was to overcome the problem of diversity and novelty in recommendation. Primary results show also the utility of exploring LOD graph in ensuring diversity when recommending personalized educational resources in social tagging systems. In order to continue and improve our work, we aim at using others principles like event detection, for example, to help capturing and analyzing the behavior of learners when new events come, this can improve recommendation and even resources ranking.

References

1. Beldjoudi, S., Seridi, H. & Benzine, A.: The Impact of Social Similarities and Event Detection on Ranking Retrieved Resources in Collaborative E-Learning Systems. In: Koch F., Koster, A., Primo, T., Guttmann, C. (eds) Advances in Social Computing and Digital Education. CARE 2016, SOCIALEDU 2016. Communications in Computer and Information Science, vol 677. Springer, Cham (2016)
2. Beldjoudi, S., Seridi, H. & Faron-Zucker, C.: Personalizing and Improving Resource Recommendation by Analyzing Users Preferences in Social Tagging Activities. Computing and Informatics 36(1): 223-256 (2017)
3. Karabadji, NEI., Beldjoudi, S., Seridi, H., Aridhi, S. & Dhifli, W.: Improving memory-based user collaborative filtering with evolutionary multi-objective optimization. Expert Syst. Appl. 98: 153-165 (2018)
4. Kopeinik, S., Lex, E., Seitlinger, P., Albert, D. & Ley, T.: Supporting collaborative learning with tag recommendations: a real-world study in an inquiry-based classroom project. In: LAK '17 Proceedings of the Seventh International Learning Analytics & Knowledge Conference Pages 409-418 (2017).
5. Vargas, S.: Novelty and diversity enhancement and evaluation in recommender systems and information retrieval. In Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR '14, pages 1281–1281, New York, NY, USA, (2014).
6. Zhang, M. and Hurley, N.: Novel item recommendation by user profile partitioning. In 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, volume 1, pages 508–515, Sept (2009).