International Journal on Advanced Science Engineering Information Technology

Vol.7 (2017) No. 6
ISSN: 2088-5334

# Using Tags for Measuring the Semantic Similarity of Users to Enhance Collaborative Filtering Recommender Systems

Ayman S. Ghabayen[#], Shahrul Azman Noah *

[#]*Department of Computer Science, College of Science and Technology, KhanYounis, Palestine*
*E-mail: a.ghabayen@cst.ps*

[*]*Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia*
*E-mail: shahrul@ukm.edu.my*

***Abstract*— Recent years have seen a significant growth in social tagging systems, which allow users to use their own generated tags to organize, categorize, describe and search digital content on social media. The growing popularity of tagging systems is leading to an increasing need for automatic generation of recommended items for users. Much previous research focuses on incorporating recommender techniques in social tagging systems to support the suggestion of suitable tags for annotating related items. Collaborative filtering is one such technique. The most critical task in collaborative filtering is finding related users with similar preferences, i.e., "liked-minded" users. Despite the popularity of collaborative filtering, it still suffers from certain limitations in relation to "cold-start" users, for example, where often there are insufficient preferences to make recommendations. Moreover, there is the data-sparsity problem, where there is limited user feedback data to identify similarities in users' interests because there is no intersection between users' transactional data a situation which also results in degraded recommendation quality. For this reason, in this paper, we present a new collaborative filtering approach based on users' semantic tags, which calculates the similarity between users by discovering the semantic spaces in their posted tags. We believe that this approach better reflects the semantic similarity between users according to their tagging perspectives and consequently improves recommendations through the identification of semantically related items for each user. Our experiment on a real-life dataset shows that the proposed approach outperforms the traditional user-based collaborative filtering approach in terms of improving the quality of recommendations.**

***Keywords*— recommendation system; collaborative filtering; social tagging system**

## I. INTRODUCTION

The World Wide Web (WWW) has undergone exponential growth over the past two decades; the first generation enabled Internet users to have direct access to a large variety of available knowledge. The second generation of the WWW typically denoted as "Web2.0" and also referred to as "participatory Web2.0," has led to a significant change in the way in which people interact with and through the Web. Web2.0 can be characterized as a paradigm that facilitates communication, interoperability and information sharing and collaboration on the Web [1].

Web2.0 allows users to easily annotate any item (websites, articles, media, etc.) that somebody else has published. These annotations (tags) take many forms such as rating, editing, classification, and organizing. These. These annotations enable users to easily retrieve, search or filter these items in the future. Moreover, the social phenomenon of collaborative tagging (also known as "folksonomies" or "social tags") is a big shift from earlier local and solitary to global and collaborative Internet activity. This shift has enabled users to be information producers, rather than just information browsers. However, rich information is increasing exponentially in social web systems, a substantial amount of new information being produced every day. Since this phenomenon is already exceeding human processing capabilities, it is becoming difficult for users to find the needed information quickly because they face the problem of information overload [2], [3]. Consequently, recommender systems have emerged in response to the information overload problem, providing users with recommendations for items that are relevant and likely fit their needs [4]. Furthermore, collaborative tagging systems such as Delicious, YouTube, Flicker, and Twitter allow the creator or visitors to assign freely chosen keywords or tags to such content [5].

Many recent studies have focused on using recommender systems with social tagging [2], [6]-[8] to mitigate limitations such as "cold start" and sparsity [3], which are present in the traditional systems. However, without considering the semantics of user tags in the recommendation process, the recommender system cannot distinguish and interoperate the user's interests for the same tags. Furthermore, almost all of the studies integrate recommender systems in social tagging to offer tag recommendations in order to assist users annotate, search and organize their own content, as well as search other content shared by other users.

We believe that semantic tags can tackle the limitations inherent in traditional collaborative filtering and improve the quality of collaborative filtering by capturing users' semantic preferences based on the user tags. In traditional user-based collaborative filtering, two users are similar if they co-rate a particular item with similar score values. Consider two users, u1 and u2, both of whom rate the movie "Avatar" with a similarly high score. In traditional user-based approaches, u1 and u2 are considered similar and "like-minded" users. However, traditional methods discard the semantic perspective of the respective users, where u1 awards a high score because he likes "science fiction" movies, whereas u2 likes "Avatar" because she likes "adventure movies."

To elaborate the problem further, in state-of-the-art social tagging and collaborative filtering, two or more users are considered similar if they both annotate a particular item with similar tags. For example, let user u1 post tags (Java, tour) on an item, u2 post tags (Java, XML) on an item, and u3 post a tag (RDF) on another item. In traditional methods, u1 and u2 are similar because both tagged (java); however, there is no similarity between u2 and u3. Unfortunately, the similarity is incorrect in this scenario because traditional methods do not distinguish between "Java", the island, for u1 and "java", the programming language, for u2. To solve this problem, our approach determines the semantic similarity between users so that u2 and u3 are identified as being more semantically similar according to their semantic tags, wherein the tags "java" and "XML" of u2 are more similar to the tag "RDF" of u3 than to the tags "java" and "tour" of u1.

This paper presents an approach to measure the similarity between two users on the basis of the semantics of the annotation or tags attached by both users. Therefore, instead of considering the co-occurring features of tags as exhibited in other research, we extend those tags into their respective semantics by exploiting available open-semantic lexical resources. The remainder of this paper is organized as follows: Section II presents the basic principle of baseline recommender system, related work, and the proposed approach, dataset and evaluation matrices used.; Section III presents the results and discussion; and finally, Section IV offers a conclusion and our recommendations for future work.

## II. Material and Methods

In this section, we describe the basic principle that underlies the baseline recommender system used to generate a recommendation, related work, and the proposed approach.

### *A. Recommender Systems*

In a common formulation, the recommender system task is reduced to the problem of discovering related items that have not been seen by the user [3]. Collaborative filtering (CF hereafter) is considered to be the most promising recommendation approach that automates the process of the "word-of-mouth" paradigm in estimating of unseen items [9]. CF compares users on the basis of the similarity of their preferences and those of other users [10], [11]. The two main recommendation approaches in CF are the item-based CF [12], [13] and the user-based CF [14]. Usually, the recommendation process in both approaches depends on discovering a similar pattern for the target user (the term used in the user-based approach) and other users having similar preferences to form a "neighborhood," the preferences from which usually are called the most similar users (or similar items in the item-based approach). Many computing methods have been used to measure the similarity between users in CF, such as the Pearson correlation coefficient and cosine similarity [4]. The most critical task in a CF recommender system is the formation of a similar neighborhood because of differences within result in different recommendations, thereby influencing the accuracy of the recommendation process. The similarity between two users, u and v, is calculated as the cosine angle between the corresponding feature vectors, as follows[15]:

$$\text{similarity}(u,v) = \cos(\theta) = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \cdot \|\vec{v}\|} \qquad (1)$$

When these neighbors have been found, the next step is estimating the predicted value of items as yet unseen and unrated by the target user. The higher number of similar users found in the recommendation neighborhood for the target user, the more influences this user has in this prediction process. The last step is the recommendation of the top M items with the greatest predicted values to the target user [16].

As a result of the growing popularity of social media sites in recent years, many researchers have investigated the recommender system domain under the social-tagging area of research, where tags have been considered as an additional information resource for designing effective recommendation systems. Social tagging systems also called folksonomies allow users to assign content with a freely chosen keyword called tag [5], which can reflect the users' cognitive preferences for the content. Hence, the tag co-occurrence properties might express similarity between users or items to build a user community and item clusters, which can be employed to estimate the likely items for targeted individuals. Therefore, tags in social tagging provide a promising way to tackle some of the limitations in recommender systems, such as the cold-start and sparsity problems [17]. The phenomenon of social tagging has resulted in two areas of research in recommender systems: i) tag recommendations and suggestions and ii) resource filtering and recommendations.

When making tag recommendations and suggestions, the main idea is to assist users by recommending appropriate tags for annotating given items. The proposed approach presented in this paper, however, falls into the latter category, i.e., resource filtering and recommendation. Therefore, it is beyond the scope of this paper to discuss various approaches to tag recommendation. Further information is available in [18].

## B. *Related Work*

The area of resource filtering and recommendations has attracted many scientists who have proposed innovative approaches for enhancing existing recommender systems. Tso-Sutter et al. [18] integrate tags into CF by reducing the three-dimensional <user, item, tag> relationship to three two-dimensional relationships, as <user, tag>, <item, tag> and <user, item>. The idea behind their approach is to consider tags as items in a two-dimensional relationship; for example, <user, tag> should be considered as a single item in the user-item rating matrix. To integrate tags with CF, the researchers then apply a fusion method to re-associate these relationships. However, the process of reducing three-dimensions into two generally leads to the discarding of potentially useful information for the recommender system.

De Gemmis et al. [19] propose a strategy that enables a content-based recommender to infer user interests. Machine-learning techniques are applied to both official content descriptions of items (static content) and the tagging data (dynamic content) to build user profiles and learn user interests. Static and dynamic content are preventively analyzed in order to capture the semantic preferences of the users behind the keywords to increase the prediction accuracy of the recommender system. De Gemmis et al. utilize the content-based method on their suggested approach since they assume that this method provides adequate recommendations [16]. Nevertheless, this method is merely valuable when the recommended items include an adequate amount of content, such as books, articles, and bookmarks that can be easily extracted.

Bao et al. [20] propose two algorithms to incorporate social tagging and web searching into CF, called SocialSimRank (SSR), used to calculate the similarity between tags and web queries), and SocialPageRank (SPR), used to tabulate the popularity of a web page by considering the number of annotations on the page. Bao et al. attempt to improve web searching by incorporating social tags into user query expansion. Liang et al. [21] integrate tags inside CF to build a tag-based similarity to boost the standard CF simply by clustering users according to their tagging behavior as different to their similar-rating. However, how many clusters have to be defined due to the fact of a huge tagging data clustering can be quite a very high-priced computation.

Sen et al. [22] suggest a tag-based recommendation algorithm known as “tagommenders” the actual underlying concept being this algorithm may be used to predict users preferences with regard to items based on their deduced tagging data. Au et al., [23] assuming that it is still possible that users can influence one another in the process of item adoption through various implicit mechanisms, capture the influence preferences among the users in a social system, considering the preferences to be related to the tagging behaviour between users for certain items. Ghabayen et al .[24] suggests calculating the similarity between target users by expanding the users’ tags on items. As this kind of similarity in between users could be enhanced just as much tags co-occurrence occurs between users.

Our approach differs from the aforementioned studies in that we aim to explore the tagging of the semantic space of users. In other words, we consider semantic tags to discover like-minded users so that semantically relevant items can be recommended to a particular user. We expect this to become more useful not just in improving the recommendations quality, but additionally in recognizing better user perception associated with relevant items.

## C. *Proposed Approach- Exploiting Semantic Similarity of Tags for Collaborative Filtering*

Fig. 1 demonstrates the overall process in our proposed approach utilizing the three main CF steps: i) semantic similarity of tags ii) generating a neighbourhood and iii) recommending relevant items.

The basic idea behind our study assumes that active users are interested in items that have been tagged by like-minded users and that these tags are similar to the tags used by the same active users. The first step in the proposed approach is to compute the semantic similarity for tags. The second step in is to look for a set of similar users who have tagged a target item. Then we compute the semantic similarity between that similar user set and the active user. On the basis of this similarity, the semantic ranking of the item is computed to decide whether to recommend the target item. In the last step, the top items are recommended to the target user on the basis of the semantic similarity with like-minded users. This process as is illustrated in Fig. 1.

### 1) *Semantic Similarity for Tags*

In our approach, the semantic similarity between tags is obtained by exploiting the well-known WordNet lexical database for English. WordNet is really a large conceptual design database associated with nouns, verbs, adjectives as well as adverbs which are grouped into sets associated with cognitive synonyms (synsets) [25]. In WordNet, there is a conceptual-semantic interlinkage and lexical relationship between synsets. The terms which hold the same meaning are referred to as synonyms, which belong to the same concept and are placed in the same synset. These hierarchical concepts can quantify the extent to which concept A is similar to concept B. For example, these concepts and relationships might indicate that an automobile is more similar to a boat than to a tree because “boat” and “automobile” share “vehicle” as a common ancestor in the WordNet structure [26].

Our proposed approach enables any user to tag any item in the collaborative tagging environment, as well as duplicate a tag for the same item as a different user. The proposed approach is based on the triple <user, item, tag> representation which is widely adopted in the collaborative tagging community. A folksonomy is a set of triples. Each triple represents a user's annotation of an item with a tag. More technically speaking, if there is a list of users U = {u1, u2, u3,…, um}, a list of items I = {m1, m2, m3,…, mk}, and a list of tags T = {t1, t2, t3,…, tn}, the folksonomy F = <U,I,T,Y>, where Y is the user tag assigned for an item [27]. This designation differs from traditional CF where each user assigns a tag to an item, rather with triples. The existence of a scope of real numbers leads us to consider the user feature as a vector of tags posted by the user. For example, in Fig. 1, user u1 posts t1, t2 for item1 which can be represented as (u1, m1, (t1, t2)).

To provide a semantic grounding for our folksonomies, we use WordNet as the external semantic space for measuring the semantic similarity between tags. Calculating the semantic similarity in WordNet can be done by measuring the distance between nodes related to the associated concepts. When the links between these nodes are considered in terms of distance, that distance indicates how similar the concepts are. We measure the similarity between tags by using Lin's semantic similarity [28], which uses information content for calculation. Lin's semantic measures relate the information content (IC) of the most informative common ancestor (MICA) to the IC of the associated concepts thus:

$$\mathrm{Sim}_{\mathrm{Lin}}\left(c_1, c_2\right) = \frac{2 \times \mathrm{IC}\left(C_{\mathrm{MICA}}\right)}{\mathrm{IC}\left(c_1\right) + \mathrm{IC}\left(c_2\right)} \tag{2}$$

Lin's similarity ranges from 0 for tags without similarity to 1 for tags with maximum similarity. Budanitsky et al. [29] mention that similarity can be viewed as a particular case regarding relatedness due to the fact both are usually semantic notations. When measuring the semantic similarity of social tags, we need to map the tags to an existing lexicon or thesaurus such as WordNet [25]. However, tags are by nature free keywords which can include many community-specific terms that do not exist in any lexicon. Therefore, we propose the use of co-occurrence distribution to identify the semantic similarity for such tags.

Christian et al. [30] propose a method to calculate the co-occurrence association between tags in social tagging systems as follows: Let n(m,t) be the number of occurrences of tag t on item m; $n(t) = \sum_m n(m,t)$ the number of tag occurrences for all items in $I$; and $N(t) = \sum_t n(m,t)$ the number of tag occurrences for item m, where $m$ is an instance of $I$ and $m \in I$. In [30], the actual similarity in between two tags tx and ty is recognized as the actual weighted average from the tag distributions for that item, which signifies the co-occurrence distribution between tags for this item. The co-occurrence distribution of a tag for all items in a social tagging system is calculated by (3);

$$p_{t_y}\left(t_x\right) = \sum_{m \in I} q\left(t_x | m\right) Q(m \mid t_y) \tag{3}$$

where,

$$Q\left(m | t_y\right) = \frac{n\left(m, t_y\right)}{n\left(t_y\right)} \text{ on I.}$$

$$q\left(t_x | m\right) = \frac{n\left(m, t\right)}{N\left(m\right)} \text{ on } m \in \text{I.}$$

*2) Generation of Semantic-Based Neighborhood*

As mentioned previously in section 2, among the critical tasks of the user-based CF recommender system is the generation of some like-minded or even nearest-neighbor users having preferences like the target users. Consider two users, $u$ and $v$, where u, v∈U. First, we obtain items $m_{u,v} \in I$, which are sharable in terms of tagging behaviour between the related users. For each item $m \in m_{u,v}$ we present each user with the tags posted by users $u$ and $v$. The tags of users $u$ and $v$ or item m are presented as (u, m, ($t_{u1}$,..., $t_{un}$)) and (v, m, ($t_{v1}$,..., $t_{vn}$)) respectively, where both $t_{un}$ and $t_{vn} \subseteq T$. For each tag $t_u$ and $t_v \in \left(t_{v1}, .., t_{vn}\right)$ we calculate the semantic-similarity-of-tag (STSim) value. $\mathrm{STSim}\left(t_u, t_v\right)$ can be calculated by using (2) if both $t_u$ and $t_v$ exist in the WordNet lexicon; otherwise, the value can be calculated by using (3) if one of the tags does not exist in WordNet. On the basis of the STSim value for the tags given by both user $u$ and user $v$ on $m_{u,v}$ we can determine the semantic similarity between the two users by (4):

$$\mathrm{SUsim}(u, v) = \sum_{m \ni m_{u,v}} \sum_{t_u, t_v} \mathrm{STSim}\left(t_u, t_v\right) \tag{4}$$

Where, $t_u$ denotes the tags posted by user u on item m, and $t_v$ denotes user v's tags on item m, where m ∈ $m_{u,v}$ the higher the SUsim value between the two users $m_{u,v}$ the greater their similarity.

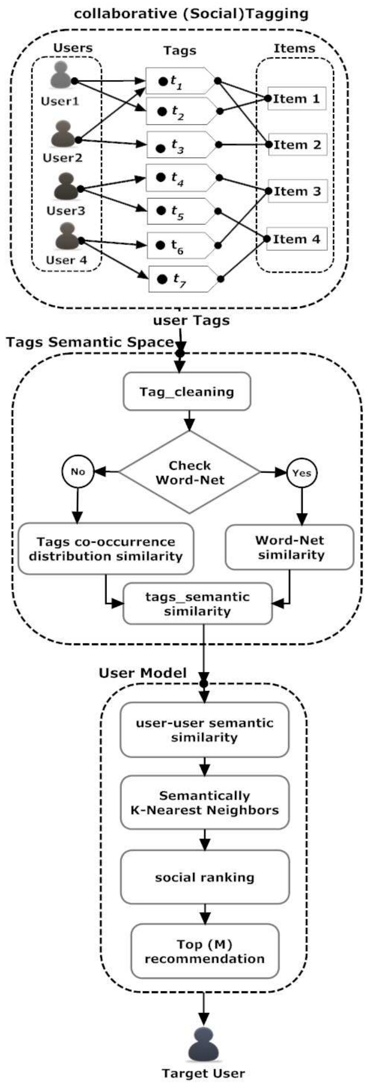


Fig. 1 System overview of collaborative filtering based on semantic tags

*3) Item Recommendation*

Once the set of N semantically like-minded users are identified, the final step includes the actual prediction for every item and also the generation of the top M recommended items. In 0ur proposed approach, the basic concept of estimating appropriate unseen items for that active users starts in the assumption which users prefer items which have been tagged through their like-minded users. We identify this assumption as a semantic sociable rank from the set regarding SSU, as follows:

$$SSR(u, r) = \sum_{v \ni SSU_N(u)} SUsim(u,v) \times social\ rank(v) \quad (6)$$

Where, $r \in I - m_u$ and $r$ denotes the items that have not been seen by user u; $m_u$ the items tagged by user u; and $I$ the set of all items. The social rank is equivalent to 1 if the item has been tagged through semantically similar users; else, it acquires a value equal to 0. Finally, a set of top M ranked items which obtained greater SSR values is recommended to user u.

*4) Dataset and Evaluation Matrices*

The dataset used in our experiments is the hetrec2011-movielens-2k dataset dated May 2011 which has been made available to the public by Cantador et al. [31]. It is based on the original MovieLens10M dataset, published by the Group Lens1 research group. This dataset has been used in previous studies such as [32], [33]. One of the main difficulties when coping with tagging data is the quality of the tags mainly because tags are words or a combination of words that are that are freely assigned by users. In order to ensure the quality of our experiments and findings, it is necessary to remove meaningless data by filtering the dataset. Since our proposed approaches depend on co-occurrence distribution between tags, we apply a dataset filtering implemented by previous research in this area [22], [30], [34]. We did not consider meaningless tags, i.e., tags that had not been assigned to at least two items and items that had not been annotated by at least two tags because this would lead to a zero co-occurrence score with other tags. We eliminate tags that had not been given to at least five items because such assignment would lead to a low co-occurrence score with other tags. According to the high sparsity of the tagging dataset, we considered items that had at least 15 tags [35]. The final abridged dataset used in our study consisted of the following: 2,013 users, 500 items, 14,800 tagging records and 1,400 tags.

The main problem when trying to map tags in the MovieLens dataset to WordNet is that not all the tags are recognized by the lexicon. Specifically, 51% of the tags in the dataset were not in WordNet. Therefore, we tried to increase this percentage by stemming the original tags in the dataset. Then we used Edit Distance-based Word Similarity (Levenshtein distance) [36], one of the well-known edit distance functions. The Levenshtein distance is defined as the number of deletions, insertions, or substitutions of characters required to transform string s1 into another string, s2; the shorter the distance between the strings, the greater the similarity. MovieLens tags can be mapped to WordNet lemmas if the edit distance ratio between the test tag and the WordNet lemma is greater than 85%. Undertaking this cleaning process resulted in 58.8 % of the tags in the dataset being mapped to the WordNet lexicon.

To evaluate the performance of our proposed approach, As a consequence Herlocker et al. [37] proposed to use Decision-support accuracy which includes classical information retrieval metrics such as: precision and recall. These matrices judge how well a recommender system can make a prediction of high relevance items. These metrics are suitable for evaluation Top-M recommendation list. Precision is a metric that represents the probability that an item recommended as relevant is truly relevant. It is defined as the ratio of items correctly predicted as relevant among all the items selected. The recall is a metric that represents the probability that a relevant item will be recommended as relevant. It is defined as the ratio of items correctly predicted as relevant among all the items known to be relevant. the Precision and Recall for user u as follows:

$$Precision(u) = \frac{|Test(u) \cap TopM(u)|}{|TopM(u)|} \quad (7)$$

$$Recall(u) = \frac{|Test(u) \cap TopM(u)|}{|Test(u)|} \quad (8)$$

where Test (u) is the relevant items for user u in the test set and TopM (u) is the top M recommended items for user u. To make our result realistic, we considered that some users would have large tagging records, whereas, others would make only a few tags for items. Next, we calculated the precision and recall for each user in the user space. Then the Average Precision@M (AP) and the Average Recall@M (AR) was calculated by the following equations:

$$AP(U) = \frac{1}{U} \sum_{u \ni U}^{U} Precision(u) \quad (9)$$

$$AR(U) = \frac{1}{U} \sum_{u \ni U}^{U} recall(u) \quad (10)$$

However, according to the number of recommended items, the values of precision and recall conflict with each other. Generally, an increment in the number of items recommended tends to increase recall but decreases precision [37]. Therefore, we also considered the F1 measure, which combines both recall and precision with equal weight in a single value [37]. The F1 measure is denoted by the following equation:

$$F1 = \frac{2 \times recall \times precision}{recall + precision} \quad (11)$$

In order to compare the performance of our proposed approach, we compared our approach with a popular tagging approach [38], [39] based on classical cosine similarity and presented as (cosine_CF), which depends on users' tagging histories. The results of this comparison are discussed in the next section.

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

## III. RESULTS AND DISCUSSION

This section presents the experimental results with respect to the quality of item recommendations. We compared the performance of the top M recommendations for our proposed approach with a popular tagging approach based on the cosine similarity measure [38], [39]. The experiment on top M recommended items was done with a variant number of recommended items by considering M from 5 to 100 with an increment of 5. Furthermore, we also considered the number of K similar users.

The results are shown in Fig. 2 and Fig. 3 yield an interesting finding. We can observe that precision gradually decreases while recall increases with the increment in the top M recommended items. One possible reason with this outcome is actually with the increment of M recommended items, more false positives are likely to be returned in the recommendation, thus resulting in low precision; whereas, more true positives are likely to be returned for the increment of M recommended items that obtain higher recall. This pattern of findings is popular in information retrieval research. However, our proposed approach (denoted semantic_CF in Fig. 2 and Fig. 3) outperforms traditional CF (denoted cosine_CF) in terms of precision and recall, as shown in Fig. 2 and Fig. 3.

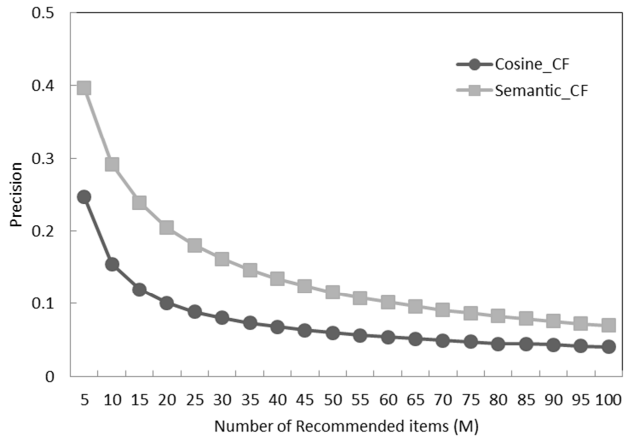


Fig. 2 Precision with increment of top M recommended items

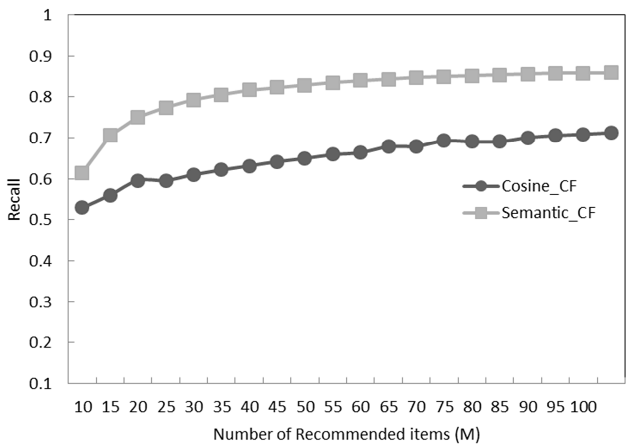


Fig. 3 Recall with increment of top M recommended items

Fig. 4 and Fig. 5 also indicate that the proposed approach outperforms the cosine-based approach in terms of F1 with regard to a number of recommended items and number of neighbors. The M = 10 shows the highest F1 measure, indicating that higher values of M will result in more 'junk' recommendations.

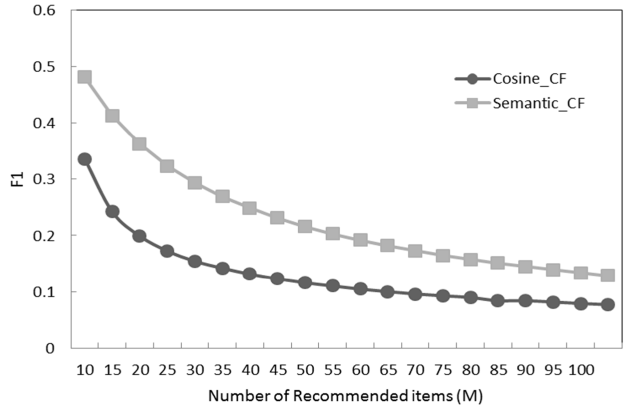


Fig. 4 F1 measure values with increment of Top M recommended items

We also examined the F1 measure with various numbers of top K similar users. The neighborhood selection of our proposed approach was found to differ from that of the classical cosine-based similarity. We presumed that the difference occurred because the proposed approach is based on users' semantic perspectives on tags; whereas, the cosine-based approach depends solely on the co-occurrence tagging between users.

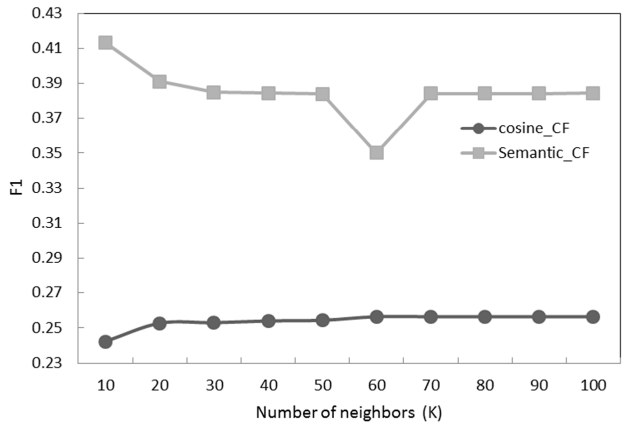


Fig. 5 F1 measures with variant size of top K similar users

Our approach outperforms the cosine-based in all variants of top K users. The superiority of our approach, however, decreases when the number of top K users ranges between 50 and 70, due to the effect of lower precision; whereas, the cosine-based is hardly affected by the increment in the neighbor size.

## IV. CONCLUSION

This paper has presented an approach for deriving semantic similarity between users (i.e., neighborhood) by exploiting user tags. The main idea comes from the belief that, 'similar' tags allocated by different users could possibly indicate relatedness and potential input for recommender systems. However, the majority of tagging activities are subject to little-to-no control with regard to terms and vocabulary used. Therefore, different tags can be semantically equivalent. In order to defeat such a situation, we utilized WordNet to determine the semantic relatedness between tags. In the case of words not existing in the WordNet database, the co-occurrence distribution measure was used.

Evaluation of the MovieLens dataset has shown interesting and promising results. The proposed approach outperforms the conventional tag-driven user-based CF on the basis of the cosine-based similarity in terms of precision, recall and harmonic-means (F1) measure. Hence, it demonstrates that representing simple semantic information is capable of enhancing the performance of recommendation systems. However, there is no doubt that the complexity and extra processing required to implement the semantic analysis might be a disadvantage of this approach.

Our future works include evaluating the approach on a different or larger dataset for further comparison with other state-of-the-art approaches. With the emergence of Semantic Web and particularly Linked Open Data (LOD) [40], expansion of tags to such open data is another potential work in this area. Furthermore, explore other types of users generated data that may exist in LOD such as rating, blogs, reviews or demographic data could be integrated with tags to improve the recommendation quality

## REFERENCES


[1] P. Sharma, "Core characteristics of web 2.0 services," *Online at* http://www.techpluto.com/web-20-services, 2008.

[2] S. Siersdorfer and S. Sizov, "Social recommender systems for web 2.0 folksonomies," presented at the Proceedings of the 20th ACM conference on Hypertext and hypermedia, Torino, Italy, 2009.

[3] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," *IEEE Trans. on Knowl. and Data Eng.,* vol. 17, pp. 734-749, 2005.

[4] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Advances in Artificial Intelligence,* vol. 2009, p. 4, 2009.

[5] S. A. Golder and B. A. Huberman, "Usage patterns of collaborative tagging systems," *Journal of Information Science* vol. 32, pp. 198-208, 2006.

[6] Z.-K. Zhang, T. Zhou, and Y.-C. Zhang, "Personalized recommendation via integrated diffusion on user–item–tag tripartite graphs," *Physica A: Statistical Mechanics and its Applications,* vol. 389, pp. 179-186, 2010.

[7] H. Han, Y. Cai, Y. Shao, and Q. Li, "Improving Recommendation Based on Features' Co-occurrence Effects in Collaborative Tagging Systems,Web Technologies and Applications." vol. 7235, Q. Sheng, G. Wang, C. Jensen, and G. Xu, Eds., ed: Springer Berlin / Heidelberg, 2012, pp. 652-659.

[8] D. H. Lee and P. Brusilovsky, "Improving recommendations using WatchingNetworks in a social tagging system," presented at the Proceedings of the 2011 iConference, Seattle, Washington, 2011.

[9] J. B. Schafer, J. A. Konstan, and J. Riedl, "E-commerce recommendation applications," *Data mining and knowledge discovery,* vol. 5, pp. 115-153, 2001.

[10] V. Krishnan, P. K. Narayanashetty, M. Nathan, R. T. Davies, and J. A. Konstan, "Who predicts better?: Results from an online study comparing humans and an online recommender system," 2008, pp. 211-218.

[11] X. N. Lam, T. Vu, T. D. Le, and A. D. Duong, "Addressing cold-start problem in recommendation systems," presented at the Proceedings of the 2nd international conference on Ubiquitous information management and communication, Suwon, Korea, 2008.

[12] S. Mahapatra, A. Tareen, and Y. Yang, "A Cold Start Recommendation System Using Item Correlation and User Similarity," *ACM Transactions on Information Systems (TOIS),* 2011.

[13] G. Linden, B. Smith, and J. York, "Amazon. com recommendations: Item-to-item collaborative filtering," *Internet Computing, IEEE,* vol. 7, pp. 76-80, 2003.

[14] J. S. B. D. H. CarlKadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," *Microsoft Research Microsoft Corporation One Microsoft Way Redmond, WA,* vol. 98052, 1998.

[15] A. S. Ghabayen and S. A. Noah, "Enhancing the Accuracy of Social Collaborative Filtering based on the Relevance Weight of Tagging," *International Journal of Information Processing and Management,* vol. 4, pp. 127-136, 2013 2013.

[16] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Journal ,Advances in Artificial Intelligence,* vol. 2009, pp. 2-2, 2009.

[17] S. Hamouda and N. Wanas, "PUT-Tag: personalized user-centric tag recommendation for social bookmarking systems," *Social Network Analysis and Mining,* vol. 1, pp. 377-385, 2011/11/01 2011.

[18] K. H. L. Tso-Sutter, L. B. Marinho, and L. Schmidt-Thieme, "Tag-aware recommender systems by fusion of collaborative filtering algorithms," presented at the Proceedings of the 2008 ACM symposium on Applied computing, Fortaleza, Ceara, Brazil, 2008.

[19] M. d. Gemmis, P. Lops, G. Semeraro, and P. Basile, "Integrating tags in a semantic content-based recommender," presented at the Proceedings of the 2008 ACM conference on Recommender systems, Lausanne, Switzerland, 2008.

[20] S. Bao, G. Xue, X. Wu, Y. Yu, B. Fei, and Z. Su, "Optimizing web search using social annotations," presented at the Proceedings of the 16th international conference on World Wide Web, Banff, Alberta, Canada, 2007.

[21] H. Liang, Y. Xu, Y. Li, and R. Nayak, "Tag Based Collaborative Filtering for Recommender Systems," in *Rough Sets and Knowledge Technology*. vol. 5589, P. Wen, Y. Li, L. Polkowski, Y. Yao, S. Tsumoto, and G. Wang, Eds., ed: Springer Berlin Heidelberg, 2009, pp. 666-673.

[22] S. Sen, J. Vig, and J. Riedl, "Tagommenders: connecting users to items through tags," presented at the Proceedings of the 18th international conference on World wide web, Madrid, Spain, 2009.

[23] C. Au Yeung and T. Iwata, "Capturing implicit user influence in online social sharing," in *Proceedings of the 21st ACM conference on Hypertext and hypermedia*, Toronto, Ontario, Canada, 2010, pp. 245-254.

[24] A. Ghabayen and S. A. M. Noah, "Exploiting social tags to overcome cold start recommendation problem," *Journal of Computer Science,* vol. 10, pp. 1166-1173, 2014.

[25] C. Fellbaum. (1998) WordNet: an electronic lexical database. *MIT Press*.

[26] T. Pedersen, S. Patwardhan, and J. Michelizzi, "WordNet::Similarity: measuring the relatedness of concepts," presented at the Demonstration Papers at HLT-NAACL 2004, Boston, Massachusetts, 2004.

[27] J. Hu, B. Wang, Y. Liu, and D. Y. Li, "Personalized Tag Recommendation Using Social Influence," *Journal of Computer Science and Technology,* vol. 27, pp. 527-540, 2012.

[28] D. Lin, "An Information-Theoretic Definition of Similarity," presented at the Proceedings of the Fifteenth International Conference on Machine Learning, 1998.

[29] A. Budanitsky and G. Hirst, "Evaluating WordNet-based Measures of Lexical Semantic Relatedness," *Comput. Linguist.,* vol. 32, pp. 13-47, 2006.

[30] C. Wartena, R. Brussee, and M. Wibbels, "Using Tag Co-occurrence for Recommendation," presented at the Proceedings of the 2009 Ninth International Conference on Intelligent Systems Design and Applications, 2009.

[31] n. Cantador, I. Konstas, and J. M. Jose, "Categorising social tags to improve folksonomy-based recommendations," *Web Semantics,* vol. 9, pp. 1-15, 2011.

[32] C. Jones, J. Ghosh, and A. Sharma, "Learning multiple models for exploiting predictive heterogeneity in recommender systems," presented at the Proceedings of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems, Chicago, Illinois, 2011.

[33] A. Said, E. W. D. Luca, B. Kille, B. Jain, I. Micus, and S. Albayrak, "KMulE: a framework for user-based comparison of recommender algorithms," presented at the Proceedings of the 2012 ACM international conference on Intelligent User Interfaces, Lisbon, Portugal, 2012.

[34] Z.-K. Zhang, T. Zhou, and Y.-C. Zhang, "Tag-Aware Recommender Systems: A State-of-the-Art Survey," *Journal of Computer Science and Technology,* vol. 26, pp. 767-777, 2011/09/01 2011.

[35] H.-N. Kim, A. Alkhaldi, A. El Saddik, and G.-S. Jo, "Collaborative user modeling with user-generated tags for social recommender systems," *Expert Systems with Applications*, vol. 38, pp. 8488-8496, 2011.
[36] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals," *Soviet Physics Doklady*, vol. 10, pp. 707-710, 1966 1966.
[37] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems (TOIS)*, vol. 22, pp. 5-53, 2004.
[38] F. Durao and P. Dolog, "Extending a hybrid tag-based recommender system with personalization," presented at the Proceedings of the 2010 ACM Symposium on Applied Computing, Sierre, Switzerland, 2010.
[39] R. Jäschke, L. Marinho, A. Hotho, L. Schmidt-Thieme, and G. Stumme, "Tag recommendations in social bookmarking systems," *AI Communications* vol. 21, pp. 231-247, 2008.
[40] T. Heath and C. Bizer, "Linked Data: Evolving the Web into a Global Data Space," *Synthesis Lectures on the Semantic Web: Theory and Technology*, vol. 1, pp. 1-136, 2011/02/09 2011.