Extracting and Exploiting Topics of Interests from Social Tagging Systems

Felice Ferrara and Carlo Tasso

University of Udine, Via delle Scienze 206, 33100 Udine, Italy
{felice.ferrara,carlo.tasso}@uniud.it

Abstract. Users of social tagging systems spontaneously annotate resources providing, in this way, useful information about their interests. A collaborative filtering recommender system can use this feedback in order to identify people and resources more strictly related to a specific topic of interest. Such a collaborative filtering approach can compute similarities among tags in order to select resources associated to tags relevant for a specific interest of the user. Several research works try to infer these similarities by evaluating co-occurrences of tags over the entire set of annotated resources discarding, in this way, information about the personal classification provided by users.
This paper, on the other hand, proposes an approach aimed at observing only the set of annotations of a single user in order to identify his topic of interests and to produce personalized recommendations. More specifically, following the idea that each user may have several distinct interests and people may share just some of these interests, our approach adaptively filters and combines the feedback of users according to a specific topic of interest of a user.

Keywords: Recommender systems, collaborative filtering, social tagging, adaptive, personalization

1 Introduction

The social Web is constituted by services aimed at promoting socialization and communication among users which are allowed to create, share, and organize information. In particular, social tagging systems deal with sharing and organizing resources: each user can both share her resources with the other peers and assign some labels (named tags) to resources in order to simplify future retrieval.

The collection of resources tagged by a user is called *personomy*. A personomy is a personal classification of resources which are interesting to the specific user and it can be explored by other users by means of tags. However, the classification generated by a user usually is not very precise since users freely choose tags without following rules to avoid ambiguities and, for this reason, the tags applied by a user cannot have a clear semantic. On the other hand, the union of all personomies (referred as *folksonomy*) can be analyzed in order to infer semantic relations taking into account how the community of Web 2.0 users combines tags.

This analysis can extract semantic relations among tags which emerge from the collaborative social work of users. Such 'social semantic' relations among tags can potentially be used in order to enhance the access to information by empowering, for example, information filtering systems such as Collaborative Filtering (CF) recommender systems.

CF recommender systems implement the word-of-mouth mechanism: they simulate the behavior of humans which usually share their opinions with friends when they need to take a decision. In a user-based CF recommender system [1], the simulation of this social process is exploited in two steps. The first step is usually referred as '*neighbor selection*' and it is the phase when the system identifies the set of people which share an interest with the target user (referred in this paper as '*active user*'). In the second step, the system generates the list of recommendations by combining feedback (i.e. information about what is relevant for a user) provided by the best (most similar) neighbors: resources more relevant to the community of the best neighbors are suggested to the active user. The rationale of this model is that users with a common interest are interested in the same resources and could be interested in the same information also in the future. In order to accomplish this mechanism, a CF system models the opinions of each user by a vector which contains her ratings (rating vector). In this way, the similarity between two users depends on the similarity between the two rating vectors associated to them.

However, this basic model does not take into account the fact that each user may be interested in more than one Topic of Interest (ToI). This can reduce the accuracy of both the first and the second steps exploited by a user-based CF system. In fact, by computing just one similarity value for each pair of users, the system ignores that the active user could be similar to a set of users for a certain specific topic, but she could share another ToI with a completely different neighborhood. This issue has been recognized also in [2], where the authors proposed the BIPO (Best Item Per Overlapping) framework aimed at finding for each resource a locally adaptive neighborhood. Following the 2-step workflow described above it is straightforward to recognize a further criticality, since the system considers all the feedback provided by the neighborhood, without taking into account that such feedback could partly refer to different ToIs. This means that the system could suggest resources which are interesting to the community of neighbors but which are completely not related to the ToIs of the active user.

In this work we describe an approach aimed to overcome these limitations in social tagging systems by using social semantic relations among tags.

More specifically, we face the first criticality by analyzing the personomy of the active user: tags applied by the active user are grouped according to the strategy used by the active user to combine tags. Given a user, this first analysis produces a set of clusters of tags where each cluster contains tags which have been frequently used together to annotate resources. Different clusters of tags are considered to correspond to different ToIs of the active user. Given the list of the ToIs of the active users, we select the resources associated to a specific ToI (associated to the tags belonging to the same cluster associated to that ToI) in order to detect a set of neighbors (users) interested in that specific ToI. This is accomplished by selecting the users which have tagged the specific resources associated to the considered ToI.

To tackle the second criticality we take into account the tags utilized by neighbors. More specifically, given a neighbor, the collection of her tagged resources is filtered discarding the resources which have not been labeled by tags relevant to the specific ToI. In order to compute the relevance of the tags exploited by a neighbor, we consider the set of resources that the neighbor shares with the active user for the specific ToI: tags applied on shared resources are more 'trustworthy' than others for finding new items related to that specific ToI.

The choice of extracting relations among tags by analyzing the set of annotations of the specific user (instead of evaluating the co-occurrences over the entire collection of annotations) is aimed at filtering the user feedback according to the personal classification provided by the active user in order to better fit his specific informative needs.

The paper is organized as follow: in Section 2 related works are discussed, the construction of the user profile is the object of Section 3, Section 4 deepens the discussion on the mechanism used to compute recommendations for a specific topic of interest, the ongoing evaluation activities are illustrated in Section 5 and, finally, conclusions and future works conclude the paper in Section 6.

2 Related Work

The growing usage of Web 2.0 applications and the social dimension introduced by these tools are appealing features for researchers interested in systems devoted to personalize the access to Web information.

In fact, the collaborative work of Web 2.0 users in social tagging systems have been analyzed to extract social semantic relations since semantic information generated by the social work of users can be used to infer similarities among tags, resources and users [3]. These relations are inferred by evaluating co-occurrences among tags and the majority of the work in the literature takes into account co-occurrences over the entire folksonomy, i.e., by collapsing the annotations produced by all the users.

The resulting 3-dimensional relation involving tags, users and resources cannot be trivially managed to infer social semantic relations since it merges relations among objects of the same type as well among objects of different types. A common approach to overcome such limitation is to project projecting the 3-dimensional relationship among users, items and tags into two 2-dimensional relationships by hiding information about one dimension. Following this approach, the social semantic relations can be described by three matrices:

1. the *Tag-Resource* (TR) matrix which describes the two-way relation between tags and resources. Each row of the matrix, associated to a tag, is a vector, which counts how many times a tag has been applied on each resource.

2. The *User-Tag* (UT) matrix which describes the two-way relation between users and tags. Each row of the matrix, associated to a user, is a vector, which counts how many times a user applied each tag.
3. The *User-Resource* (UR) matrix which describes the two-way relation between users and resources. Each row of the matrix, associated to a user, is a unary vector which allows to describe if a user tagged or not a resource.

As we described in [4], several approaches aimed at inferring similarities among users, tags, and resources by using these matrices have been proposed and integrated in recommender systems.

For example, by computing the cosine similarity between the rows of TR matrix associated to the tag t_i and t_j the similarity between the two tags, $sim(t_i, t_j)$, is quantified according to the number of times the two tags co-occur on the same resources. On the other hand, the similarity between the users u_i and u_j, $sim(u_i, u_j)$, can be estimated by computing the cosine similarity between the rows of the UT matrix associated to the two specific users.

These two measures of similarity have been used in SocialRanking [5], a user based CF recommender system which, given a tag t_k, assigns a score $R(p)$ to each resource p for the user u as

$$R(p, t_k) = \sum_{u_i} \left(\sum_{t_x} sim(t_x, t_k) \right) \cdot (sim(u, u_i) + 1)$$

where t_x is a tag that the user u_i assigned to the resource p. In this case, given a user u and a tag t_k, the relevance of a resource p is higher if

- p has been tagged by people which used many tags applied by u;
- p has been labeled by tags which have been often used (by the entire community of users) to classify the same resources tagged with the tag t_k.

We also proposed a user based CF recommender framework [6] which groups 'similar' tags utilized by the active user for modeling his interests. In this CF framework, the similarity among two tags depends on the number of times the two tags co-occur on the same resources in the folksonomy and the relevance of a new resource is computed by evaluating also the relevance of the tags assigned to it. However, both these approaches uses tag relations introduced by all users without taking into account how the active user specifically combines tags. In fact, a folksonomy merges the annotations provided by the entire community without taking into account specific personal interests and tagging strategies of different users. On the other hand, a personomy embodies information strictly related to the personal interests of a user. This means that the analysis of a personomy can reveal relations among tags which are meaningful just for that user.

Authors of [7] proposed a framework to catch semantic relations built by each user. This CF framework is based on a community detection algorithm for clustering tags that the active user applied frequently together: each user is modeled by sets of tags and the similarity among users is computed by evaluating the similarity among the sets of tags they used. However, users that share an interest could label the same concept using different set of tags, and this limits the possibility to identify similarities among users.

In this work we also group tags utilized by the active user by taking into account the semantic relations built by her. However, differently from the approach described in [7], we recognize that some tags are used to refer also to other different interests. This means that some tags utilized by the active user are more relevant than others for describing a specific ToI. We also face the limitation of neighbor selection by adaptively choosing a set of neighbors interested in the specific ToI. To reach this aim we take into account the set of resources tagged within a specific ToI since we recognize that different users may use different tags to refer to the same concept. Finally, we also filter feedback from neighbors by using tags (not necessarily the tags applied by the active user) that the neighbor applied on resources relevant for the specific ToI.

3 Identifying the Topic of Interests

The set of ToIs $\left\{ToI_{au}^{1}, \ldots, ToI_{au}^{t}\right\}$ identified within the personomy of the active user, constitutes her interest profile. The ToI_{au}^{k} is defined as

$$ToI_{au}^{k} = \left(T_{au}^{k}, R_{au}^{k}\right)$$

where

$$T_{au}^{k} = \left\{(t_1, w_{t_1}^{k}), \ldots, (t_n, w_{t_n}^{k})\right\}$$

is the set of weighted tags used by the active user au to annotate his resources in the topic k and

$$R_{au}^{k} = \left\{(r_1, w_{r_1}^{k}), \ldots, (r_m, w_{r_m}^{k})\right\}$$

is the set of resources tagged by the user au with the tags in T_{au}^{k}.

More specifically, T_{au}^{k} is defined on a set of semantically related tags

$$tag(T_{au}^{k}) = \{t_1, \ldots, t_n\}$$

applied by the active user, where two tags are considered to be in a semantic relation if the active user has applied them together to classify one or more resources. The weight associated to each tag represents the relevance of the tag with respect to that ToI and it is used to compute the relevance of each resource $res(R_{au}^{k}) = \{r_1, \ldots, r_m\}$ tagged by the active user within that ToI.

In order to identify the semantic relations defined by the active user we analyze her personomy throwing out information about the tagged resources. In particular, given the personomy of the active user we build an undirected weighted graph P where: each node represents a tag; an edge connects two tags if they have been used together to label one or more resources; an edge connecting two tags is weighted by the number of times two tags have been used together.

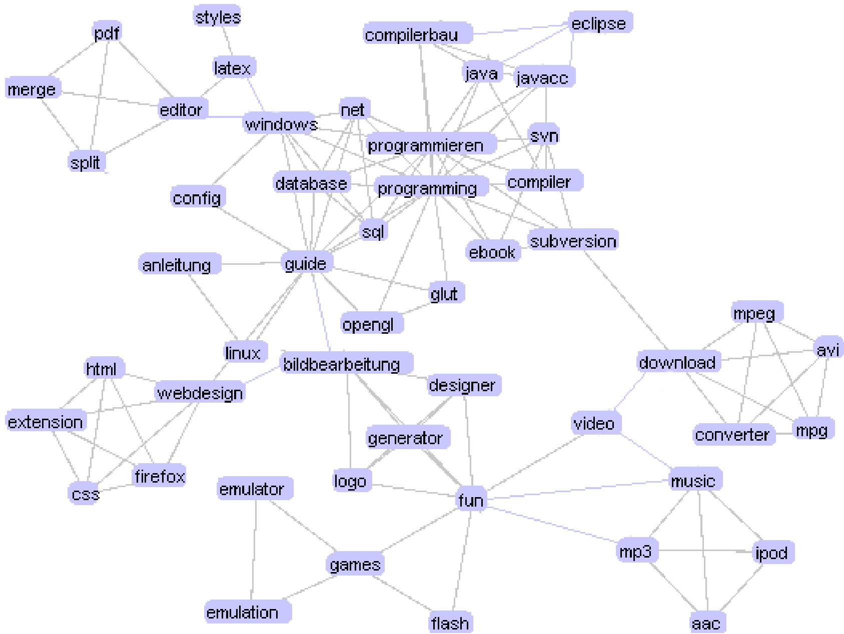

Fig. 1. An example of the graph P for a user of the BibSonomy system

Figure 1 shows the graph P for one of the user of the BibSonomy social tagging system [8], where we do not show the weights associated to edges to make the graph readable.

Given the graph representation of a personomy, we apply a graph clustering technique for grouping tags with a shared semantic. In particular, we follow the idea proposed in [9] where a node (representing a tag in our model) may be in more than one cluster identifying, in this way, overlapping clusters of tags. This clustering technique identifies clusters of tags by identifying subgraphs from the starting graph P, where each subgraph G maximizes the following fitness property

$$f_G = \frac{K_{in}}{(K_{in} + K_{out})^{\alpha}}$$

where K_{in} is the sum of the weights of the edges which connect two tags belonging to G, K_{out} is the sum of the weights which connect tags in G with the rest of the graph, and α is a parameter which controls the size of clusters. In other words, the fitness of a subgraph increases when we add to the subgraph a tag that the user has exploited frequently in co-occurrence with tags in the subgraph and rarely with the other. The algorithm builds, for a given node, a cluster of tags by adding, at each step, the node which maximizes the following fitness function $f_G^a = f_{G+a} - f_{G-a}$,where $G+a$ $(G-a)$ is the subgraph obtained by adding (removing) the node a to the subgraph G. The process which adds tags to the cluster stops when there are not tags with a positive fitness value.

Using this method, our approach defines the cluster for the most used tag. Then, the approach identifies the cluster for the most used tag which has not been yet included in a cluster. The clustering algorithm ends when each tag is at least in one cluster. At the end, each subgraph detected by the clustering phase contains the set of tags associated to a certain ToI for the active user.

However, given a subgraph G^k, some of the tags in G^k are less relevant than others since they possibly are used also for referring to other different ToIs. Therefore, we associate a weight w_t^k to each tag t in the subgraph G^k by computing the betweeness centrality [10] of t in the specific subgraph identifying, in this way T_{au}^k. Applying this strategy on the example showed above for the most used tag 'programming' with $\alpha = 1.0$ we obtain the weighted set of tags: (programming, 1.0), (programmieren, 0.77), (guide 0.54), (java, 0.45), (javacc, 0.40), (compilerbau,0.36), (windows,0.31), (database, 0.27), (net,0.27), (sql, 0.27), (glut,0.22), (opengl,0.22), (eclipse,0.22), (compiler, 0.18), (svn, 0.18), (subversion 0.18), (ebook, 0.18), (anleitung, 0.09), (linux,0.09).

The set T_{au}^k is used to infer the set R_{au}^k such that $res(R_{au}^k)$ is composed by the resources that the active user labeled by tags in $tag(T_{au}^k)$ and the weight w_r^k for the resource r is equal to the maximum weight of tags in $tag(T_{au}^k)$ which the active user associated to r.

4 Recommending resources for a ToI

This section focuses on the recommendation process by describing how the approach filters and ranks resources for a specific $ToI_{au}^k = (T_{au}^k, R_{au}^k)$. We will show how the set of weighted resources R_{au}^k can be used to select adaptively the neighbors (Section 4.1) and then how feedback from neighbors are filtered and combined (Section 4.2).

4.1 Adaptive neighbor selection

Given the $ToI_{au}^k = (T_{au}^k, R_{au}^k)$ of the active user, the set of weighted resources R_{au}^k is used to filter the set of neighbors for the ToI. In particular, the approach identifies people interested in the specific ToI by taking into account only the users who tagged the resources in $res(R_{au}^k)$. We assume that people interested in ToI_{au}^k share with the active user relevant resources within the specific ToI. For this reason, let $R_{shared}(u, R_{au}^k)$ be the set of resources that the user u share with the active user in $res(R_{au}^k)$, we compute how much the specific interest of the active user is matched by the neighbor u by computing the following

$$InterestMatch(u, ToI_{au}^k) = \frac{\sum_{r_i \in R_{shared}(u, R_{au}^k)} w_{r_i}^k}{\sum_{r_i \in res(R_{au}^k)} w_{r_i}^k}$$

The logic behind this formula is that higher is the number and the relevance of the resources in R_{au}^k that the neighbor u tagged, higher is the interest of u in the specific ToI. By using the InterestMatch function, we can filter the set N_{au}^k of neighbors interested in ToI_{au}^k.

4.2 Filtering and combining feedback for the ToI

The neighbor selection phase takes in account only resources in the ToI of the active user. In order to produce recommendations we need to identify new resources labeled by neighbors which are related to the specific ToI. Therefore, to do this, we consider the tags that the neighbor u applied on the set of shared resources $R_{shared}(u, R_{au}^k)$: the resources labeled by u with these tags are considered relevant for the specific ToI. e follow the idea that, some tags in the personomy of the neighbor u are more trustworthy than others for finding relevant resources for the ToI. In fact also the neighbor may have several ToIs and, for this reason, we are interested in discovering which tags utilized by u better account for the ToI of the active user. We consider more trustworthy the tags which have been used by the neighbor to label many relevant resources within ToI_{au}^k and, specifically, we measure the trustworthiness of a tag t_j in the collection of the neighbor u with respect to ToI_{au}^k as follow:

$$trustworthiness_u(t_j, ToI_{au}^k) = \frac{\sum_{r_i \in R_{shared}(u, R_{au}^k)} w_{r_i}^k \cdot \phi(u, t_j, r_i)}{\sum_{r_i \in R_{shared}(u, R_{au}^k)} w_{r_i}^k}$$

where $\phi(u, t_j, r_i) = 1$ if the user u applied the tag t_j on the resource r_i, 0 otherwise. Following the principle that trustworthy tags are associated to relevant resources of the neighbor u, we assign an higher relevance to resource labeled by more trustworthy tags. Specifically, we compute $rel_u(r_j, ToI_{au}^k)$, which is the relevance of the resource r_j in the personomy of the neighbor u with respect to ToI_{au}^k, as the highest trustworthiness associated to tags that the neighbor u assigned to r_j.

Finally, the relevance of a resource r_j for the active user with respect to ToI_{au}^k is computed summing the relevance of r_j over the collections of the neighbors N_{au}^k as follow:

$$rel(r_j, ToI_{au}^k) = \sum_{u \in N_{au}^k} InterestMatch(u, ToI_{au}^k) \cdot rel_u(r_j, ToI_{au}^k)$$

This allows to produce the ranked list of resources that are recommended to the active user.

5 Evaluation

Two main approaches can be used to evaluate the accuracy of the suggestions produced by a recommender system: a live approach and an off-line approach. In the first case, experimentation involves human participants which explicitly provide information about their interests. The recommender system can, in turn, build user profiles and produce a set of recommendations for each user. Finally, the involved humans can personally judge if the suggested items match their interests. On the other hand, an off-line evaluation uses an historical dataset which contains the rate history (the tagging history) of a certain set of users. In this case, the dataset is divided into two chunks of data where a block of data is used as training set in order to build user profiles and the other block is used as test set, in order to compare the resources rated/tagged by the user to the ones predicted/suggested by the system. Both the live and the off-line approach have some advantages and shortcomings since live experimentations are more precise but are also more expensive while, on the other hand, off-line evaluation is cheaper but cannot be very precise. More specifically, the main limitation of off-line evaluation depends on the high number of unrated resources: the users in the dataset usually evaluated just a small part of the available resources. This means that the system can produce meaningful recommendations which have not been rated/tagged by the active user: therefore an off-line evaluation considers these recommendations as not relevant for the active user (i.e. he did not rate/tag them) and this lowers the estimated precision. Obviously, this prevents an effective evaluation of the recommender system. For this reason, we have planned to evaluate our approach by using both a live and an off-line approach. At the moment we have exploited only some off-line experimentations in order to demonstrate the validity of our technique.

More specifically, in this work, we are interested in evaluating if the idea of filtering feedback for a single topic of interest by using the semantic relations built by the active user and his neighbors can significantly improve the approaches in the literature. For this reason, we have implemented the SocialRanking recommender system [5], which similarly to our approach produces recommendations for a specific tag/topic, and we have used it as a baseline reference for performance measurement.

We used a dump of the BibSonomy system [11] to compare the results provided by the two recommender systems. More specifically, we divided the BibSonomy dataset into two chunks of data: the training set which includes all bookmarks until the first of January 2008; the test set which has the bookmarks from the first of January 2008 to the 31 December 2008. We created the user profile only for the m users who tagged at least 60 resources until the first of January 2008 and who tagged at least 10 resources during the 2008. This is reasonable since collaborative filtering approaches produce effective recommendations only when users rated a significant number of items [1].

For each user, we used our approach to identify his topic of interests. Then, we computed a set of recommendations for the ToI which contained the tag that he used most frequently in the subsequent period (i.e. in the test set).

We used the same tag as input in the SocialRanking case. Both the SocialRanking mechanism and our approach have been exploited for producing 10 recommendations, using feedback from the top 10 and from top 20 neighbors.

The quality of the computed recommendations produced by the two mechanisms was evaluated by adopting two measures, named respectively *hit-rate*(HR) and *average reciprocal hit-rank* $(ARHR)$, which have been used also in [12] to compare two collaborative filtering recommender systems. More specifically, the HR measure is defined as follow

$$\textit{hit-rate} = \frac{\textit{Number of hits}}{m}$$

where m is the total number of users considered in the evaluation and we count a hit when the system produces at least one correct recommendation (i.e. a recommendation for a resource that the active user has actually tagged in the subsequent period). Given the lists of recommendations for the m users produced by a recommendation mechanism, the hit-rate is a value in $[0, 1]$ which is higher when there is a larger number of users who received at least one recommendation for a resource that they will tag in the test period.

Table 1 shows that our approach has an higher HR both when 10 neighbors and 20 neighbors are used: according to this metric the approach proposed in this paper outperforms SocialRanking.

Table 1. Hit-rate with 10 and 20 neighbors

	HR (10 neighbors)	HR(20 neighbors)
SocialRanking	0.13	0.15
Our Approach	0.34	0.40

The main limitation of the HR measure is given by the fact that hits are evaluated regardless of their position, i.e, a hit that occurs in the first position of the list of recommendations is treated equally to a hit that occurs in the last position. In other words, the capability of the recommender system to better rank resources is not recognized. In order to face this limitation we also used the *ARHR* measure which is defined as

$$ARHR = \frac{1}{m} \sum_{i=1}^{h} \frac{1}{p_i}$$

where h is the number of hits and p_i is the position of *i-th* hit. $ARHR$ is still a value in $[0, 1]$ but it represents a measure of how well the recommender mechanism is capable to rank a hit in high-score positions. In Table 2 we show that our approach outperforms SocialRanking also when we use the $ARHR$ metric.

The main advantage of SocialRanking is that it can face the sparsity problem. Such limitation is due to the fact that users in CF filtering systems usually rate only a small part of the total number of items globally considered in the system and this makes harder the task of finding similarities on the basis of shared resources. Our approach can actually increase the sparsity since only a part of

Table 2. ARHR with 10 and 20 neighbors

	ARHR (10 neighbors)	ARHR (20 neighbors)
SocialRanking	0.05	0.07
Our Approach	0.14	0.19

the user feedback is used to produce the set of recommendations (i.e., only the part relevant to a specific ToI).

However, Table 1 and Table 2 show that by inferring topic of interests from the set of the user annotated resources we obtain a more adequate description of the user interests since the number of hits increases as well as the position of the hits in the lists of recommendations. This depends on the fact that by computing similarities among tags by using only the similarities inferred from the *Tag-Resource* matrix, SocialRanking cannot account personal tagging strategies, i.e., semantic relations which are not adopted by a significant number of users in the community. Moreover, by taking into account only the number of times two tags co-occur in order to infer their semantics, approaches (such as SocialRanking) ignore that the tagged resources embed the information able to clarify the meaning of tags. For this reason we compute similarities among users by also taking into account the number of shared resources: higher is the number of shared resources, higher is the probability that the users are using two set of (possibly distinct) tags to describe a shared ToI.

6 Conclusions

In this work we propose a novel method to improve the precision of recommendations computed by a user-based CF system for social tagging systems. In order to reach this aim we address two main open issues.

In particular, traditional user-based CF recommender systems do not take into account that each user may have several distinct interests and people may share just some of these interests. In order to face this limitation, our approach adaptively filters and combines the feedback of neighbors according to a specific topic of interest identified for the active user. More specifically, in our approach, a topic of interest of the active user is defined by a set of tags with a shared meaning. In order to identify it we do not use approaches based on the analysis of the entire folksonomy, since they tend to completely discard information about the specific user. On the other hand, we group the tags of the active user according to the personal tagging strategy adopted by the user.

The results show that the proposed approach is reasonable and it outperforms other approaches proposed in literature. At the moment, we are planning a more effective evaluation to validate this claim by exploiting off-line evaluations on new datasets and by exploiting an on-line evaluation.

Moreover, we are interested both in:

- extending the approach following some idea proposed in [9] in order to identify hierarchical organization of user interests;

– adding a more semantic layer by means of content/ontology based analysis.

This step could potentially empower our approach by extracting semantic information able to disambiguate and enrich the description of user interests, merging more strictly the social and the semantic perspectives.

References

1. Schafer, J., Frankowski, D., Herlocker, J., Sen, S.: Collaborative filtering recommender systems. In Brusilovsky, P., Kobsa, A., Nejdl, W., eds.: The Adaptive Web. LNCS (4321). Springer-Verlag, Berlin, Germany (May 2007) 291–324
2. Baltrunas, L., Ricci, F.: Locally adaptive neighborhood selection for collaborative filtering recommendations. In: Proc. of the 5th Int. Conf. on Adaptive Hypermedia and Adaptive Web-Based Systems, Hannover, Germany (July 2008) 22–31
3. Marinho, L., Nanopoulos, A., Schmidt-Thieme, L., Jaschke, R., Hotho, A., Stumme, G., Symeonidis, P.: Social tagging recommender systems. In Ricci, F., Rokach, L., Shapira, B., Kantor, B., eds.: Recommender Systems Handbook. Springer (2011) 615–644
4. Dattolo, A., Ferrara, F., Tasso, C.: On social semantic relations for recommending tags and resources using folksonomies. Human-Computer Systems Interaction. Backgrounds and Applications 2. In printing
5. Zanardi, V., Capra, L.: Social ranking: Finding relevant content in web 2.0. In: Proc. of the 2nd ACM Int. Conf. on Recommender Systems, Lausanne, Switzerland (October 2008)
6. Dattolo, A., Ferrara, F., Tasso, C.: Supporting personalized user concept spaces and recommendations for a publication sharing system. In: Proc. of the 17th Int. Conf. on User Modeling, Adaptation, and Personalization (UMAP 2009), Trento, Italy (July 2009) 325–330
7. Zhou, T., Ma, H., Lyu, M., King, I.: Userrec: A user recommendation framework in social tagging systems. In: Proc. of the 24th AAAI Conf., Atlanta, Geogia, USA (July 2010) 1486–1491
8. Benz, D., Hotho, A., Jaschke, R., Krause, B., Mitzlaff, F., Schmitz, C., Stumme, G.: The social bookmark and publication management system bibsonomy. The VLDB Journal **19**(6) (2010) 849–875
9. Lancichinetti, A., Fortunato, S., Kertsz, J.: Detecting the overlapping and hierarchical community structure in complex networks. New Journal of Physics **11**(033015) (2009)
10. Opsahl, T., Agneessens, F., Skvoretz, J.: Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks **32**(45) (2010)
11. BibSonomy: Bibsonomy dataset, dumps for research purposes. `http://www.kde.cs.uni-kassel.de/bibsonomy/dumps`
12. Deshpande, M., Karypis, G.: Item-based top-n recommendation algorithms. Transaction on Information Systems **1**(22) (2004) 143–177