A Semantic Model for Social Recommender Systems

Heung-Nam Kim[1,2], Andrew Roczniak[1], Pierre Lévy[1], Abdulmotaleb El-Saddik[2]

[1] Collective Intelligence Lab, University of Ottawa
{hnkim, roczniak}@mcrlab.uottawa.ca, pierr.levy@mac.com
[2] Multimedia Communication Research Lab, University of Ottawa
abed@mcrlab.uottawa.ca

Abstract. Social recommender systems, which have emerged in response to the problem of information overload, provide users with recommendations of items suited to their needs. To provide proper recommendations to users, social recommender systems require accurate models of characteristics, interests and needs for each user. In this paper, we introduce a new model capturing semantics of user-generated tags and propose a social recommender system that is incorporated with the semantics of the tags. Our approach first determines semantically similar items by utilizing semantic-oriented tags and secondly discovers semantically relevant items that are more likely to fit users' needs.

Keywords: Social Recommender System, IEML Semantic Model

1 Introduction

With the popularity of social tagging (also known as folksonomies), recently a number of researchers have concentrated on recommender systems with social tagging. Recommender systems incorporated with the tags can provide promising possibilities to better generate personalized recommendations [3]. However, if the systems do not take into consideration the semantics of tags themselves, they suffer from fundamental problems: polysemy and synonymy of the tags. Without the semantics of the tags used by users, the systems cannot differentiate the various social interests of the users from the same tags.

To address the discussed issues, we introduce a new model capturing semantics of user-generated tags and propose a social recommender system that is incorporated with the semantics of the tags. Our approach first determines similarities between items by utilizing semantic-oriented tags that is associated to tags that users collectively annotate, called *Uniform Semantic Locator* (*USL*) [4] and subsequently identifies semantically similar items for each item. Finally, we recommend items (*e.g.*, text, picture, video) based on the semantically similar items. The main contributions of this study toward social recommender systems can be summarized as follows: 1) We present a model for semantic-oriented social tagging by using IEML [4]. We illustrate how the model can be adapted and applied to existing social tagging systems. 2) We propose a method in semantic space that aims to find semantically similar items and discover (recommend) items semantically relevant to users' needs.

2 A Semantic Model for Social Recommender Systems

In this paper, we exploit Information Economy MetaLanguage (IEML) [4] for social recommender systems. Formally, folksonomy F is a tuple $F = \langle \tilde{U}, \check{T}, \tilde{I}, Y \rangle$ where $\tilde{U}$ is a set of users, $\check{T}$ is a set of tags, $\tilde{I}$ is a set of items, and $Y \subseteq \tilde{U} \times \check{T} \times \tilde{I}$ is a ternary relationship called tag assignments, respectively [2]. Beyond the tagging space, in our study, there is another space where the tags are connected to *USL*s according to their semantics. We label this space the IEML semantic space. According to IEML model, a *USL* is composed of a set of semantic categories of different layers, called catset [4]. Therefore an extended formal definition of the folksonomy, called *semantic folksonomy*, is defined as follows:

Definition 1 (Semantic Folksonomy) Let $Ł$ be the whole IEML semantic space. A *semantic folksonomy* is a tuple $SF = \langle \tilde{U}, \check{T}, \tilde{I}, Y, \check{N} \rangle$ where $\check{N}$ is a ternary relationship such as $\check{N} \subseteq \tilde{U} \times \check{T} \times Ł$.

From *semantic folksonomies*, we present a formal description of a semantic item model as follows:

Definition 2 (Semantic Item Model) Given an item $i \in \tilde{I}$, a formal description of a semantic item model for item i, $M(i)$, follows: $M(i) = \langle \check{T}(i), \check{N}(i) \rangle$, where $\check{T}(i) = \{(u, t) \in \tilde{U} \times \check{T} \mid (u, t, i) \in Y\}$ and $\check{N}(i) = \{(t, \upsilon) \in \check{T} \times Ł \mid (u, t, \upsilon) \in \check{N}\}$.

2.1 Recommendation based on the Semantic Model

In our social recommender system, we first look into the set of similar items that the target user has tagged and then compute how semantically similar they are to the target item, called a *semantic item-item similarity*. Based on the semantically similar items, we recommend relevant items to the target user through capturing how he/she annotated the similar items. We define semantically similar items as a group of items that tagged categories of IEML close to those of the target item. Note that a *USL* can have at most seven distinct layers. Therefore, semantic similarity between two items, i and j, can be computed by the weighted sum of layer similarities from layer 0 to layer 6. Formally, the semantic item similarity measure is defined as:

$$semISim(i, j) = \omega \cdot \sum_{l=0}^{6} \frac{(l+1)}{7} \times \frac{|USL_*^i(l) \cap USL_*^j(l)|}{|USL_*^i(l) \cup USL_*^j(l)|} \tag{1}$$

where ω is a normalizing factor such that the layer weights sum to unity. $USL_*^i(l)$ and $USL_*^j(l)$ refer to the union of *USL*s for item i and j at layer l, $0 \le l \le 6$, respectively. The layer similarity between two *USL* sets is defined as the weighted *Jaccard coefficient* of two *USL* sets. Here we give more layer weights at higher layer when computing the semantic item similarity. That is, the intersections of higher layers present more contribution than intersections of lower layers.

Once we have identified a group of semantically similar items, the final step is a prediction, this is, attempting to speculate upon how a certain user would prefer unseen items. In our study, the basic idea of discovering relevant items starts from assuming that a target user is likely to prefer items which are semantically similar to items that he/she has tagged before. Formally, the prediction value of the target user u for the target item i, denoted as $P(u, i)$, is obtained as follows:

$$P(u,i) = \sum_{j \in SSI_k(i)} \frac{|USL_*^i \cap USL_u^j|}{|USL_u^j|} \times semISim(i, j) \quad \textbf{(2)}$$

where $SSI_k(i)$ is a set of k *most similar items* of item i grouped by the semantic item similarity and USL_u^j is the union of *USLs* connected to tags that user u has annotated item j. $semISim(i, j)$ denotes the semantic similarity between item i and item j. Finally, a set of top-N ranked items that have obtained the higher scores are identified for user u, and then, those items are recommended to user u.

3 Experiment Result

In this section, we compare the performance of our method (*SM*) against that of the benchmark algorithms: a user-based collaborative filtering (*UCF*), an item-based collaborative filtering (*ICF*), and a most popular tags approach (*MPT*) [2]. The dataset used in this study is the *p*-core at level 5 from *BibSonomy* (http://bibsonomy.org) [2]. To evaluate the performance of the recommendations, we randomly divided the dataset into *a training set* (80%) and *a test set* (20%) for each user. We used a 5-fold cross validation scheme. Therefore, the result values reported in the experiment section are the averages over all five runs. We adopted *precision* and *recall* to measure the performance of the recommendations [1].

Prior to the experiment, we measured the performance of *UCF*, *ICF* and SM according to the variation of the neighborhood size. As a result, the best values of the neighborhood size were 20 for *UCF*, 40 for *ICF*, and 30 for SM, respectively.

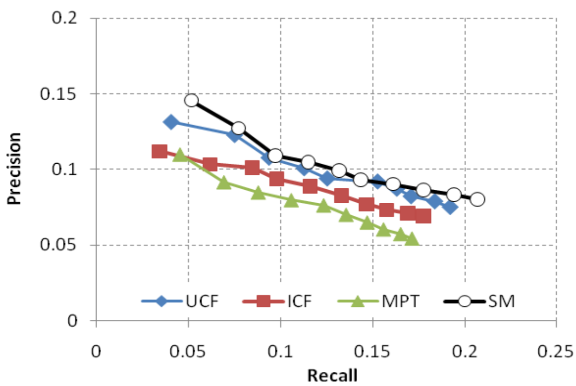

Fig. 1. *Recall* and *precision* as the value of the number of recommended items *N* increases

For evaluation the top *N* recommendation, we measured *precision* and *recall* obtained by *UCF*, *ICF*, *MPT*, and *SM* according to the variation of *N* value from 1 to

10. Fig. 1 depicts the precision-recall plot, showing how *precisions* and *recalls* of four methods changes as *N* value increases. Data points on the graph curves refer to the number of recommended items. Comparing the results achieved by *SM* and the benchmark methods, both *recall* and *precision* of the former was found to be superior to that of the benchmark methods in all cases. Only *UCF* achieves comparable results on some occasions. An important observation is that *SM* significantly outperforms the other methods when a relatively small number of items were recommended (e.g., top-1, top-2). For example, when *N* is 1 (starting points on the left of the curves), with respect to *recall*, *SM* obtains 1.12%, 1.73%, and 0.64% improvement compared to *UCF*, *ICF*, and *MPT*, respectively. With respect to *precision*, similar results are demonstrated. SM outperforms *UCF*, *ICF*, and MPT by 1.43%, 3.37%, and 3.59%, respectively. That is, *SM* provides more suitable items with a higher rank in the recommended item set, and thus can provide better quality of items for the target user than the other methods. We conclude from the comparison results that our semantic model can provide better performance of recommendations than other methods.

4 Conclusions and Future Work

In this paper, we have presented a semantic model and a method of applying the model to social recommender systems. As noted in our experimental results, our model can successfully enhance the performance of item recommendations. Moreover, we also observed that our approach can provide more suitable items for user interests, even when the number of recommended is small. For the future work, we intend to explore semantic relations of IEML and apply the relations to the semantic item similarity and item recommendations.

Acknowledgment. The work was mainly funded since 2009 by the Canada Research Chair in Collective Intelligence at University of Ottawa.

References

1. Herlocker, J. L., Konstan, J. A., Terveen, L. G., Riedl, J. T.: Evaluating Collaborative Filtering Recommender Systems. ACM Transactions on Information Systems 22(1): 5-53 (2004)
2. Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., Stumme, G.: Tag Recommendations in Social Bookmarking Systems. AI Communications 21(4): 231-247 (2008)
3. Kim, H.-N., Ji, A.-T., Ha, I., Jo, G.-S.: Collaborative Filtering based on Collaborative Tagging for Enhancing the Quality of Recommendation. Electronic Commerce Research and Applications, 9(1): 73-83 (2010)
4. Lévy, P.: From Social Computing to Reflexive Collective Intelligence: The IEML Research Program. Information Sciences 180(1): 71-94 (2010)