International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 11
July 2016

# Privacy-Preserving Enhanced Collaborative Tagging

[1] AMMINENI NARMADA, [2] RAMANAGOUDA S PATIL

[1] *M.Tech Student, Department of CSE, BIT Institute Of Technology, Hindupur.*
[2] *Professor & HOD, Department of CSE, BIT Institute Of Technology, Hindupur.*

***Abstract*** Collaborative tagging is one of the most popular and dif-fused services available online. The main purpose of collaborative tagging is to loosely classify resources based on end-users feedback, expressed in the form of tags. Con-tent/resource categorization has been seen a challenging research topic in recent year. Tag suppression is a privacy enhancing technique for the semantic Web. In this paper, users are assigned a tag to resources on the Web revealing their personal preferences. However, in order to avoid privacy attackers from profiling users based on their interests, they may wish to refrain from tagging certain resources. Consequently, tag suppression protects user privacy to a certain manner, but at the cost of semantic loss incurred by suppressing tags. In a nutshell, this technique poses a trade-off between privacy and suppression. In this paper, this trade off is investigated in a systematic fashion and provides an extensive theoretical analysis. User privacy is measure as the entropy of the users tag distribution after the suppression of some tags.

**KEYWORDS:** social bookmarking, tag suppression, privacy-enhancing technology, Shannon's entropy, privacy-utility tradeoff.

## INTRODUCTION

Collaborative tagging became popular with the launch of sites like Flickr and Delicious. Since then, different social systems have been built that support tagging of a variety of resources. For a particular web object or resource, tag-ging is a process where a user assigns a tag to an object. A user can assign tags to a particular bookmarked URL on Delicious and on Flickr, users can tag photos uploaded by them or by others. Whereas Delicious allows each user to have her personal set of tags per URL, Flickr has a single set of tags for any photo. On blogging sites like Blogger, Live journal, Word press, blog authors can add tags to their posts. The main purpose of collaborative tagging is to classify re-sources based on user feedback in the form of tags. It is used to annotate any kind of online and offline resources, such as Web pages, images, videos, movies, music, and even blog posts. Nowadays collaborative tagging is mainly used to support tag-based resource browsing and discovery. Consequently, collaborative tagging would re-quire the enforcement of mechanisms that enable users to protect their privacy by allowing them to hide certain user generated contents, without making them useless for the purposes they have been provided in a given online service. This means that privacy preserving mechanisms must not negatively affect the accuracy and effectiveness of the service, e.g., tag-based filtering, browsing, or personalization. Tag suppression is the privacy-enhancing technology (PET) is used to protect privacy of end user. Tag suppression is a technique that has the purpose of pre-venting privacy attackers from profiling users interests on the basis of the tags they assign. It can affect the effectiveness of policy based collaborative tagging systems.

## 2. Literature Survey:

There are numerous approaches for collaborative tagging like data perturbation, tag prediction and tag recommendation.

### 2.1 Data Perturbation:

Collaborative filtering techniques are becoming increasingly popular in E-commerce recommender systems as data filtration is most demanding way to reduce cost of searching in E-commerce application. Such techniques suggest items to users employing similar users preference data. People use recommender systems to deal with information overload.

#### 2.1.1 Randomized Perturbation Techniques:

In this paper, H. Polat and W. Du propose a random-ized perturbation technique to protect individual privacy while still producing accurate recommendations results. Although the randomized perturbation techniques attach randomness to the original data to prevent the data collector from learning the private user data, the method can still provide recommendations with decent accuracy. These approaches basically suggest perturbing the in-formation provided by users. In this, users add random values to their ratings and then submit these perturbed ratings to the recommender system. After receiving these ratings, the system performs an algorithm and sends the users some information that allows them to compute the prediction

#### Advantage:

This approach makes it possible for servers to collect private data from users for collaborative filtering purposes without compromising users privacy requirements. This solution can achieve nearly accurate prediction compared to the prediction based on the original data. The accuracy of this scheme can be provide most ac-curate result if more aggregate information is disclosed along with the concealed data, especially those aggregate information whose disclosure does not compromise much of users privacy. This kind of information includes distribution, mean, standard deviation, true data in a permuted manner.

#### 2.1.2 SVD (Singular Value Decomposition):

In this paper, H. Polat and W. Du proposed SVD Based collaborative filtering technique to preserve rivacy.The method used is a randomized perturbation-based System to protect users privacy while still providing recommendations with decent accuracy. In this, the same perturbative technique is applied to collaborative filtering algorithms based on singular-value decomposition.Even though a user disguises all his/her ratings, but the items themselves may uncover sensitive information. The simple fact of showing interest in a particular item may be more revealing than the ratings assigned to that item.

### 2.2 Tag Prediction:

Tag prediction concerns about the possibility of identifying the most probable tags to be associated with a non tagged resource. Tags are predicted based on resources content and its similarity with already tagged resources.

#### 2.2.1 Social Tag Prediction:

In this paper, D. Ramage, P. Heymann, and H. Garcia-Molina proposed a tag prediction technique. Tag is predicted based on anchor text, page text, surrounding hosts, and other tags applied to the URL. An entropy-based metric which captures the generality of a particular tag and informs an analysis of wellness of the tag which can be predicted. Tag-based association rules can produce very high-precision predictions and giving the deeper under-standing into the relationships between tags. The predictability of a tag when the classifiers are given balanced training data is negatively correlated with its occurrence rate and with its entropy. More popular tags and higher entropy tags are harder to predict. When considering tags in their natural (skewed) distributions, data scar-city issues lead to dominate, so each tag improves classifier

performance. This method performs poor in case of popular tags and distribution becomes poor with overall performance.

### 2.2.2 Granularity of User Modelling:

In this paper, Frias-Martinez, M. Cebrian, and A. Jaimes proposed a tag prediction technique based on granularity. One of the characteristics of tag prediction mechanisms is that, all user models are constructed with the same granularity.In order to increase tag prediction accuracy, the granularity of each user model has to be adapted to the level of usage of each particular user. In this, canonical, stereotypical and individual are the three granularity. levels which are used to improve accuracy. Prediction ac-curacy improves if the level of granularity matches the level of participation of the user in the community.This approach doesn't investigate the following two areas:

1)How to identify the scope of information used in the construction of the models (i.e., size and shape of clusters in the stereotypical case).
2)How and when user models evolve from one granularity to the next.

### 2.3 Recommendation Approach:

In this paper, G. Adomavicius and A. Tuzhilin proposed a tag recommendation approach. It suggests to users the tags to be used to describe resources they are bookmarking. It is enforced by computing tag based user profiles and by suggesting tags specified on a given resource by users having similar characteristics/interest.

### 2.3.1.Content-based Recommendation Approach:

Content-based recommendation systems try to recommend items similar to those a given user has preferred in the past. The basic process performed by a content-based recommender consists in matching up the attributes of a user profile in which preferences and interests are stored, with the attributes of a content object (item), in order to recommend to the user new interesting items.

**a) Heuristic-based:**

In this item profile is searched by using TF-IDF (Term Frequency-Inverse Document Frequency). User profile
(weights of keywords for each user) and cosine similarity are calculated.

**b) Model-based:**

In this Bayesian classifiers and Probability measures are used in content-based approach. Some of the model-based approaches provide rigorous rating estimation methods utilizing various statistical and machine learning techniques.
1.Limited Content Analysis (insufficient set of features).
2.Overspecialization (recommend too similar items).
3.New User Problem (not enough information to build user profile).

### 2.3.2 Collaborative based:

In this, the user is recommended items that people with similar tastes and preferences liked in the past. Collaborative recommender systems (or collaborative filtering systems) try to predict the utility of items for a particular user based on the items previously rated by other users. The utility u(c, s) of item s for user c is calculated based on the utilities u (cj , s) assigned to item s by those users cj € C who are similar to user c.

**a) Heuristic-based:**

In this, correlation coefficient and cosine-based Similarity measurements are used. Heuristic based methods are also known as memory based methods. Memory-based algorithms essentially are heuristics that make rating pre-dictions based on the entire collection of previously rated items by the users.

**b) Model-based:**

In this, Cluster models and Bayesian networks are used. Some of the model-based approaches provide

various rating estimation methods utilizing various statistical and machine learning techniques.

1.New User Problem (not enough information to build user profile).

2.New Item Problem (too few have rated on new items).

3.Sparsity (too few pairs of users have sufficient both-rated items to form a similar group among them).

## 3. Implementation Details:

The architecture consists of privacy and policy layer. The aim of privacy layer is to preserve privacy of end user by applying tag suppression techniques and the aim of policy layer will be to enforce user preferences.

### *3.1. CONTENT TRUST MODELING*

Content trust modelling is used to classify content (e.g., Web pages, images, and videos) as spam or legitimate. In this case, the target of trust is a content (resource), and thus a trust score is given to each content based on its content and/or associated tags. Content trust models reduce the prominence of content likely to be spam, usually in query-based retrieval results. They try to provide better ordering of the results to reduce the exposure of the spam to users. Koutrika et al proposed that each incorrect content found in a system could be simply removed by an administrator. The administrator can go a step further and remove all content contributed by the user who posted the incorrect content, on the assumption that this user is a spammer (polluter).

### 3.2. *USER TRUST MODELING (static)*

The aforementioned studies consider users' reliability as static at a specific moment. However, a user's trust in a social tagging system is dynamic, i.e., it changes over time. The tagging history of a user is better to consider, because a consistent good behavior of a user in the past can suddenly change by a few mistakes, which consequently ruins his/her trust in tagging.

### *3.3USER TRUST MODELING (Dynamic)*

A dynamic trust score, called Social Trust, is derived for each user. It depends on the quality of the relationship with his/her neighbours in a social graph and personalized feedback ratings received from neighbours so that trust scores are updated as the social network evolves. The dynamics of the system is modelled by including the evolution of the user's trust score to incent long-term good behaviour and to penalize users who build up a good trust rating and suddenly "defect." It was shown that Social Trust is resilient to the increase in number of malicious users, since the highly trusted users manage to keep them under control thanks to the trust aware feedback scheme introduced in this approach. It was also shown that Social Trust outperforms Trust Rank-based models, because Social Trust model incorporates relationship quality and feedback ratings into the trust assessment so that bad behaviour is punished.

### *3.4. DATA SET*

Data sets used for development and evaluation of trust modeling techniques have a wide range of diversity in terms of content, numbers of resources, tags and users, and type of spam. Social bookmarking is the most popularly explored domain for trust modeling, especially user trust modeling.

## CONCLUSIONS

In this article, we dealt with one of the key issues in social tagging systems: combatting noise and spam. We classified existing studies in the literature into two categories, i.e., content and user trust modeling. Representative techniques in

each category were analyzed and compared. In addition, existing databases and evaluation protocols were re viewed. An example system was presented to demonstrate how trust modeling can be particularly employed in a popular application of image sharing and geotagging. Finally, open issues and future research trends were prospected. As online social networks and content sharing services evolve rapidly, we believe that the research on enhancing reliability and trustworthiness of such services will become increasingly important.

## REFERENCES

[1]Javier Parra-Arnau, Andrea Perego, Elena Ferrari, Jor-di Forne´, and David Rebollo-Monedero," Privacy-Pre-serving Enhanced Collaborative Tagging", IEEE Transactions on Knowledge and Data Engineering, VOL. 26, NO. 1, January 2014.

[2]H. Polat and W. Du, SVD-Based Collaborative Filtering with Privacy, Proc. ACM Intl Symp. Applied Computing (SASC), pp. 791-795, 2005.

[3]P. Heymann, D. Ramage, and H. Garcia-Molina, Social Tag Prediction, Proc. 31st Ann. Intl ACM SIGIR Conf. Research Development Information Retrieval, pp. 531-538, 2008.

[4]E. Frias-Martinez, M. Cebrian, and A. Jaimes, A Study on the Granularity of User Modelling for Tag Prediction, Proc. IEEE/ WIC/ACM Intl Conf. Web Intelligence Intelligent Agent Technology (WIIAT), pp. 828-831, 2008.

[5]H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar, On the Privacy Preserving Properties of Random Data Perturbation Techniques, Proc. IEEE Intl Conf. Data Mining (ICDM), pp. 99- 106, 2003.

[6]Z. Huang, W. Du, and B. Chen, Deriving Private In-formation from Randomized Data, Proc. ACM SIGMOD Intl Conf. Management Data, pp. 37-48, 2005.

[7]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. Knowledge Data Eng., vol. 17, no. 6, pp. 734- 749, June 2005.

[8]H. Polat and W. Du, Privacy-Preserving Collaborative Filtering Using Randomized Perturbation Techniques, Proc. SIAM Intl Conf. Data Mining (SDM), 2003.