Volume 5, Issue 6, June 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Evaluating the Privacy of User Profiles in Personalized Information Systems

[1]U. Yobu, [2]B. Lalitha
[1]M. Tech (Artificial Intelligence), [2]Assistant Professor
[1, 2] CSE Dept., JNTUA College of Engineering, Anantapuramu, A.P, India

Abstract -Collaborative tagging is one of the most well-known and widespread services available online. The key point of collaborative tagging is to distinguish the resources based on user opinion, stated in the form of tags. Collaborative tagging supplies the source for the semantic Web, network will connect all online resources based on their meanings. While this information is a valued source, its total volume limits its value. Most of the research projects and corporations are discovering the use of personalized applications that control this overflow by modifying the information obtainable to individual users. These applications altogether utilize some information about individuals in directive to be active. This zone is generally called user profiling. In this paper some of the most standard techniques for gathering information about users, signifying, and constructing user profiles. This paper mainly focus on measuring the privacy of user profiles through kl divergence and Shannon entropy techniques showing the tag suppression that protects the end user privacy.

Keywords: Policy-based collaborative tagging, tag annihilation, privacy-enhancing technology, social networking.

I. INTRODUCTION

Recent years have seen the accelerated growth of a rich variety of personalized information systems (PISs) of unparalleled simplicity, whose aim is to help users deal with information overload. Examples of these systems cover personalized Web search and news, resource tagging in the semantic Web and multimedia recommendation systems. The key enabling technology of such systems is personalization, a research area that has received great attention currently and whose objective is to tailor information-exchange functionality to the particular interests of their users. To achieve this functionality, most personalized information systems exploit on, or lend themselves to, the assembly of profiles, either directly confirmed by a user, or concluded from past activity, not only of the user in question, but also from the profiles of users with whom social relationships are known to the information system. Personalized services therefore allow users to deal with the irresistible excess of information, but certainly at the expense of privacy, especially when profiling is showed across several information systems. Besides, the enhancement of these services with data from social networks creates additional opportunities with respect to information distribution but, at the same time, rises the user privacy risks.
The key point of collaborative tagging is to distinguish the resources based on user opinion, stated in the form of tags. In recent years it is observed that the novelty of such an approach in content and resource categorization ,it is considered as an exciting research topic. Additionally collaborative tagging supplies the source for the semantic web, such that the semantic network will connect all online resources based on their meanings[1].

II. RELATED WORK

COLLABORATIVE tagging is one of the most widespread and popular online accommodation available. First it available only in social bookmarking sites. For Example, Dainty (http://ambrosial.com), shovel (http://digg.com) Stumble- Sur (http://stumbleupon.com) -there is currently applicable to any type of web application social, and it is used to comment on any type of online and offline resources (e.g., web pages, pictures, videos, films, music, and even blog mails)[9]
The key point of collaborative tagging is to distinguish the resources based on user opinion, stated in the form of tags. Now a tag-based resource discovery and browsing are supported by collaborative tagging, it can also be exploited for other purposes. Semantic tagging is intimately related to personalization and that many collaborative tagging systems have recently begun to offer personalized services. In these systems, the user's tastes and interests are inferred implicitly, based on the tags they submit [8].
In fig 2.1 .By proposing an architecture that combines two layers on support of increase and private collaborative tagging. Most significantly, the proposed architecture contains of a bookmarking service and two extra services construct on it. The arrangement of these two services allows us to extend the functionality of collaborative tagging systems and parallely, to provide users with a mechanism to maintain their privacy when tagging[1].

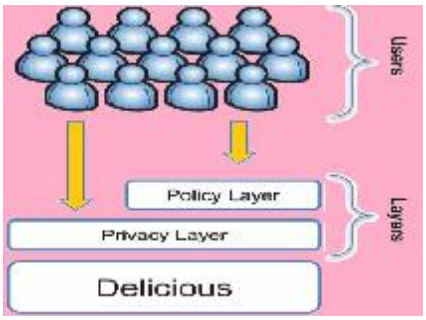

Fig 2.1: Proposed System Architecture**.**

Consequently, collaborative tagging would require the enforcement of mechanisms that allow users to preserve their privacy by permitting them to hide some user produced contents (unless they desire otherwise), without construction them unworkable for the goals that have been provided in an online service given. This means that privacy preserving mechanisms should not adversely affect the accuracy and efficiency of the service, for example, based navigation tags, altering, or customization[5].

We also note that labeling of collaboration is not exploited to its full potential, as it is normally used only to support resource exploration based on tags and search, despite the fact that labeling systems collaboration can be easily improved without modify its core architecture as providing access to information collected through the API, which can be easily exploited by external applications. One reason is that the size of the sets of data collected is too large to allow the enforcement of even simple mechanisms, concerning, e.g., personalization, content filtering and quality assessment. In addition, we commented that current collaborative tagging systems do not enable users to explicitly convey their preferences In fact, the exploitation of explicit relationships and preferences of users has been studied only in where a multi-layer architecture is proposed to integrate basic social services labeled, trust relationships and preferences user. A notable feature of this framework is supported by a rule layer, which can be used to express and enforce user preferences. Such preferences are coded in the policies that explicitly specify the set [3]

III. PROPOSED SCHEME

3.1 Tag Supression:

In our collaborative tagging scenario, users tag Resources on the web, for example, music, images, videos and bookmark's, them depending on their personal preferences. Users therefore help to describe and transfer these resources, but this is inevitable detrimental to reveal their profile. To avoid writing precisely match by tagging systems, or in general by an attacker able to mass data, users can take a technology to reinforce compliance connoted on disruption information[8].

The Data perturbation technology discussed in this section is considered as tag removal technique, a conceptually simple strategy that allows a user to refrain from marking some resources in such a way that the profile resulting from this disturbance does not capture their interests with as much precision. Our approach protects user's privacy to some degree, but at the cost effectiveness of enhanced collaboration marking system. Here we adopted the same model of the opponent and privacy metrics.

The adversary model contemplates the ultimate goal of profiling. Here we assume that the attacker aims to individuate users, that is, its objective is to capture users whose interests deviate from the average profile.

3.2 KL Divergence And Shannon Entropy:

Under this interpretation, the KL divergence is thus interpreted as an (inverse) indicator vulgarity similar profiles in this population. As such, we must hasten to emphasize that the KL divergence is a extent of anonymity instead of intimacy, in the sense that confused information is the individuality of the profile behind the online business, rather than the profile itself. Indeed, an interest profile already corresponding to the public not require disruption. In conclusion, our justification of entropy and divergence as the anonymity of measures is based on these two ideas:

- user-profile density may be regarded as a measure of anonymity.
- The probabilistic model describing the distribution of profiles is frequently unknown to users. Therefore Shannon's entropy and KL divergence as measures of user-profile density.

We model the tags posted by a user as taking values on a common finite alphabet of categories or topics, namely the set {1…….. n} for some integer n >2. In our mathematical model, allows us to describe user profiles by means of the PMF according to which such are distributed, which leads to an equivalent representation than that used in tagging systems. Accordingly, we define q as the probability distribution of the tags of a particular user and σ ∑

[0,1] as a tag suppression rate, which is the ratio of suppressed tags to total tags that the user is willing to eliminate. We define the user's apparent tag distribution We define the distribution of apparent label user t as(q-s/1-σ) for some suppression strategy s = (s1……..sn) satisfying 0 ≤ si ≤ qi and for i =(1…….. n). Conceptually, the distribution of apparent label of the user can be interpreted as the result of a hand, the removal of certain tags from the actual user profile, that is, q - s, and one of the other, the subsequent normalization 1/1 -σ so that we define the delete function of privacy

Max H ((q-s/1-σ)) s
$0 \leq si \leq qi \sum si = \sigma$

3.3 Tag Arrangement:

The representation of a utilizer profile as a tempered histogram across these 59,505 tags would be certainly infeasible from sundry practical positions, mainly concerning the unavailability of data to reliably, exactly evaluate fascinates across such fine-grained arrangement, and, should the data be available, its overwhelming computational complexity. Further, in our experiments but additionally in data mining procedures, a coarser arrangement makes it more facile to have an expeditious overview of the utilizer fascinates.

We determined to categorize the tags in our data set into a coarser representation with just a few high-level tag families. Un conditionally, we used Lloyd's algorithm to group tags into 20 families; and then, for each of those families, we clustered its tags into 10 subcategories. The result of this hierarchical clustering yielded a total of 200 subcategories. In the cessation, we classified the tags in each subcategory in decrementing order of closeness to the center of mass.

Fig 3.3: Under arrangement used in the example policies for parental control scenario.

IV. EXPERIMENTAL RESULTS

In fig 5.1 Module Utilizer share data to his friends with integrating the URL and Tag name and already integrated friends Email Id. The quota data module is subsidiary for Utilizer share their views and fascinates with his friends

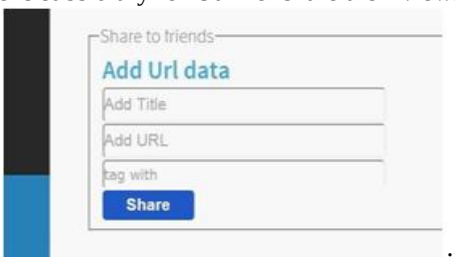

Fig 4.1: Share data to friends.

Fig 4.2: Tag Arrangement Page.

In fig 5.2 Selecting the tag which contains sensitive information that you don't want to share with your friends.

V. EXPERIMENTAL ANALYSIS

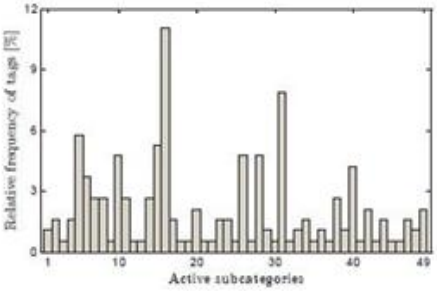

Fig 5.1: σ = 0:00, H(t) ˜3.4834.

fig 6.1 :The figure shows the user's apparent profile just for the active subcategories. For convenience, we rearranged these subcategories and indexed them from 1 to 49. Clearly, when no suppression is applied, the apparent profile is in fact the actual user profile q.

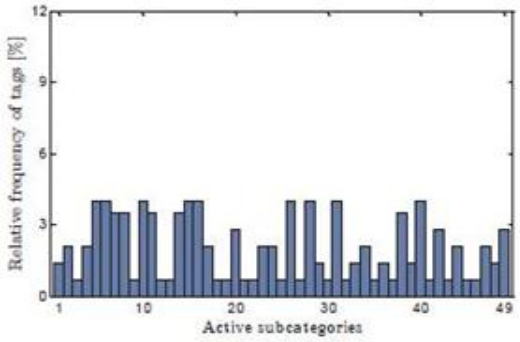

Fig 5.2 : σ = 0:25, H(t) ˜- 3.6863

Fig 6. 2: On the other hand, when σ = 0.25 we observe that the subcategories affected by suppression are those with a percentage of tags furthest away

VI. CONCLUSION

In this paper, the privacy of end utilizer is preserved utilizing tag suppression. The enhanced collaborative tagging architecture is proposed that consists of a bookmarking accommodation and two additional accommodations built on it. The former accommodation enables users to set policies both to block unsought web content and to denote resources of concern. The Tag suppression is a privacy preserving technology predicated on data perturbation. The coalescence of these two accommodations sanctions broadening the functionality of collaborative tagging systems and, concurrently, providing users with a mechanics to maintain their privacy while tagging. Future scope is an extensive performance evaluation of collaborative tagging system architecture, exhibiting its efficacy in terms of privacy measures ,information utility, and filtering capabilities.

REFERENCES

[1] B. Carminati, E. Ferrari, and A. Perego, "Combining social networks and Semantic Web technologies for personalizing Web access," in Collaborative Computing: Networking, Applications and Worksharing, ser. LNICST. Springer, 2009, vol. 10, pp. 126-144.

[2] S. P. Lloyd, "Least squares quantization in PCM," IEEE Trans. Inform. Theory, vol. IT-28, pp. 129-137, Mar. 1982.

[3] E. Michlmayr and S. Cazer, "Learning user profles from tagging data and leveraging them for personal(ized) information access," in Proc. Workshop Tagging and Metadata for Social Inform. Org. Workshop in Int. www. Conf., 2007.

[4] B. Markines, C. Cattuto, F. Menczer, D. Benz, A. Hotho, and G. Stum, "Evaluating similarity measures for emergent semantics of social tagging," in Proc. Int. WWW Conf. ACM, 2009, pp. 641-650.

[5] Z. Yun and F. Boqin, "Tag-Based User Modeling Using Formal Concept Analysis," Proc. IEEE Eighth Int'l Conf. Computer Information Technology (CIT), pp. 485-490, 2008.

[6] R. A.Shepitsen, J.Gemmell, B. Mobasher, and Burke, "Personalized Recommendation in

[7] Social Tagging Systems Using Hierarchical Clustering," Proc. ACM Conf. Recommender Systems (RecSys), pp. 259-266, 2008

[8] M. Bundschus, S. Yu, V. Tresp, A. Rettinger, M. Dejori, and H.-P. Kriegel, "Hierarchical Bayesian Models for Collaborative Tagging Systems," Proc. IEEE Int'l Conf. Data Mining (ICDM), pp. 728-733,2009.

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

[10] C. Marlow, M. Naaman, D. Boyd, and M. Davis, "HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, to Read," Proc. 17th Conf. Hypertext and Hypermedia (HYPERTEXT), pp. 31-40, 2006.

[11] T. Berners-Lee, R. Fielding, and L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax," RFC 3986 (INTERNET STANDARD), Internet Engineering Task Force, Jan. 2005, updated by RFC 6874. [Online]. Available: http://www.ietf.org/rfc/rfc3986.txt

ABOUT AUTHORS

U. YOBU received B.Tech degree in Computer Science and Engineering from sri venkateswara University, College of engineering, A.P, India, during 2009 to 2013. Currently pursuing M.Tech in Computer Science specialization in artificial intelligence from JNTUA College of Engineering, Anantapuramu, A.P, India. His Area of interests include data mining, Network Security.

B. LALITHA is an Assistant Professor of Computer Science and Engineering at Jawaharlal Nehru Technological University College of Engineering, Ananthapuramu. She obtained her Bachelor degree in Computer Science Engineering from Sitams, Chittor, Master of Technology in Computer Science from Jawaharlal Nehru Technological University Anantapur and pursuing Ph.D. in Computer Science and Engineering from Jawaharlal Nehru Technological University Anantapuramu. She has published several Research papers in National International Conferences and Journals. Her research interests include Distributed Computing and Cloud Computing

