Leveraging Semantic Similarity for Folksonomy-Based Recommendation

To recommend interesting resources such as webpages or pictures that are available through social tagging sites, recommender systems must be able to assess such resources' similarity to user profiles. Here, the authors analyze the role semantic similarity plays in calculating the resemblance between user profiles and published resources in folksonomies. Experiments carried out using data from two social sites show that associating semantics with tags results in more accurate similarities among elements in tagging systems and, consequently, enhances recommendations.

Daniela Godoy
Argentinian National Scientific and Technical Research Council

Gustavo Rodriguez and Franco Scavuzzo
Universidad Nacional del Centro de la Provincia de Buenos Aires

Social tagging refers to collaboratively annotating Web resources using textual labels, also known as *tags*. Del.icio.us, Flickr, and CiteULike are examples of social sites in which users share various resources, including webpages, pictures, videos, and bibliographic references. The result of this collaborative tagging process is a social classification scheme — known as a *folksonomy* — that relates users and resources through tag assignments.

The rapid growth of communities that use social sites, as well as the myriad shared resources available in folksonomies, make discovering relevant content a time-consuming and difficult task for users. Unsupervised tagging and the lack of a control vocabulary for annotating resources exacerbate this problem; social tags are naturally noisy and ambiguous, reducing their effectiveness for content indexing and searching.

In this context, recommender systems that support users in tagging, searching, and discovering novel resources are becoming not only valuable but also extremely necessary tools. In fact, multiple recommender systems have emerged for social tagging sites. Traditional approaches for developing recommender systems are basically built on content-based and collaborative filtering techniques, which both rely heavily on the notion of similarity. In the content-based approach, potentially interesting items are predicted according to their similarity to items the user liked in the past; in the collaborative approach, items are recommended if they're interesting to people the user shares interests with.

In social tagging, we can represent user interest profiles with the tags that

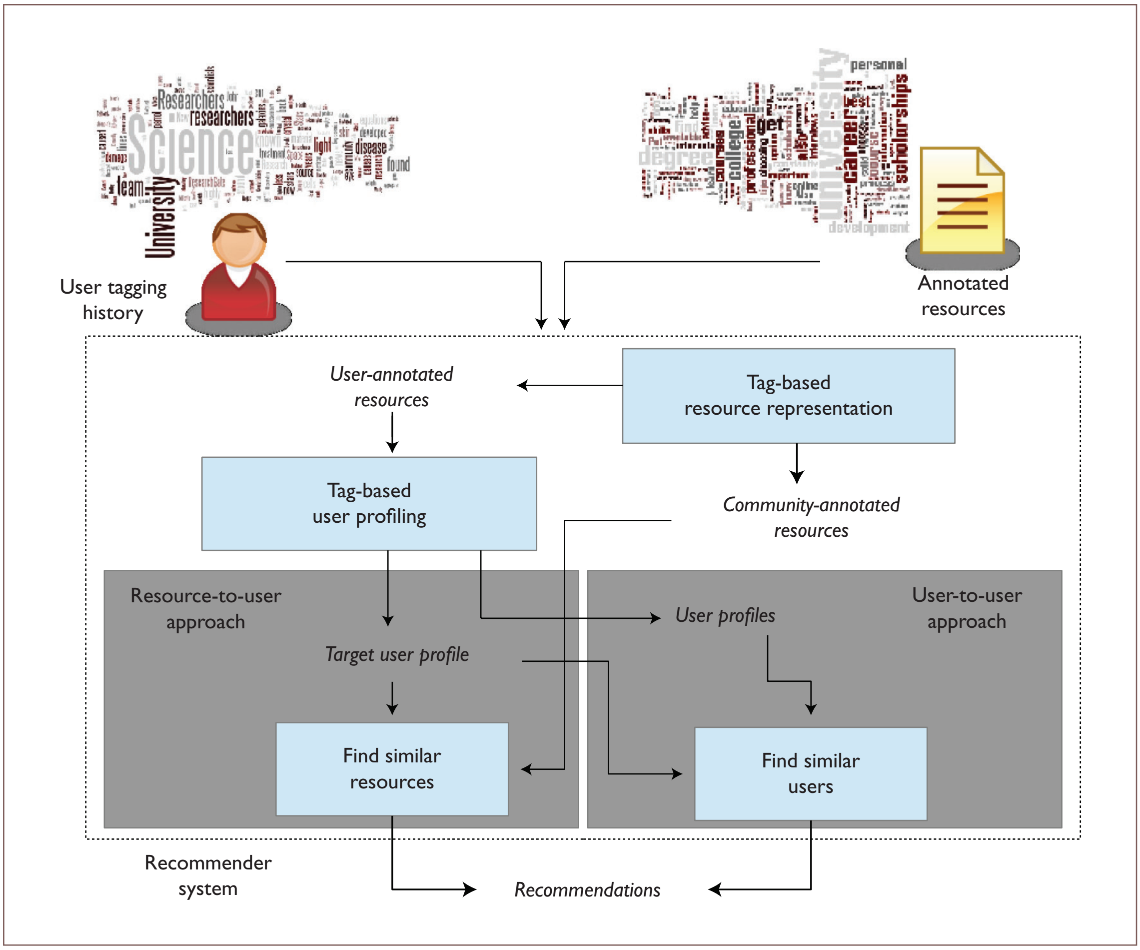

Figure 1. Resource recommendation approaches in social tagging systems. We can see how recommendations are generated starting from tag-based representations of resources and users.

the user tends to apply when annotating resources; we can represent resources with the specific tags with which they're annotated. Thus, similarity assessments of tag-based representations are affected by the syntactic variations in tags stemming from the use of different verb forms, plurals, acronyms, and synonyms, among other issues. By relating tags to semantic entities from lexical resources, we can enrich tag-based representations, reducing these problems' effects.

Quantifying how closely tags, resources, and users are related is essential to developing recommender systems for folksonomies. Enhancing similarity measures with semantic knowledge extracted from lexical databases such as WordNet lets such systems get better similarity assessments and, in turn, improve recommendations. Here, we analyze the role of semantic similarity in the context of traditional recommendation approaches for suggesting interesting resources in folksonomies. In particular, we focus on empirically evaluating the impact that associating concepts to tag-based representations of folksonomy elements (users and resources) has on the precision of the recommendations delivered to users.

Similarity Measures in Folksonomies

To apply classical, content-based, and collaborative filtering recommendation techniques, we must estimate the similarity among users and resources or users and other users starting from the folksonomy. Both users and resources are represented by the third dimension involved in tagging systems — that is, the social tags used for annotation purposes. Users who assign many of the same tags are similar to each other, whereas resources are similar to a user profile if they're annotated with tags the user tends to apply.

Figure 1 depicts a recommendation scheme in social tagging systems for suggesting potentially interesting resources. On one hand, we assume users are interested in the resources they annotate, so similarly tagged resources would also be interesting to them. On the other hand, users resemble each other if they tend to use the same tags, so one user can receive as a recommendation a resource that another user annotates. In this scheme, similarities among profiles and resources are derived from how similar their tag-based representations are.

Semantic-Based Tag Similarity

We can obtain the semantic similarity of two tags by associating them with semantic entities or concepts. For this purpose, we use WordNet (http://wordnet.princeton.edu), a large, lexical, English-language database that groups words into sets of synonyms called *synsets* and records various semantic relations between these sets.

For nouns and verbs, we can obtain a subsumption hierarchy based on the "is-a" relationships that connect *hyponyms* (more specific synsets) to *hypernyms* (more general synsets). Because a synset can have multiple hypernyms, the network becomes a directed acyclic graph with a top-level node that subsumes all the roots of the disconnected hierarchies to fully connect the graph. Approaches to measuring semantic relatedness between concepts base the measure of similarity on the properties of paths in this graph.

Philip Resnik's information-based approach[1] is based on the intuition that the shorter the path from one concept to another, the more similar they are, which we can determine in an is-a taxonomy by inspecting the relative position of the most specific concept that subsumes both concepts.

Let C be the set of concepts in an is-a taxonomy that permits multiple inheritance. The taxonomy is augmented with a function $p : C \rightarrow [0,1]$, such that for any $c \in C$, $p(C)$ is the probability of encountering an instance of concept c. Following the standard definition from information theory, the information content of c, denoted $IC(c)$, is then

$$IC(c) = -\log p(c), \tag{1}$$

where $p(c)$ is the probability of finding c in a given corpus:

$$p(c) = \frac{\sum_{w \in W(c)} frequency(w)}{N}, \tag{2}$$

where w is a word, $W(c)$ is the set of words describing the concept c, and N is the total number of words in the corpora. In this way, polysemous words contribute with the frequency of all their meanings. In other words, we obtain IC through statistical analysis of a corpus from which we can infer probabilities of concept occurrences. In the experiments we discuss here, we derived IC from SemCor, a manually sense-tagged subset of the Brown Corpus. We assumed that each word is used in its most often-occurring sense.

We can then define the semantic similarity of a pair of concepts c_1 and c_2 as

$$sim_{\text{Resnick}}(c_1, c_2) = \max_{c \in S(c_1, c_2)} IC(c), \tag{3}$$

where $S(c_1, c_2)$ is the set of concepts that subsume both c_1 and c_2.

Following this idea, the Jiang and Conrath similarity measure postulates that the semantic distance of the link connecting a child concept c to its parent concept c_p is proportional to the conditional probability $p(c|c_p)$ of encountering an instance of c given an instance of c_p.[2] From this postulate, it is possible to derive the following formula for the semantic similarity between concepts c_1 and c_2:

$$sim_{J\&C}(c_1, c_2) = \frac{1}{IC(c_1) + IC(c_2) + 2 * sim_{\text{Resnick}}(c_1, c_2)}. \tag{4}$$

This measure proves to be the most appropriate for measuring concept relatedness in WordNet.[3]

Similarity of Tag-Based Representations

We can translate user profiles and resources from social tagging systems into a bag-of-words representation that identifies each element by a feature vector with a numerical value or weight indicating its importance. Each element (user or resource) from a folksonomy $\mathbb{F}$ is then identified by a vector v in the t-dimensional space, where each vector component w_{ij} represents the weight of the tag t_i in the element v_j:

$$\vec{v_j} = \left(w_{1j}, w_{2j}, \ldots, w_{|T|j}\right), \tag{5}$$

where $t \in T$, $|T|$ represents the total number of tags in the folksonomy, and weights are assumed to be zero if the tag isn't present in the element description.

To represent a user u in the folksonomy, we create a vector $\overrightarrow{v_u}$, constituting the user profile, based on all tags this user employs to annotate resources. The amount of this vector's non-zero entries corresponds to the set of all tags from T_u in the user personomy $\mathbb{P}_u$, which is the part of the folksonomy corresponding to a single user. In tag-based representations, tag weights correspond to how frequently they appear in the resources the user annotates.

Likewise, a resource r in the folksonomy is represented by a vector $\overrightarrow{v_r}$ in which non-zero entries correspond to all tags assigned to the resource by members of the community. We assign tag weights according to the number of users in the folksonomy $\mathbb{F}$ that annotated the resource r with each tag — that is, how many times users assigned the tag to the resource.

We evaluated three similarity measures to determine the degree of resemblance between two users or a user and a resource and thus generate recommendations.

Overlap similarity. Overlap similarity measures the amount of tags two vector representations have in common, where a vector represents either a user or a resource. We define this measure as

$$sim_{overlap}\left(\overrightarrow{v_i},\overrightarrow{v_j}\right) = \frac{|V_i \cap V_j|}{\min(|V_i|,|V_j|)}, \tag{6}$$

where v_i and v_j are the sets of all tags of user i and j, respectively.

Cosine similarity. Commonly used in information retrieval, the cosine measure estimates the similarity between two vectors as the cosine of the angle they form in a vector space. Given two vectors $\overrightarrow{v_i}$ and $\overrightarrow{v_j}$, each representing either a user or a resource, we compute the cosine similarity as follows:

$$sim_{\cos}\left(\overrightarrow{v_i},\overrightarrow{v_j}\right) = \frac{\overrightarrow{v_i}\cdot\overrightarrow{v_j}}{\left\|\overrightarrow{v_i}\right\|\left\|\overrightarrow{v_j}\right\|} = \frac{\sum_{t\in T} w_{ti} * w_{tj}}{\sqrt{\sum_{t\in T} w_{ti}^2}\cdot\sqrt{\sum_{t\in T} w_{tj}^2}}, \tag{7}$$

where w_{ti} is the weight of the tag t in the vector representation $\overrightarrow{v_i}$ of a user or resource i.

Semantic similarity. To define a semantic similarity, we extended the cosine similarity to consider concepts associated with tags in WordNet. Normally, the cosine similarity will consider only products of those dimensions with an exact match in both vectors, because for non-matching tags, the corresponding dimensions in the second vector will be zero.

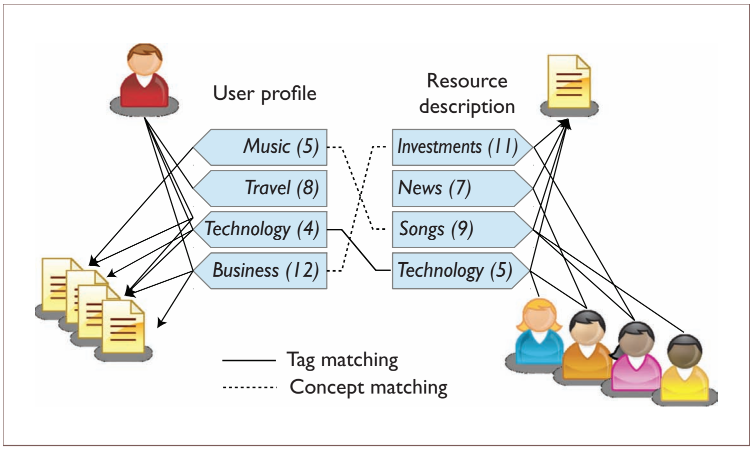

Figure 2. Example of semantic similarity calculation. We can see the inter-relationships among tags, users, and resources.

With semantically enriched tags, if a tag in the first vector matches one in the second vector, we multiply their weights. Otherwise, we multiply the tag weight by the weight of the most similar tag in the second vector found with the Jiang and Conrath similarity, provided it exceeds a certain threshold. If no tag is similar enough, we multiply the tag in the first vector by zero.

Figure 2 shows an example of how we calculate semantic similarity. Consider a user, represented by all the tags he's assigned to his resources, and a resource, represented by all the tags users in the community have annotated it with. In the example, the tag *technology* is present in both vector representations. For the tags *music* and *business*, we can find some similar tags, *songs* and *investments*, respectively. In contrast, the *travel* and *news* tags have neither a direct match nor a semantically similar tag in the other vector.

More formally, we can define the semantic similarity measure as

$$sim_{sem}\left(\overrightarrow{v_i},\overrightarrow{v_j}\right) = \sum_{t\in T} w_{t_{ij}}, \tag{8}$$

where we calculate $w_{t_{ij}}$ according to the following:

$$w_{t_{ij}} = \begin{cases} w_{ti} * w_{tj} & \text{if } t_i = t_j \\ w_{ti} * w_{tk} & \text{if } t_i \sim t_k, \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

and

$$t_k : \arg\max_{t\in T_j} sim_{J\&C}\left(t_i,t\right). \tag{10}$$

Related Work in Semantic Similarity within Social Tagging Systems

Several recent works have addressed the problem of how to infer relatedness between social tags, which has potential for a wide range of applications, such as tag clustering, ontology learning, query expansion, and search assistance, in addition to recommendation. One study analyzed tag-weighting models that take advantage of the three dimensions of folksonomies on which to base similarity. The study used classical (non-semantic) similarity measures such as the Dice Similarity Coefficient, cosine, and mutual information.[1] Another approach uses co-occurrence and its distributional version to define relatedness directly on the network structure of folksonomies instead of using a lexical resource.[2] In both these works, the authors used WordNet as a gold standard to evaluate the inferred relationships. In contrast, we exploit semantic relationships in the WordNet structure to assess tag similarity. More importantly, we empirically evaluate the impact that semantic enrichment has in calculating user-user and user-resource similarities in the context of classical recommendation approaches.

Closely related to our work is a study on content-based and collaborative methods using different similarity metrics.[3] For item-based filtering, using tags to calculate similarity alleviated sparsity and improved the recommendation results. User-based filtering didn't lead to the same results. Our work differs from this study in that we include a semantic similarity measure in the comparison with other measures in the context of resource-to-user (similar to the traditional content-based approach), user-to-user (similar to the collaborative filtering approach), and hybrid recommendation methods.

Existing recommendation approaches in folksonomies are based on techniques such as user-based or item-based collaborative filtering,[4] content-based analysis,[5] information retrieval,[6] tag-based user profiling,[7,8] graph-based methods,[9] and hybrids,[10] among others.[11] Whereas most of these works handle tag variations with syntactic transformations such as stemming, we relate tags to concepts in a lexical database to improve the accuracy of similarity. Our experimental results show that semantically enriched similarity improves the precision of recommendation in classical content-based and collaborative filtering approaches. We can thus expect it to produce the same effect in the context of other recommendation approaches.

References

1. G. Srinivas, N. Tandon, and V. Varma, "A Weighted Tag Similarity Measure Based on a Collaborative Weight Model," *Proc. 2nd Int'l Workshop Search and Mining User-Generated Contents* (SMUC 10), ACM, 2010, pp. 79–86.
2. B. Markines et al., "Evaluating Similarity Measures for Emergent Semantics of Social Tagging," *Proc. 18th Int'l Conf. World Wide Web* (WWW 09), ACM, 2009, pp. 641–650.
3. T. Bogers and A. van den Bosch, "Collaborative and Content-Based Filtering for Item Recommendation on Social Bookmarking Websites," *Proc. Workshop on Recommender Systems & the Social Web*, 2009, pp. 9–16.
4. K.H.L. Tso-Sutter, L.B. Marinho, and L. Schmidt-Thieme, "Tag-Aware Recommender Systems by Fusion of Collaborative Filtering Algorithms," *Proc. 2008 ACM Symp. Applied Computing* (SAC 08), ACM, 2008, pp. 1995–1999.
5. M. Lipczak, "Tag Recommendation for Folksonomies Oriented Towards Individual Users," *Proc. European Conf. Machine Learning and Principles and Practice of Knowledge Discovery in Databases Discovery Challenge* (RSDC 08), 2008, pp. 84–95.
6. A. Hotho et al., "FolkRank: A Ranking Algorithm for Folksonomies," *Proc. 2006 Workshop of the Special Interest Group on Information Retrieval* (FGIR 06), 2006, pp. 111–114.
7. S. Sen, J. Vig, and J. Riedl, "Tagommenders: Connecting Users to Items through Tags," *Proc. 18th Int'l Conf. World Wide Web* (WWW 09), ACM, 2009, pp. 671–680.
8. R. Wetzker et al., "I Tag, You Tag: Translating Tags for Advanced User Models," *Proc. 3rd ACM Int'l Conf. Web Search and Data Mining* (WSDM 10), ACM, 2010, pp. 71–80.
9. Z. Guan et al., "Document Recommendation in Social Tagging Services," *Proc. 19th Int'l Conf. World Wide Web* (WWW 10), ACM, 2010, pp. 391–400.
10. A.B. Barragáns-Martínez et al., "Exploiting Social Tagging in a Web 2.0 Recommender System," *IEEE Internet Computing*, vol. 6, no. 14, 2010, pp. 23–30.
11. L.B. Marinho et al., "Social Tagging Recommender Systems," *Recommender Systems Handbook*, Springer, 2011, pp. 615–644.

For setting the threshold of relatedness for two tags, we considered the study from Alexander Busanitsky and Graeme Hirst, which analyzed how well several similarity measures reflect human judgments of semantic relatedness.[3] According to Jiang and Conrath's findings, distance values up to ~10 mean high semantic similarity between two terms in WordNet, so we apply this threshold in our experiments. However, we've yet to determine how this threshold affects the precision of recommendation approaches.

Empirical Evaluation

To empirically evaluate and compare similarity measures, we used two datasets gathered from different social tagging systems: CiteULike, a social bookmarking system for tagging academic papers, and CABS120k08,[4] a collection of webpages with annotations extracted from one of the main social bookmarking sites on the Web, Del.icio.us. The first system is a domain-specific social bookmarking service, and the second a general-purpose one; together, they provide a good perspective on the types of folksonomies available on the Web.

The CiteULike dump we used for our experiments numbers 45,028 users, 299,112 tags, and 1,464,648 articles, related with a total of 5,323,631 tag assignments collected from November 2004 to April 2009. The CABS120k08 dataset contains 117,434 URLs with additional metadata. It comprises 388,963 users, 175,910 unique tags, and 117,434 webpages, related with a total of 3,528,875 tag assignments collected from January 2004 to October 2007.

Social tagging systems on the Web owe their success to the possibility of users freely determining tags for resources without being constrained by a controlled vocabulary, lexicon, or predefined hierarchy. Uncontrolled vocabularies can lead to several problems in the resulting tags, however, such as misspellings, synonyms, and morphological variety. In turn, syntactic mismatches in tags interfere with recommendation algorithms.

We applied a filtering approach to alleviate these problems. First, a filter of compound words replaces symbols such as -, _, or & with whitespace characters to divide individual terms. Second, a dictionary filter verifies whether each individual term exists in an English dictionary. If the word appears in the dictionary, it passes directly to the stemming filter; if not, it's spell-checked using Yahoo's Spelling Suggestion Web service, which provides a suggested spelling correction. The last filter stems the remaining words using the Porter algorithm to solve morphological variations.

To enable the calculation of the semantic similarity of tag pairs, both tags must be present in the WordNet dictionary. WordNet covers 89 percent of the top 100 tags (ordered by frequency) in the CiteULike collection, and 97 percent in the CABS120k08 collection. Considering the top 500, the percentages are 90 and 90.8, and in the top 1,000, they drop to 88.3 and 89 percent, respectively. WordNet's high level of coverage ensures that we can calculate semantic similarity on the basis of available semantic relationships.

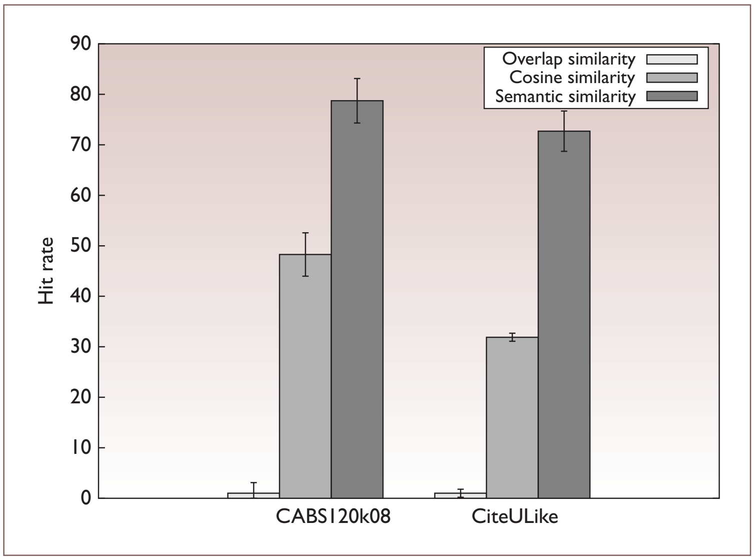

Figure 3. Hit rates of the content-based approach for recommendations in the CABS120k08 and CiteULike datasets. The simple overlap measure performs poorly at recommending resources, and the cosine similarity measure based on tag matching can only recommend less than half of the resources the user is interested in. The semantic similarity measure leads to a higher number of hits.

Resource-to-User Recommendation

In folksonomies, the presence of tags led to the emergence of tag-based profiling approaches, which assume that users expose their preferences for certain content through tag assignments. The resource-to-user recommendation approach focuses on building a representation of resources published in folksonomies and learning user interest profiles to recommend resources matching a profile.

This approach obtains user interest representations by creating a vector of all tags users employed to annotate their resources, weighted according to frequency of use. Likewise, resource representations are given by the overall set of tags assigned to resources in the system with the same weighting strategy. Recommender systems compare both representations using one of the aforementioned similarity measures, and suggest resources exceeding some similarity threshold to users.

For our experiments, we used personomies in both datasets, which provide ground truth about relevant resources, assuming that users are interested in the resources they tagged. Conversely, irrelevant resources aren't available, because this would necessitate acquiring explicit user judgments. We thus focused our experimental evaluation on determining how well semantic similarity can recognize interesting resources that a recommender system can suggest to users with some certainty. We conducted experiments using a holdout strategy that randomly splits a user personomy into 80 percent for training, in which the system learns user interest profiles or vector representations of the user, and 20 percent for testing, used for validation. To make the results less dependent on data splitting, we report the average and standard deviation of 10 runs for each user in all experiments.

We evaluated the recommendations' quality considering the number of hits — that is, the number of resources in the test set that were also present in the recommendation list. The hit rate grants high values to an algorithm that can predict user interests and low values otherwise. If N is the size of the test set, we compute a recommendation algorithm's hit rate as *number of hits*/N.

Figure 3 shows the results we obtained for the resource-to-user recommendation scheme using

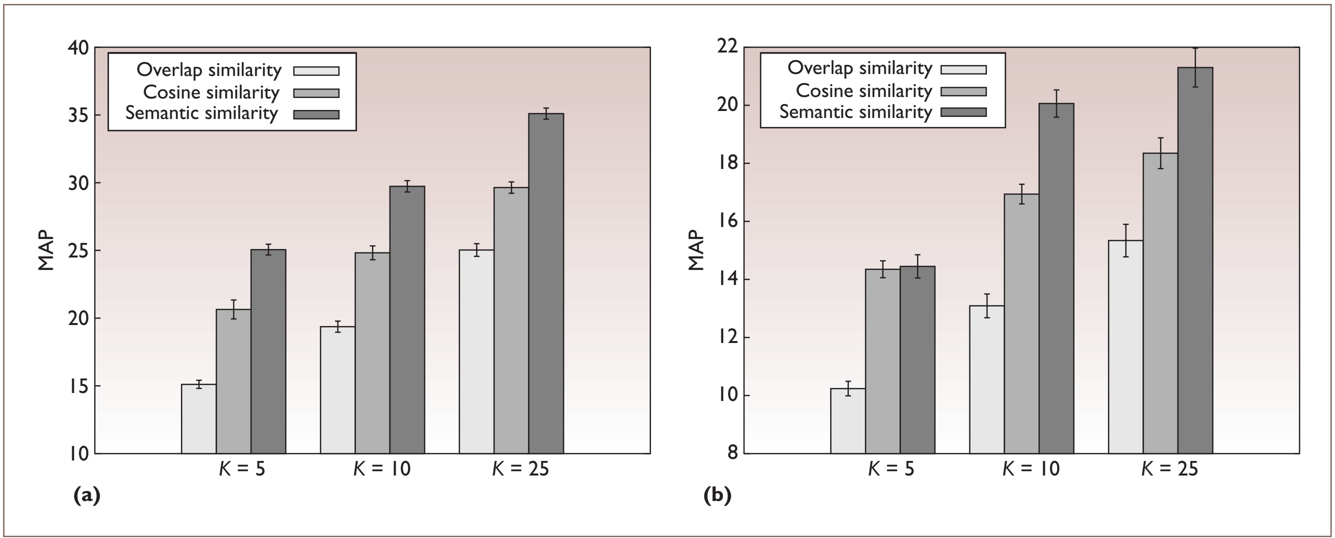

Figure 4. Mean average precision (MAP) scores for the user-to-user approach. We evaluated results for several values of K *in the (a) CABS120k08 and (b) CiteULike datasets.*

the CABS120k08 and CiteULike datasets. We can see that a simple overlap measure performs poorly at recommending resources. Although considerably better, the cosine similarity measure based on tag matching can only recommend less than half of the resources the user is interested in. Once enriched with semantic knowledge, tags lead to a higher number of hits.

User-to-User Recommendation

User-based recommendation assumes that users assigning similar tags share information interests. For recommendation, this approach learns user profiles by building the weighted tag vectors with the tags the users assigned to their resources and searches for the *K* more similar users by comparing the active user's profile with the remaining users in the folksonomy. Finally, the approach employs the resources annotated by the *K* more similar users to generate recommendations, ordering them according to two factors: the distance of the user providing the candidate resource (that is, a resource is more important if a nearer user suggests it) and the number of votes the resource received (the more neighbors that have annotated the resource, the more important it is).

As in the previous experiments, we used a holdout strategy. To reduce the number of users involved in the search for neighbors, we considered only those with at least one resource tagged in common with the active user.

Because this approach produces a ranked list of suggestions, the order in which it presents recommendations becomes important. To evaluate the results, we used mean average precision (MAP), a metric that emphasizes ranking relevant recommendations higher. This measure averages the precisions computed at the point of each relevant recommendation in the ranked sequence.

Figures 4a and 4b show the results obtained for several values of *K* in the CABS120k08 and CiteULike datasets, respectively. User-to-user results show the same trend as resource-to-user ones. The semantic similarity of tag vectors outperforms cosine and overlap similarities based on simple matching.

Hybrid Recommendation

A well-known disadvantage of content-based recommendation approaches (resource-to-user) is that they produce recommendations very similar to the items the user has already seen in the past. In a hybrid approach, combining content-based with collaborative recommendations (user-to-user) adds diversity to the list of suggestions.

To obtain hybrid recommendations, the third approach ranks the candidate resources of the user-to-user approach based on their resource-to-user similarity with user profiles. In this setting, the recommender system must assess two similarities: the collaborative approach's user-user similarity to obtain the *K* nearest neighbors, and the similarity of the candidate resource (only those the target user hasn't already tagged) with the user profile to determine the resource position in the ranked recommendations list.

We conducted experiments using different combinations of cosine and semantic similarities; we discarded the overlap approach because it had the poorest performance. Figures 5a and 5b depict

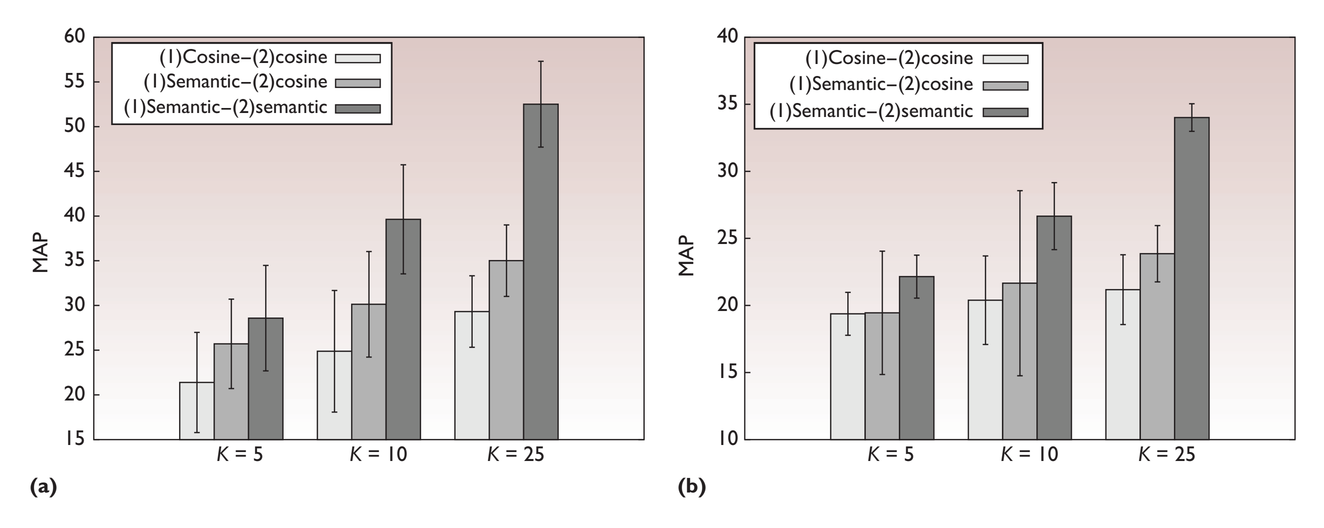

Figure 5. Mean average precision (MAP) scores for the hybrid approach. We evaluated results for several values of K *in the (a) CABS120k08 and (b) CiteULike datasets.*

the MAP scores achieved with CABS120k08 and CiteULike, respectively. MAP scores using semantic similarity in one or both approach steps outperform the use of cosine similarity exclusively. The best-performing combination is the one that uses semantic similarity in both steps.

The empirical study we conducted showed that incorporating semantic knowledge produced better-quality recommendation lists. We thus expect this approach to enhance other recommendation approaches involving similarity calculations in folksonomies. In future research, we will consider other sources of semantic information, such as Wikipedia, and compare them with the WordNet results.

Acknowledgments

This research was partially supported by Argentina's National Scientific and Technical Research Council (CONICET) under grant PIP number 114-200901-00381.

References

1. P. Resnik, "Using Information Content to Evaluate Semantic Similarity in a Taxonomy," *Proc. 14th Int'l Joint Conf. Artificial Intelligence* (IJCAI 95), Morgan Kaufmann, 1995, pp. 448–453.
2. J.J. Jiang and D.W. Conrath, "Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy," *Proc. Int'l Conf. Research in Computational Linguistics* (ROCLING X), 1997, pp. 19–33.
3. A. Budanitsky and G. Hirst, "Evaluating WordNet-Based Measures of Lexical Semantic Relatedness," *Computational Linguistics*, vol. 32, no. 1, 2006, pp. 13–47.
4. M.G. Noll and C. Meinel, "The Metadata Triumvirate: Social Annotations, Anchor Texts, and Search Queries," *Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence and Intelligent Agent Technology* (WI/IAT 08), vol. 1, IEEE, 2008, pp. 640–647.

Daniela Godoy is a full-time professor in the computer science department at the Universidad Nacional del Centro de la Provincia de Buenos Aires (UNICEN), Argentina, as well as a member of the ISISTAN Research Institute and a researcher at the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Her research interests include user profiling, text mining, and recommender systems. Godoy has a PhD in computer science from UNICEN. Contact her at daniela.godoy@isistan.unicen.edu.ar; www.exa.unicen.edu.ar/~dgodoy.

Gustavo Rodriguez received a BSc in systems engineering from the Universidad Nacional del Centro de la Provincia de Buenos Aires (UNICEN), Argentina, where he was a graduate teaching assistant at the time of this work. His thesis was about the integration of semantic knowledge in recommendation technologies applied to social tagging systems. Contact him at rodriguezgustavoandres@gmail.com.

Franco Scavuzzo received a BSc in systems engineering from the Universidad Nacional del Centro de la Provincia de Buenos Aires (UNICEN), Argentina, where he was a graduate teaching assistant at the time of this work. His thesis was about the integration of semantic knowledge in recommendation technologies applied to social tagging systems. Contact him at francoscavuzzo@gmail.com.