Priyanka Chakraborty
‡ • ‡ ƒ ” … Š • IIM Calcutta, Joka, Diamond Harbour Road, Kolkata 700104, India

Assistant Professor, IIM • † ‘ ” ‡ á ƒ — æ ‹ – Š ƒ • ’ — ” ‘ ƒ † á • † ‘ ” ‡ v w u u u s

&

Professor, IIM Calcutta, Joka, Diamond Harbour Road, Kolkata 700104, India

ã

Priyanka Chakraborty
‡ • ‡ ƒ ” … Š • • ‹ • – ƒ • – á– ƒ•• ƒ ˆ ƒ ‰ ƒ – ‡ ‘•• › • – ‡ • • ” ‘ — ’
• † ‹ ƒ • • • – ‹ – — – ‡ ‘ ^ ƒ • ƒ ‰ ‡ • ‡ • – ƒ Ž … — – – ƒ
‹ ƒ • ‘ • † ƒ ” „ ‘ ” ‘ ƒ †
‘ Ž • ƒ – ƒ y r r s r v á
• – ƒ – ’ ” ‹ › ƒ • • ƒ 7 ‰ • ƒ ‹ Ž ä … ‘ •

Sanjog Ray
ƒ … — Ž – › á ƒ • ƒ ‰ ‡ • ‡ • – • ^ ‘ ” • ƒ – ‹ ‘ • › • – ‡ • • ” ‘ — ’
• † ‹ ƒ • • • – ‹ – — – ‡ ‘ ^ ƒ • ƒ ‰ ‡ • ‡ • – • † ‘ ” ‡
ƒ — æ ‹ – Š ƒ • ’ — ” ‘ ƒ †
• † ‘ ” ‡ v w u u u s á
• ƒ • Œ ‘ ‰ ” 7 ‹ ‹ • ‹ † ” ä ƒ … ä ‹ •

Ambuj Mahanti
” ‘ ^ ‡ • • ‘ ” á ƒ • ƒ ‰ ‡ • ‡ • • ƒ — ‹ ‘ •• › • – ‡ • • ” ‘ — ’
• † ‹ ƒ • • • – ‹ – — – ‡ ‘ ^ ƒ • ƒ ‰ ‡ • ‡ • – ƒ Ž … — – – ƒ
‹ ƒ • ‘ • † ƒ ” „ ‘ ” ‘ ƒ †
‘ Ž • ƒ – ƒ y r r s r v á
ƒ • 7 ‹ ‹ • … ƒ Ž ä ƒ … ä ‹ •

ã

Recommender systems are one of the most popular systems that help in personalization of ecommerce sites and collaborative web sites. Tags are textual description given to resources by users and are prevalent in most web 2.0 websites as they help in organization and retrieval of resources. In this paper we present an overview of the way tags are used in recommender systems and also on the aspect of recommendation of tags for items. This paper surveys the work done on two aspects of recommendations based on tags (a) Recommending items to users and (b) Recommending tags for items. This paper also discusses possible extensions that can improve recommendation capabilities using tags and make better tags recommendation for items.

Tagommenders, Tag suggestion, Tags, Recommender Systems

Recommender systems are technology-based systems that provide personalized recommendations to users. In these systems, opinions and actions of other users with similar tastes are used to generate recommendations. Recommender systems primarily use ratings data given by a user to different items present in the system to make personalized recommendations. Recommender systems are a ubiquitous feature in most ecommerce sites such as Amazon.com, Ebay.com, Netflix.com, Last.fm etc. Recommendation systems popularity is not only because of their ability to provide personalization features but also due to their impact in higher sales and profits. In[Chen, Wu et al. 2004] , it has been shown empirically on Amazon.com dataset that recommender systems indeed improved sales. However, with increasing popularity of recommender systems in ecommerce sites they have become susceptible to attacks by malicious users who try to influence the systems by inserting biased data into the system [Mobasher, Burke et al. 2007]. Recent research on trust aware recommender systems [Massa and Avesani 2004; Golbeck 2006; Massa and Avesani 2007] has shown that they are more robust against shilling attacks and are more capable of generating recommendations for new users in the system. Trust aware systems also have been shown to produce recommendations which are better than or as accurate as collaborative filtering based recommender systems. Trust

aware systems are able to make more accurate recommendation compared to traditional systems as they use the concept of trust propagation over a trust network. Because of these advantages over traditional systems, trust aware recommender systems are generating much research interest.

Traditional recommender systems help users in tackling the problem of information overload by recommending products and services that will be of interest to them from the vast universe of choices available. Popular ecommerce sites like Amazon.com suggests products users may like based on their ratings, clicked items, and purchased items [1]. Users of Digg.com receive news articles based on other articles they find interesting [2]. Netflix provides movie recommendations based on movie ratings [3]. Because of the enormous use of recommender systems in practical applications on the web, research interest in the area is immense.

Traditional recommender systems primarily use only ratings for recommending items but recent work has shown that using tags data along with the ratings seems to improve upon the traditional method of recommending. Tags are widely used for information organization on the web. A social tag is a piece of brief textual information given by users explicitly to describe and group items, thus it implies users' interests or preferences information. The social tags in web 2.0 are becoming another important information source to profile users' interests and preferences for making personalized recommendations. Tagging has emerged as a powerful tool that enables users to find, organize and understand online entities. Numerous efforts have been made to better understand and exploit the use of tags and their usage patterns.

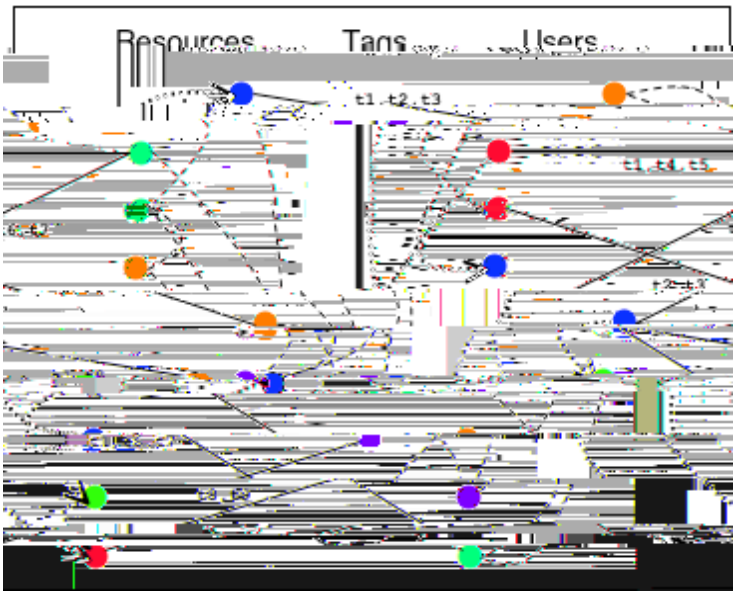
Tagging offers users an alternate way to organize items. As users create tags, they represent concepts meaningful to them. Tags offer flexibility to the users in describing the item content and as a result provide personalized information about the user to the recommender system that leads to improved recommendation quality. A model of tagging system is shown in figure1. There are three ways in which tags play a role in recommender systems. First tags information is used in addition to item rating and content information for recommending items to users, secondly, algorithms are designed for recommending tags for items and finally tags information can be used to recommend users.

In this paper, we present a comprehensive survey of different approaches used in solving the problem of recommending items to users using tags and the problem of recommending tags for items. We first describe the approaches used for tag based item recommendation in section 2 and in section 3 we survey the approaches used in

recommending tags for item. In section 4 we discuss way in which existing capabilities of tag based recommender systems can be extended.

Definition of popular terms used in tagommenders.

Folksonomy	The term folksonomy describes the taxonomy-like structures that emerge when large communities of users collectively tag resources. A folksonomy is a system of classification derived from the practice and the method of collaboratively creating and managing tags to annotate and categorize content.
Tags	A 'tag' is a descriptive keyword or phrase often used to categorize a piece of content. Tag is a metadata and it helps in describing an item and it allows it to be found again by browsing or searching.
Web 2.0	The term "Web 2.0" is commonly associated with web applications that facilitate interactive information sharing , and collaboration on the Web. A Web 2.0 site allows its users to interact with each other as contributors to the website's content, in contrast to websites where users are limited to the passive viewing of information that is provided to them.
Tag Cloud	A tag cloud or word cloud (or weighted list in visual design) is a visual depiction of user-generated tags, or simply the word content of a site. A tag cloud is a box containing a list of tags with the most prominent or popular tags receiving a darker and bigger font than less popular tags.
Tagging	Tagging is the process of assigning personal keywords (“tags”) to resources by users. The related concept folksonomy is the set of labels that emerges from the tagging process.



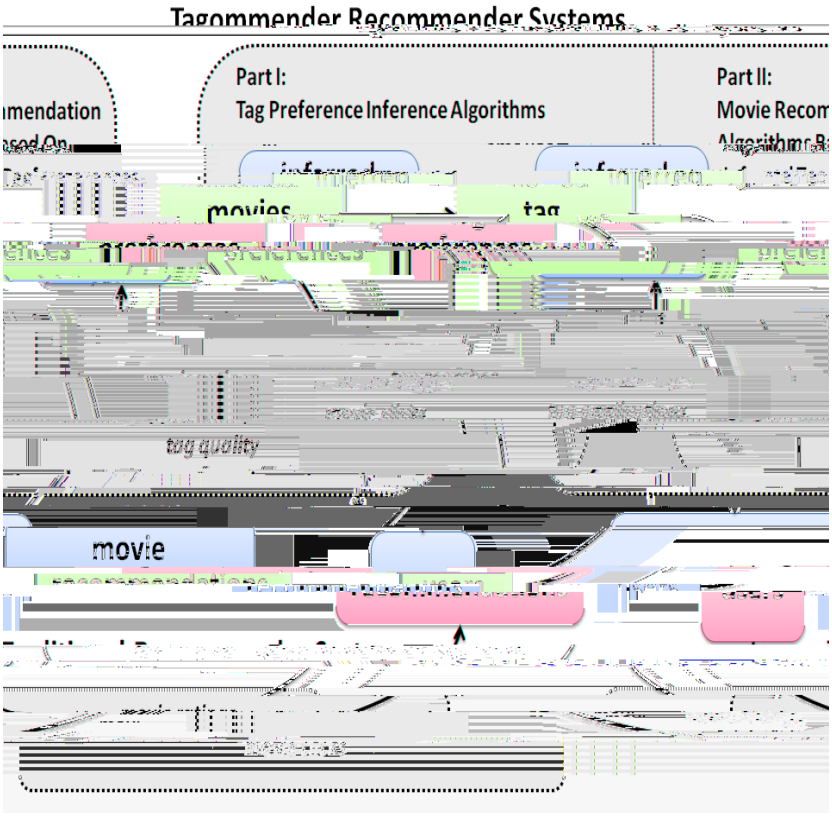
. Model for tagging systems [40].

.

Earlier approaches for prediction in trust aware system make predictions in trust-aware systems utilizing all the trust statements present in the data. The reason explained for the superiority of trust based recommendation over traditional recommendation approach has been attributed to the fact that there is high correlation between trust and user similarity. In [Abdul-Rahman and Hailes 2000], it has been shown that user develop social connections with people that have similar tastes. In [Ziegler and Lausen 2004] they have empirically shown correlation between trusoeen i 169.ql8 TD

2 [illegible]

Collaborative filtering (CF) based tagommenders are those systems which integrate tags information into standard CF systems based on user based and item based algorithms. Tags information integrated into the CF systems can be explicit tagging information or inferred tag preference information. Most of the approaches that are categorized into CF based tagommenders can be explained by the model given in [10]. Figure 2 illustrates the model. In [10] tag preferences were inferred from tag searches, tag applications, tag quality, movie ratings and movie clicks. Combining inferred tag preferences data and ratings data final recommendations were generated. It was shown that the accuracy of the recommender system that integrates hybrid tags preferences data (i.e. tag preference algorithm that combining explicit and implicit preferences data) into CF based recommendation outperforms existing CF algorithms. An approach to incorporate tagging information into a model based CF procedure has been proposed in [11]. Tagging information is used to regularize the matrix factorization step in the framework proposed in [11]. The proposed algorithm was evaluated on the movielens data set and shown to outperform CF algorithms. It is also shown as a solution for the cold start problem i.e. the problem of recommending items to new users. Similarly, in [12] it is shown that social tags information combined with CF based methods can help improve recommendations made to cold start users. In [13] a fusion mechanism is proposed that captures the relations between the three dimensions of users, items and tags by extending the existing user-item correlation matrix with tags data. It applies an algorithm that fuses user based and item based CF algorithms so that the correlations between users, items and tags can be captured simultaneously. [14] also develops a unified framework that tries to model the relationship between the three dimensions users, items and tags. This proposed method results show significant improvement over CF algorithms when evaluated over on real life data sets of Last.fm and BibSonomy.



Model for Collaborative Filtering based tagommenders [10].

2

Profiling based tagommender systems are primarily based on refining the profile of each customer through tags data.

Profiling based tagommenders systems may use ratings data

provided by users. The recommendation process then exploits the profiles learned. An approach to integrate content-based profiles representing long-term user interests with tags based user profiles acquired by capturing users tagging data has been proposed in [19]. In [20], the approach uses the three-dimensional relationship between users, items and tags as used in [13, 14] to profile users and generate most like-minded neighbors or similar items.

The are few more approaches used for recommending items to users based on tags data. In [21] it is shown that hierarchal clustering of tags is an effective means of generating recommendations in collaborative tagging systems. This approach is shown to be more effective in generating recommendations in spare folksonomies as it removes clusters that are not directly related user's context. Similar to approach used in [19], [22] proposes an approach of integrating users tags and expert developed item taxonomy together to make personalized recommendations.

3

Tags are applied by users to items for the purpose of organization and retrieval of items. High quality relevant tags also help in generating better recommendations. Most of the responsibility of tagging lies with the user. Research has shown that even though users realize the benefit of tagging, most people do not tag their resources [23]. It is seen that [24] the perceived benefits of tagging are vaguely defined as a result the user is reluctant to give the required effort. Tag suggestions algorithms help solve the problem by suggesting tags that are relevant thus enabling the user to effectively and efficiently attach tags with items. In [25] the motivations behind tagging and the role of tag suggestion in the systems were studied. The study shows that tag suggestions has a large impact on users tagging behavior as relevant tag suggestions can encourage tagging and provides ideas to users for other possible tags. How social tagging can be enhanced by tags suggestion from a controlled vocabulary in described in [26]. The results show the importance of tag suggestions for vocabulary and retrieval.

Tag co-occurrence is the most popular tag suggestion algorithm. [27,28, 29,30] are some approaches proposed to extract co-occurring tags. In [28] tag-tag correlation network construction is proposed to find co-occurring tags, graph mining approaches like spectral clustering [29] and k-way graph partitioning [30] has also been proposed. For tag suggestion in blogs approaches like VSM, TF-IDF [31] and resource similarity [32] has been proposed. The

algorithms have been proposed and implemented tested on the photo sharing sit Flickr.com data set. In [34, 35] a ranked list of tags is presented to the user based on his tagging activity and tagging activity of other users in the systems. In [34] the proposed algorithm forms group-tags matrix, groups that are formed topic wise are associated with tags by using cosine similarities both on the user and the group profiles and the item for which suggestions has to be made. In [35] a hybrid algorithm is proposed by combining Naïve-Bayes and tf-idf approaches. A tag co-occurrence strategy is proposed in [36]. Tag co-occurrence statistics is first extracted, and then based on two tag aggregation strategies and a promotion function tags are suggested to the user. While approaches explained earlier primarily used the collective knowledge of users present in system i.e. the collaborative tagging activity of the users to generate tag suggestions, [37,38] use geo tags i.e. location data to generate tag suggestions for flickr datasets. In [37] the proposed system spirit tagger suggests tag that captures the spirit of the location. Spirit tagger suggests tags that are popular at that particular location. Geographically representative tags are discovered for locally prevalent tags at a particular geographic location. Zonetag [38] uses information about users own tagging history, tagging activity of other users in the flickr community at that particular location to generate a prioritized suggested tag list. To generate the final list several heuristics considering spatial, social and temporal attributes of the tags are applied. Users social network data is used to suggest tags in [39]. Tagging information from four sources are used to make tag suggestions, tags applied to all photos in the systems, tags applied by the users, tags applied to photos of user's social contacts and tags applied to photos in the groups in which the user is a member. Experimentally they show that relevant recommendations are possible through this approach.

4 [illegible]

Introduction of tag template could be kept as future work, as these templates would minimize the task of the users and would enhance an option for better tag collection and proper classification of tag clouds. Although the focus of the analysis of tag selection methods revolves around three basic types of signals (implicit user behavior, a user's own ratings, aggregate user ratings), more complex techniques may lead to improved accuracy. Future work could explore more sophisticated methods for estimating tag preference. The exploration of more complex algorithms, such as those based on machine learning techniques are left, as future research work. One would also like to validate the techniques using other tagging applications, whether the design principles that is present

generalize to other types of content such as images, articles and bookmarks can be taken as part of the future thoughts. Further research may also show that rating

[illegible]

Š‹• ”‡•‡ƒ”…Š Šƒ• „‡‡› •Š‡’‘”–‡†‡•‡ƒ”…ä Šä’”‘ŒŒ‡r s.{– ã uu{y trrzær{ä

[illegible]

[1] G. Linden, B. Smith, and J. York. Amazon.com recommendations: Item-to-item collaborative filtering. IEEE Internet Computing, 7(1):76–80, 2003.

[2] K. Rose. Digg: Recommendation engine rolling out this week. http://http://blog.digg.com/?p=127, 2008. Retrieved on October 30, 2008

[3] J. Bennett and S. Lanning. The Netflix Prize. In Proceedings of KDD Cup and Workshop, 2007.

[4] Morgan Ames, Mor Naaman, “Why We Tag: Motivations for Annotation in Mobile and Online Media” CHI 2007, April 28–May 3, 2007, San Jose, California, USA.

[14] Panagiotis Symeonidis, Alexandros Nanopoulos, and Yannis Manolopoulos, « A Unified Framework for Providing Recommendations in Social Tagging Systems Based on Ternary Semantic Analysis" IEEE transactions on knowledge and data engineering, vol. 22, no. 2, February 2010.

313.6(stbAlexa)]5 Tc[15]467 0 TD 0.003 Tc 0.029 Tw [(Liang, H., Xu, Y., Li, Y., and Nayak, R.) Popular item v10.

[29] G. Begelman, P. Keller, and F. Smadja. Automated Tag Clustering: Improving search and exploration in the tag space. In WWW '06: Proceedings of the 15th international conference on World Wide Web, 2006..

[30] G. Karypis and V. Kumar. A Fast and High Quality Multilevel Scheme for Partitioning Irregular Graphs. SIAM J. Sci. Comput., 20(1):359–392, 1998..

[31] C. H. Brooks and N. Montanez. An Analysis of the Effectiveness of Tagging in Blogs. In AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs, 2006.

[32] P. Heymann and H. Garcia-Molina. Collaborative Creation of Communal Hierarchical Taxonomies in Social Tagging Systems. Number 2006-10. Department of Computer Science, April 2006..

[33] S. Oldenburg, L. Zielinski, M. Garbe and C. Cap. Comparative analysis of tag suggestion algorithms. 2nd ACM Workshop on Social Network Mining and Analysis, 2008

[34] Nikhil Garg, Ingmar Weber, "Personalized, Interactive Tag Recommendation for Flickr" RecSys'08, October 23–25, 2008, Lausanne, Switzerland.

[35] Nikhil Garg, Ingmar Weber, "Personalized Tag Suggestion for Flickr" WWW 2008, April 21–25, 2008, Beijing, China.

[36] Borkur Sigurbjornsson , Roelof van Zwol, "Flickr Tag Recommendation based on Collective Knowledge" WWW 2008, April 21–25, 2008, Beijing, China

[37] E. Moxley, J. Kleban, and B.S. Manjunath, Spirittagger: a geo-aware tag suggestion tool mined from flickr. In Proc. ACM Multimedia Information Retrieval, pages 24--30, 2008

[38] Mor Naaman and Rahul Nair, "ZoneTag's Collaborative Tag Suggestions: What is This Person Doing in My Phone?" IEEE February 8, 2010