



An Incremental Clustering Approach to Personalized Tag Recommendations

Yen-Hsien Lee¹ and Tsai-Hsin Chu²(✉)

¹ Department of Management Information System,
National Chiayi University, Chiayi City, Taiwan
yhlee@mail.ncyu.edu.tw

² Department of E-Learning Design and Management,
National Chiayi University, Chiayi City, Taiwan
thchu@etech.ncyu.edu.tw

Abstract. Volumes of user-generated contents have caused the problem of information overload and hindered Internet users from browsing and retrieving information. Social tagging that allows users to annotate resources with free preferred keywords to ease the access to their collecting resources. Though social tagging benefits users managing their resources, it always suffers the problems such as diverse and/or unchecked vocabulary and unwillingness to tag because tags are freely and voluntarily assigned by users. Tag recommender systems, which follow some criteria to select from the tag space the most relevant tags to the user's annotating resource, drastically transfer the tagging process from generation to recognition to reduce user's cognitive effort and time. This study takes personalized tag recommendation as an incremental clustering problem and proposes a Progressive Expansion-based Tag (PET) recommendation technique. The incremental clustering assumes each object appears in sequence and then is incrementally clustered into either an appropriate existing category or a created new category. The PET technique can classify each resource into multiple categories (i.e., tags) or label it as new. While a resource is labelled as new, it will recommend a set of tags that have been used by other users and are relevant to the target user's practices. Finally, our empirical evaluation results suggest that the proposed PET technique outperforms the traditional popularity-based tag recommendation methods, while the performance rates achieved by both techniques are not satisfying.

Keywords: Tag recommender systems · Personalized tag recommendation · Incremental clustering · Progressive tag expansion · Social tagging

1 Introduction

Applications of Web 2.0 enable people to create and share information on the Internet; however, volumes of user-generated contents have in turn caused the problem of information overload and hindered users from browsing and retrieving information [1]. If not properly addressed, the users would be frustrated by the increasing number of online resources. Recently, some Web 2.0 platforms provide tagging mechanism, namely social tagging, that allows users to annotate resources (e.g., websites, articles,

photos, videos, music, and etc.) with free, preferred keywords to ease the access to their collecting resources in the future. For example, Del.icio.us, a social bookmarking website, enables individuals to bookmark any URLs on the World Wide Web. CiteU-Like, a digital library, allows users to upload abstracts or full-texts of research articles with relevant tags (or labels) and afterward they can retrieve documents through their corresponding tags. Social tagging that takes into account users' notion of a specific resource [2] is helpful in organizing, browsing and retrieving their own resources [3]. In other words, social tagging allows resources to be categorized in the way a particular user prefers to, and therefore is considered as a substitute for taxonomy [4, 5].

Social tagging can benefit people managing their online resources; on the other hand, tags from individual users represent their personal preferences and can be used to improve the performance of personalized recommendation if properly utilized [6]. However, social tagging always suffers the problems such as diverse and/or unchecked vocabulary and unwillingness to tag because tags are freely and voluntarily assigned by users [7, 8]. The problem of diverse vocabulary may result from users' forgetfulness and bounded rationality. Ebbinghaus hypothesized that the memory retention declines over time and people may lose about 70% of the information received two days ago without any attempt on retention [9]. Though the loss of information can be mitigated by constant recall, it persists when people fail to review frequently [10]. Furthermore, the bounded rationality also limits users' ability to process information that disabled them from recalling all the tags they have used [11]. As a result, users might tend to reuse the most frequent terms (vocabularies) or use different terms each time they annotate similar resources. Thus, as the number of annotated resources increased, the tag space would become vast and the resources related to a tag would become heterogeneous. Both might frustrate users in accessing resources due to the cognitive dissonance [12, 13].

To address the problems faced by social tagging, some studies have shifted focus on tag recommender systems to assist individual users in tagging resources and converge the tags attached [3, 14–18]. Tag recommendation service has been provided by some websites, such as Delicious, BibSonomy, and Last.fm, that implies the needs in real-world situation. The task of tag recommender systems is to identify a set of tags that might be considered relevant to a resource by the focal user. Specifically, given a user u and a resource r , the task of traditional recommendation is to predict the class of $preference(u, r)$; while that of tag recommendation is to predict the set of $tags(u, r)$ what the user u will assign to the resource r [7].

For making tag recommendations, previous research assumed the number of tags was static; that is, users are limited to annotating resources with existing tags. For example, they followed collaborative filtering methods by identifying users whose used tags or annotating resources are similar to that of the focal user and suggest those tags annotating to the similar resources by these users [19]. On the other hand, some research addresses the tag recommendation problem by content-based approaches. For example, while annotating a resource, some studies tried to identify the resources that share similar content with the focal resource and then recommended the top-ranked tags annotated to them to the user [20]. Chen and Shin proposed several textual features and social features for each tag used by each particular user and use which to construct a classifier to predict the representative tags that the focal user is interested in [21].

Though prior studies have shown the effectiveness of their proposed approaches in making tag recommendations, they still have some limitations needed to be addressed. A resource will be suggested one or more used tags compulsorily no matter whether they are relevant or not. However, the tags that people used to annotate resource might evolve over time. Some tags that receive less notice will be left behind; while some tag shall emerge from as annotating new resources. Reasonably, these tags shall be relevant to the annotating resource and related to the particular user's topics of interest (the tags he or she has used to annotate resources) to better help users retrieve the resources later. That is, the new tags must to some degree conform to or associate with users' practices. In a nutshell, a tag recommender system shall make suggestion on the basis of existing tags and is able to recommend new tags that are appropriate and associated with users' practices.

Nevertheless, prior research focused most on the reuse of existing tags and accordingly attempted to recommend people those tags that are popular among the referred users or frequently used to annotate similar resources. Generally, people annotate resources one by one, and each resource will be assigned one or more tags. The assigned tags can be existing ones that people used to annotate previous resources (including the previous one), or created by the users if there is no proper tag existing, or both. As a result, this study intends to improve the personalized tag recommendations by suggesting appropriate existing tags or new tags to the target user. Instead of multi-label classification, we adopt the content-based approach and model personalized tag recommendation as an incremental clustering problem. The incremental clustering assumes each object (or resource) appears in sequence and then is incrementally clustered into either an appropriate existing category or a created new category [22, 23]. This study extends the incremental clustering approach and propose a progressive expansion-based tag (PET) recommendation technique. The proposed PET recommendation technique assumes the resources to be annotated are fed in sequence and will be assigned one or more existing categories (i.e., existing tags) and/or suggested appropriate new categories (i.e., new tags). In addition, when determining the appropriate existing tags for a resource, the PET technique will consider the focal user's topics of interest. For example, instead of identifying similar resources to make tag recommendations, PET tries to measure the relevance between a new resource and a tag. It measures the content similarity between the resource to be annotated and all resources annotated by the tag. Furthermore, to suggest new tags, the PET will identify the representative term(s) in the resource to be annotated by measuring not only the term frequency but also the relevance to existing tags. The remainder of this study is organized as follows: In Sect. 2, we review the literature relevant to this study. Then we depict our proposed Progressive Expansion-based Tag (PET) recommendation technique in Sect. 3. In Sect. 4, we describe the empirical evaluation including the data collection, evaluation design, followed by the evaluation results in Sect. 5.

2 Literature Review

In this section, we briefly review the research works relevant to our proposed progressive expansion-based tag recommendation technique, including prior research in tag recommender systems and an overview of incremental clustering.

2.1 Tag Recommender Systems

Tag recommender is one kind of recommender systems. Instead of recommending objects such as books, music, or movies, the purpose of tag recommender is to suggest appropriate tags to users who are annotating objects in the social media; especially the social bookmarking and the media sharing websites. Such websites generally provide the social tagging mechanisms that allow users to annotate objects with free keywords. For example, the social bookmarking website, Del.icio.us enables individuals to bookmark any URLs in the World Wide Web; the digital library, CiteULike allows users to upload the abstract or full-text of research articles with some relevant tags (or labels); the famous video sharing website, YouTube allows users to upload their videos with some tags. Though user-generated content is the core to Web 2.0, Internet users have been overloaded with the great volumes of information that hinders them from browsing and retrieving information [1]. Social tagging can benefit users managing and accessing their online resources; on the other hand, the tags annotated by an individual user may represent his or her notions of a resource that can facilitate the personalized recommendation if properly utilized [2, 3, 6].

Use of tags allows users to annotate resources in the way they like, and therefore, tagging is somehow considered as a substitute for the taxonomy of user's resources [4, 5]. Nevertheless, social tagging always suffers the problems such as diverse and/or unchecked vocabulary and unwillingness to tag because tags are freely and voluntarily assigned by users [7, 8]. Besides, users tend to reuse the frequent tags or to create new tags, which will diminish the coherence or distinctness of the resources with a specific tag and adversely affect users' resource searches and access due to the cognitive dissonance [12, 13]. To address the problems faced by social tagging, prior research attempts to develop tag recommender systems to support users in annotating resource to converge the tags attached [3, 14–18]. Tag recommendation may drastically transfer the tagging process from generation to recognition which reduces user's cognitive effort and time [20]. A tag recommender system follows some criteria to select from the tag space the most relevant tags to the user's uploading resource. Specifically, given a user u and a resource r , the task of tag recommendation is to predict a set of tags(u, r) from a finite set of tags T that the user u may prefer to annotate the resource r [7].

Prior research broadly divided the tag recommendations into content-based, collaborative filtering, and graph-based (or ranking-based) approaches according to their adoptive algorithms [3, 7]. The content-based approaches focus on content analysis and are mainly applied textual resources like webpages and textual documents [21, 24–31]. Instead of analyzing contents, the collaborative filtering approaches for tag recommendation resemble traditional collaborative filtering recommendation approaches which make recommendations on the basis of the preferences of a referent group [19, 20, 32]. Finally, the graph-based or ranking-based approaches are inspired from the Web ranking. They make recommendations based on the ranking score that is computed according to spectral attributes extracted from the underlying folksonomy data structure (i.e., the 3-way relationship among users, resources, tags) [7, 17, 33, 34].

Overall, prior research focused most on the reuse of existing tags and attempted to recommend people those tags that are popular among the referred users or frequently used to annotate similar resources. Though users' interests may evolve over time, they

seldom take into consideration the user's topics of interest when making tag recommendations. Besides, people annotate resources one by one and always create new tags combining with existing tags to annotate them. Tag recommendations shall be made with consideration of user's interests and that is what we intend to address in this study.

2.2 Incremental Clustering

Clustering analysis methods usually employ the batch mode strategy to discover the structure hidden in the whole unlabeled data at a time. However, the sheer volume of data available for clustering analysis has made the memory-based approach impractical, and thus raise the need of incremental clustering approaches, which process one object at a time and require less memory space for data storage [35]. One of the well-known incremental clustering algorithms is sequential k-means [36], which is an incremental variant of Lloyd's algorithm [37]. The sequential k-means algorithm targets on finding a set of cluster means M that minimizes the cost function $\sum_{\forall o_j \in O} \min_{m \in M} \|o_j - m\|^2$. It randomly initials k data points as cluster means $M = (m_1, m_2, \dots, m_k)$ and set to 1 the size of each cluster $N = (n_1, n_2, \dots, n_k)$. As an object o_j arrives, Euclidean distance between the object o_j and each of the cluster means will be calculated in sequence. Assume the object o_j is classified into its closest cluster c_i , the size of cluster c_i (i.e., n_i) will be increased by 1 and the mean of cluster c_i (i.e., m_i) will be updated by $m_i + (o_j - m_i)/n_i$.

Yang et al. [23, 38] addressed the news event detection problem by proposing INCR, a single-pass incremental clustering algorithm, produces nonhierarchical clusters incrementally for both retrospective and online detection. For supporting online detection, INCR was designed to sequentially process news documents. It employed an incremental IDF to respond the effect of continuously incoming documents on term weighting and vector normalization during online detection. The incremental IDF is defined as $idf(w, p) = \log_2(\frac{N(p)}{n(w, p)})$, where w is the focal term, p is the current time point, $N(p)$ is the number of documents accumulated up to the current time point (including the retrospective corpus if used), and $n(w, p)$ is the document frequency of term w at time point p . Furthermore, INCR incorporated a time penalty, which can be a uniformly weighted time window (i.e., a time window of m documents before x is imposed) or a linear decaying-weight function, to adjust the similarity between a document x and any cluster c in the past. The similarity measure can be cosine similarity or any distance measure like Euclidean distance. The *Similarity*'(x, c) is defined as $\begin{pmatrix} (1 - \frac{i}{m}) \times \text{Similarity}(x, c) & \text{if } c \text{ has any member in the window} \\ 0 & \text{otherwise} \end{pmatrix}$, where i is the number of documents between x and the most recent member document in c , and m is the time window of documents before x . Finally, a document x is absorbed by the most similar cluster in the past if the similarity between the document and cluster is larger than a pre-selected *clustering threshold* (t_c); otherwise, the document becomes the seed of a new cluster.

3 Progressive Expansion-Based Tag (PET) Recommendation Technique

Our study intends to propose a Progressive Expansion-based Tag (PET) recommendation technique by revising an incremental clustering algorithm. The PET technique considers a focal user's interests to recommend the appropriate categories (tags) to the resources for the focal user. On the other hand, the PET tries to recommend tags by identifying the relevant tags from the tags annotated to the focal resource by other users if the focal user's own tags are less appropriate. As shown in Fig. 1, the overall process of the PET technique comprises four phases, including feature extraction and selection, resource representation, candidate tag generation, and tag recommendation. The PET technique takes as inputs a focal user's resource profile (i.e., resources with their respective annotated tags) and the resources to be annotated and produces a list of tags to be recommended. Because the PET considers user's (resources) interests, we first group the resources in the user's profile by their attached tags. Reasonably, two resources that attached the same tag may discuss similar topic or share similar content. A set of important features will then be selected and used to represent resources in each tag cluster. Subsequently, an incremental clustering algorithm is applied to determine a set of appropriate tag clusters for the resources to be annotated. A resource will be classified into a tag cluster if the content similarity between them is over a pre-specified threshold and these tag clusters then become the candidates for recommendations. If a resource could not be classified into suitable tag cluster, the PET will access appropriate tags used by other users. In the following, we describe the preliminary design of the proposed PET technique.

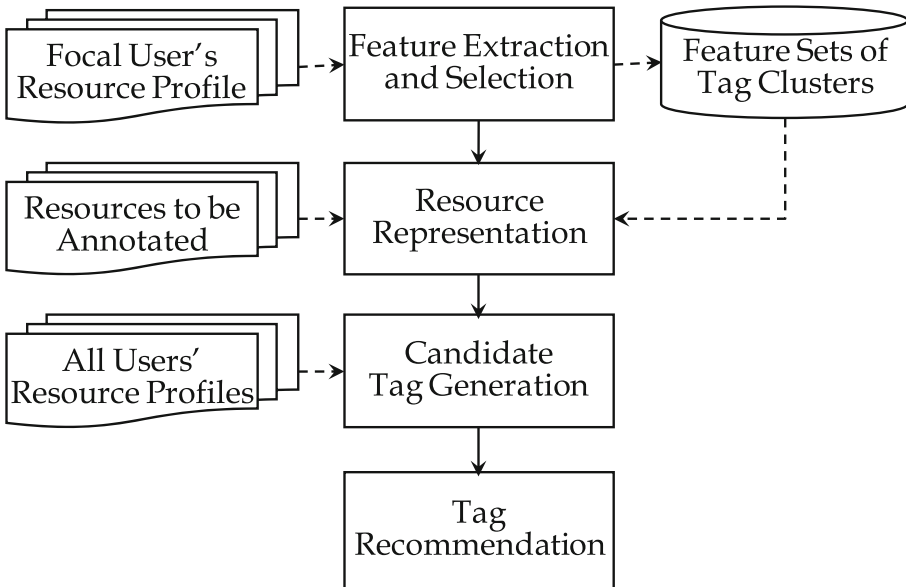


Fig. 1. Overall process of progressive expansion-based tag recommendation technique

Feature Extraction and Selection: In the feature extraction and selection phase, the resources in the user's profile are grouped by their respective attached tags to form a set of tag clusters. One resource could belong to multiple groups since it might be attached more than one tags. The PET then extracts from the textual resources a set of representative features (i.e., nouns and noun phrases) for representing the resources themselves. We adopted the rule-based part-of-speech tagger developed by Brill to syntactically tag each word in these resources [39]. Subsequently, we employed a parser for extracting nouns and verbs from each syntactically tagged document. The global dictionary scheme was adopted and the chi-square statistic was used to measure to the weight of each feature for constructing the representative feature set of each cluster [40].

Resource Representation: In the resource representation phase, the resources in each cluster are represented by its set of representative features. In this study, we employed TFxIDF measure as the representation scheme to re-represent the resources in each cluster.

Candidate Tag Generation: The purpose of candidate tag generation phase is to assess and identify the tags relevant to the resource to be annotated. This phase comprises two stages, including tag cluster identification and new tag generation. At the stage of tag cluster identification, this study revised the INCR algorithm [23, 38] to enable supporting multi-label classification. Specifically, INCR algorithm assumes each object belongs to one and only one cluster. However, in our study, a resource can belong to any number of tag cluster; that is, a resource might be different to the resources in the focal user's profile or belong to more than one tag cluster. As a result, we accommodate INCR algorithm to be able to assign a resource into multiple tag clusters or create a new cluster for it if needed. We followed the INCR algorithm by employing a clustering threshold. The tag clusters that share similarities with a resource higher than the clustering threshold will be viewed as candidate tags for recommendations. However, when a resource is labeled as new; that is, all the similarities it achieves are lower than the clustering threshold, we will try to identify suitable tags for recommendation from the annotated resources of other users. Thus, the task of new tag generation is to assess suitability of the tags that was annotated to the focal resource by other users. We rank those tags by considering their respective frequency appearing in the whole resources, their relevance associated to the resources that the focal user has annotated, and their temporal distance to the resource to be annotated. The frequency TF is defined as the number of a tag that is used to annotate resources; the relevance TR is defined as the content similarity between a specific tag cluster (i.e., the resources received the specific tag) and the resources in the focal user's profile; the temporal distance TD is defined as $e^{-\frac{|Now-Date(t_i)|}{Now-Date(T)}}$ where t_i is the tag to be assessed, T is the set of all candidate tags, $Date(t_i)$ is the starting date to use tag t_i and $Date(T)$ is the starting date to use anyone of the candidate tags. We finally defined the ranking score of a specific tag t_i as $Score(t_i) = TF \times TR \times TD$.

Tag Recommendation: The task of the final phase of PET technique is to make tag recommendations. PET will first recommend tags identified at the stage of tag cluster

identification, and if needed, the tags identified at the new tag generation stage will be recommended to satisfy the number of recommending tags. The candidate tags from focal user's profile will be ordered by their achieved similarities and those from other users' profile will be ordered by their ranking scores.

4 Empirical Evaluation

4.1 Data Collection

We adopted the MovieLens 20M database (ml-20m) as our evaluation corpus. This database contains 465,564 tag applications across 27,278 movies, created by 138,493 users who have rated at least 20 movies between January 09, 1995 and March 31, 2015. Among the database, the max, min, and average number of tags used by a user is 2,330, 1, and 58.1; the max, min, and average number of tags received by a movie is 197, 1, and 15.14; the max, min, and average number of movies that a specific tag was annotated to is 1,093, 2, and 18.03. Because the tags annotated to the movies in the evaluation corpus is sparse, we adopted the p -core scheme to tri-partite hypergraphs to trim the corpus and keep its dense part for the evaluation purpose [41, 42]. Finally, we set the level k to 3 for the p -core scheme to make sure that each user, tag and resource has/occurs at least 3 times in the evaluation corpus. After the trimming, there exists 7,801 users, 19,545 movies, and 364,804 tagging records in the evaluation corpus. Besides, we also collected the synopsis of each annotated movie for the experiments. We implemented a crawler to gather the overview of each movie from TheMovieDb website (<https://www.themoviedb.org/>) through the movie ID provided by MovieLens database.

4.2 Experiment Design

For each user in the evaluation corpus, we take his or her last annotating movie and corresponding tags as testing examples, and all users' tagging histories (i.e., all other annotating movies and corresponding tags) as training examples. In this study, we implemented two popularity-based recommendation approaches, namely PAT and PUT as the performance benchmarks. In PAT, the top- n tags that are frequently used to annotate resources by all users will be recommended; on the other hand, in PUT, the top- n tags that are frequently used to annotate resources by the focal user will be recommended. Furthermore, we adopted Precision, Recall, Hamming Loss, Mean Reciprocal Rank (MRR) [43], Average Precision (AP) [44], and Average Utility

(AU) as the evaluation criteria. These criteria are defined as $Precision = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|P_i \cap T_i|}{|T_i|}$,

$Recall = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|P_i \cap T_i|}{|P_i|}$, $Hamming\ Loss = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|P_i \Delta T_i|}{|P_i|}$, $MRR = \frac{1}{|D|} \sum_{i=1}^{|D|} \sum_{j \in P_i \cap T_i} \frac{1}{Rank_j}$,

$AP = \frac{1}{|D|} \sum_{i=1}^{|D|} \sum_{j \in P_i \cap T_i} \frac{Precision_j}{|P_i \cap T_i|}$, and $AU = \frac{1}{|D|} \sum_{i=1}^{|D|} \sum_{j \in P_i \cap T_i} \frac{Precision_j}{|P_i|}$, where $|D|$ is the number of target movies, P_i is the set of recommended tags for the target movie d_i , and T_i is the set

of true tags annotated to the target movie d_i , Δ is the XOR operation, $Rank_j$ is the rank of the recommended tag j , and $Precision_j$ is the precision at the time tag j is recommended. Finally, we set the clustering threshold for incremental clustering algorithm to 0.05 and examine the overall effectiveness of our proposed PET and benchmark techniques by averaging the recommendation performance across all users.

5 Evaluation Results

We investigate the effectiveness of both evaluation techniques when the number of recommended tags is three and five. As shown in Table 1, our proposed PET outperforms the benchmarks, i.e., PAT and PUT techniques, across all performance metrics when making recommendation of three and five tags. Though the performance of PET is advantageous over the benchmarks, the rates it achieves across all performance metrics are not satisfying. Furthermore, as the number of recommended tags increased, there is a tradeoff existing in precision and recall rates. However, almost all the rates it achieves are lower than 0.1 except for the recall rate. The evaluation results imply the difficulty of tag recommendation that must identify relevant tags among thousands of candidate tags. Overall, the performance of the proposed PET technique is better than the benchmark technique, which make tag recommendations on the basis of tag's popularity. Besides, the low performance rates may be raised by the sparse data, that is still a problem needed to be addressed in the study of tag recommendation.

Table 1. Comparative evaluation results

	Precision	Recall	HL	MRR	AP	AU
Number of recommending tags = 3						
PAT	0.021	0.015	0.990	0.011	0.042	0.015
PUT	0.034	0.064	0.973	0.038	0.059	0.021
PET	0.060	0.107	0.954	0.062	0.099	0.038
Number of recommending tags = 5						
PAT	0.022	0.027	0.987	0.013	0.050	0.011
PUT	0.037	0.097	0.969	0.046	0.069	0.019
PET	0.062	0.173	0.949	0.078	0.120	0.031

6 Conclusion

This study based on the concept of incremental clustering to propose a progressive expansion-based tag recommendation technique. The PET technique can recommend appropriate tags to the resources to be annotated in consideration of the focal user's preference and tag usage practices. The preliminary evaluation results indicated that the proposed PET technique is more effective than the popularity-based tag recommendation approaches across all evaluation criteria. The progressive expansion approach can identify tags to meet user's needs in annotating online resources. However, this

study has some limitations need to be addressed which in turns become the future research directions. First, we only adopted one database (i.e., MovieLens 20M database) to evaluate and compare the investigated techniques. More experimental datasets shall be collected from the other social bookmarking websites, such as BibSonomy, CiteULike, and Last.fm for carrying out more empirical evaluations. Second, this study employed two popularity-based recommendation approaches as the performance benchmarks. Other approaches to tag recommendation shall also be examined in the future. Finally, the experimental evaluations we conducted in this study are preliminary, and thus it requires more analyses on the effects of the proposed PET technique.

Acknowledgement. This work was supported by Ministry of Science and Technology of the Republic of China under the grant MOST 106-2410-H-415-009.

References

1. Lee, B.K., Lee, W.N.: The effect of information overload on consumer choice quality in an on-line environment. *Psychol. Mark.* **21**, 159–183 (2004)
2. Wu, H., Zubair, M., Maly, K.: Harvesting social knowledge from folksonomies. In: *The 17th Conference on Hypertext and Hypermedia*, pp. 111–114. ACM Press (2006)
3. Musto, C., Narducci, F., de Gemmis, M., Lops, P., Semeraro, G.: A Tag recommender system exploiting user and community behavior. In: Jannach, D., et al. (eds.) *The ACM RecSys 2009 Workshop on Recommender Systems & the Social Web*, vol. 532. CEUR-WS.org (2009)
4. Heckner, M., Heilemann, M., Wolff, C.: Personal information management vs. resource sharing: towards a model of information behaviour in social tagging systems. In: *The Third International AAAI Conference on Weblogs and Social Media Conference*, pp. 42–49 (2009)
5. Marvasti, A.F., Skillicorn, D.B.: Structures in collaborative tagging: an empirical analysis. In: *The Thirty-Third Australasian Conference on Computer Science*, vol. 102, pp. 109–116 (2010)
6. Yang, C.S., Chen, L.C.: Personalized recommendation in social media: a profile expansion approach. In: *The 18th Pacific Asia Conference on Information Systems* (2014)
7. Marinho, L.B., et al.: Social tagging recommender systems. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.) *Recommender Systems Handbook*, pp. 615–644. Springer, Boston, MA (2011). https://doi.org/10.1007/978-0-387-85820-3_19
8. Bischoff, K., Firan, C.S., Nejdil, W., Paiu, R.: Can all tags be used for search? In: *The 17th ACM Conference on Information and Knowledge Management*, pp. 193–202 (2008)
9. Ebbinghaus, H.: *Memory: A Contribution to Experimental Psychology*. Columbia University, New York (1885)
10. García, R.R., Quirós, J.S., Santos, R.G., González, S.M., Fernanz, S.M.: Interactive multimedia animation with macromedia flash in descriptive geometry teaching. *Comput. Educ.* **49**, 615–639 (2007)
11. Simon, H.A.: *Models of Bounded Rationality*. MIT Press, Cambridge (1997)
12. Oliver, R.L.: A cognitive model for the antecedents and consequences of satisfaction. *J. Mark. Res.* **17**, 460–469 (1980)
13. Wei, C.P., Hu, P., Lee, Y.H.: Preserving user preferences in automated document-category management: an evolution-based approach. *J. Manag. Inf. Syst.* **25**, 109–143 (2009)

14. Cattuto, C., et al.: Network properties of folksonomies. *AI Commun.* **20**, 245–262 (2007)
15. Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., Stumme, G.: Tag recommendations in social bookmarking systems. *AI Commun.* **21**, 231–247 (2008)
16. Shepitsen, A., Gemmell, J., Mobasher, B., Burke, R.: Personalized recommendation in social tagging systems using hierarchical clustering. In: *The 2008 ACM Conference on Recommender Systems*, pp. 259–266. ACM (2008)
17. Symeonidis, P., Nanopoulos, A., Manolopoulos, Y.: Tag recommendations based on tensor dimensionality reduction. In: *The 2008 ACM Conference on Recommender Systems*, pp. 43–50. ACM (2008)
18. Tso-Sutter, K.H.L., Marinho, L.B., Schmidt-Thieme, L.: Tag-aware recommender systems by fusion of collaborative filtering algorithms. In: *The 2008 ACM Symposium on Applied Computing*, pp. 1995–1999. ACM (2008)
19. Marinho, L.B., Schmidt-Thieme, L.: Collaborative tag recommendations. In: Preisach, C., Burkhardt, H., Schmidt-Thieme, L., Decker, R. (eds.) *Data Analysis Machine Learning and Applications*, pp. 533–540. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-78246-9_63
20. Sood, S., Owsley, S., Hammond, K., Birnbaum, L.: TagAssist: automatic tag suggestion for blog posts. In: *The International Conference on Weblogs and Social Media* (2007)
21. Chen, X., Shin, H.: Tag recommendation by machine learning with textual and social features. *J. Intell. Inf. Syst.* **40**, 261–282 (2013)
22. Hartigan, J.: *Clustering Algorithms*. Wiley, New York (1975)
23. Yang, Y., Carbonell, J.G., Brown, R.D., Pierce, T., Archibald, B.T., Liu, X.: Learning approaches for detecting and tracking news events. *IEEE Intell. Syst.* **14**, 32–43 (1999)
24. Brooks, C.H., Montanez, N.: Improved annotation of the blogosphere via autotagging and hierarchical clustering. In: *The 15th International Conference on World Wide Web*, pp. 625–632. ACM Press (2006)
25. Lee, S., Chun, A.: Automatic tag recommendation for the web 2.0 blogosphere using collaborative tagging and hybrid semantic structures. In: *The 6th Conference on WSEAS International Conference on Applied Computer Science*, Stevens Point, Wisconsin, pp. 88–93 (2007)
26. Heymann, P., Ramage, D., Garcia-Molina, H.: Social tag prediction. In: *The 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 531–538 (2008)
27. Tsoumakas, G., Katakis, I.: Multi-label classification: an overview. *Int. J. Data Warehouse Min.* **3**, 1–13 (2007)
28. Gonçalves, T., Quaresma, P.: A preliminary approach to the multilabel classification problem of Portuguese juridical documents. In: Pires, F.M., Abreu, S. (eds.) *EPIA 2003. LNCS (LNAI)*, vol. 2902, pp. 435–444. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-24580-3_50
29. Boutell, M.R., Luo, J., Shen, X., Brown, C.M.: Learning multi-label scene classification. *Pattern Recogn.* **37**, 1757–1771 (2004)
30. Lauser, B., Hotho, A.: Automatic multi-label subject indexing in a multilingual environment. In: Koch, T., Sølvberg, I.T. (eds.) *ECDL 2003. LNCS*, vol. 2769, pp. 140–151. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-45175-4_14
31. Zhang, M.L., Zhou, Z.H.: ML-kNN: a lazy learning approach to multi-label learning. *Pattern Recogn.* **40**, 2038–2048 (2007)
32. Mishne, G.: Autotag: a collaborative approach to automated tag assignment for weblog posts. In: *The 15th International Conference on World Wide Web*, pp. 953–954. ACM Press, New York (2006)

33. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information retrieval in folksonomies: search and ranking. In: Sure, Y., Domingue, J. (eds.) *ESWC 2006*. LNCS, vol. 4011, pp. 411–426. Springer, Heidelberg (2006). https://doi.org/10.1007/11762256_31
34. Rendle, S., Marinho, L.B., Nanopoulos, A., Schmidt-Thieme, L.: Learning optimal ranking with tensor factorization for tag recommendation. In: *The 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Paris, France, pp. 727–736. ACM Press (2009)
35. Ackerman, M., Dasgupta, S.: Incremental clustering: the case for extra clusters. In: *2014 Annual Conference on Neural Information Processing Systems*, Montreal, Quebec, Canada, pp. 307–315 (2014)
36. Charikar, M., Chekuri, C., Feder, T., Motwani, R.: Incremental clustering and dynamic information retrieval. In: *The Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pp. 626–635. ACM Press (1997)
37. Lloyd, S.P.: Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28**, 129–137 (1982)
38. Yang, Y., Pierce, T., Carbonell, J.G.: A study on retrospective and on-line event detection. In: *21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 28–36, Melbourne, Australia. ACM Press (1998)
39. Brill, E.: Some advances in rule-based part of speech tagging. In: *Proceedings of the 12th National Conference on Artificial Intelligence (AAAI-94)*, pp. 722–727. AAAI Press (1994)
40. Apté, C., Damerau, F., Weiss, S.: Automated learning of decision rules for text categorization. *ACM Trans. Inf. Syst.* **12**, 233–251 (1994)
41. Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., Stumme, G.: Tag recommendations in folksonomies. In: Kok, J.N., Koronacki, J., Lopez de Mantaras, R., Matwin, S., Mladenič, D., Skowron, A. (eds.) *PKDD 2007*. LNCS (LNAI), vol. 4702, pp. 506–514. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-74976-9_52
42. Vladimir, B., Matjaž, Z.: Generalized cores. arXiv preprint (2002)
43. Deshpande, M., Karypis, G.: Item-based top-n recommendation algorithms. *ACM Trans. Inf. Syst.* **22**, 143–177 (2004)
44. Chowdhury, G.: *Introduction to Modern Information Retrieval*. Facet Publishing, London (2010)