R. Yuvarani *et al*, International Journal of Computer Science and Mobile Applications,
Vol.2 Issue. 9, September- 2014, pg. 45-50 **ISSN: 2321-8363**

PROTECTING USER PRIVACY USING OPINION BASED COLLABORATIVE TAG FILTERING

R. Yuvarani[1], P. Narendran[2]

[1]M.Phil Research Scholar, PG & Research Dept. of Comp. Science,
Gobi Arts & Science College (Autonomous), Gobichettipalayam – 638453
yuvishrimugundan@gmail.com

[2]Associate Professor & Head, Department of Comp. Science,
Gobi Arts & Science College (Autonomous), Gobichettipalayam – 638453
narendranp@gmail.com

Abstract

In online co-operative tagging is that the most subtle and widespread services. Most analysis work has investigated the way to effectively reprocess tag collections within the linguistics internet framework and analyzed cooperative tagging practices to enforce methods addressing the linguistics ambiguity issue by statistically analyzing tag collections to infer, whenever doable, a linguistics alignment of a minimum of a set of tags. Tag suppression is employed to preserve the privacy of registered users by concealment the particular characteristics of their profiles. The distinctive purpose of our tag suppression technique is to reinforce user privacy, given a constraint on utility. to supply a services that enables America to broaden the practicality of co-operative tagging systems and, at constant time, offer users with a mechanism to preserve their privacy whereas tagging. and eventually opinion based mostly tagging system and content filtering is projected here, that permits the user to filter the info supported users choice.

***Keywords*:** Tag suppression, Content filtering, Collaborative tagging

---

1. Introduction

The issue of however information for internet resources ought to be generated with the best potency and efficaciousness continues to be a central concern because the quantity of knowledge on the net grows. a little however more and more prestigious set of internet applications, together with the social bookmarking web site delicious, Flickr, Furl, Rojo, Connotea, Technorati, and Amazon permit users to tag objects with keywords to facilitate retrieval each for the user and for alternative users. Sets of classes that are derived supported the tags that are accustomed characterize some resource are normally mentioned as folksonomies [1].

There are each advantages and disadvantages to the tagging approach. Tagging is taken into account a categorization method, in distinction to a preoptimized classification method as exampled by expert-created linguistics net ontologies. Jacob defines the excellence between categorization and classification within the following way: Categorization divides the globe of expertise into teams or classes whose members share some perceptible similarity among a given context. That this context might vary and with it the composition of the class is that the terribly basis for each the edibleness and therefore the power of psychological feature categorization [2]. whereas categorization involves the orderly and systematic assignment of every entity to at least one and just one class among a system of reciprocally exclusive and non-overlapping categories, it mandates consistent application of those principles among the framework of a prescribed ordering of reality.

The rest of this paper is organized as follows. Section 2 explains the proposed methodology. In section 3 is tag suppression. In section 4 describes the user profile model. Section 5 explains the measuring the privacy of user profile. Section 6 explains the experimental results. The last section draws the conclusions and point out future directions of research.

2. Proposed Methodology

Collaborative tagging describes the method by that several users add data within the type of keywords to shared content. Recently, cooperative tagging has full-grown in quality on the net, on sites that permit users to tag bookmarks, pictures and different content. Proponents of cooperative tagging, generally within the weblogging community, typically distinction tagging-based systems from taxonomies. whereas the latter square measure hierarchical and exclusive, the previous square measure non-hierarchical and comprehensive [3]. acquainted taxonomies embody the Linnaean system of classifying living things, the Dewey decimal classification for libraries, and file systems for organizing electronic files.

The problem here isn't only to search out the proper tradeoff between these two issues. In fact, since collaborative tagging is employed to find/browse resources supported the associated tags, suppressing tags might decrease accuracy, and increase the amount of false positives/negatives. Our aim is to verify whether and the way tag suppression may be effectively applied also in an enhanced collaborative tagging service, because the one illustrated during this paper. Next, we first illustrate our tag suppression technique, then we describe the reference scenarios we've got accustomed do our experiments.

3. Tag Suppression

In our state of affairs of cooperative tagging, users tag resources on the net, as an example, music, pictures, videos or bookmarks, in line with their personal preferences. Users thus contribute to explain and classify those resources, however this can be inevitably at the expense of showing their profile. To avoid being accurately profiled by tagging systems or normally by any offender ready to collect such info, users could adopt a privacy-enhancing technology supported knowledge perturbation [4]..

Our conceptually easy technique protects user privacy to a definite degree, however at the value of the linguistics loss incurred by suppressing tags. Different approaches supported knowledge perturbation embody the submission of false tags. as an example, a user wish to tag the webpage web.mentalhelp.net with "depression" might use the tag "sports" instead, to hide their interest for this resource [5]. In doing thus, the user distorts their actual profile, though at the expense of a way larger impact on linguistics practicality than suppression will resources area unit allotted tags that don't describe, in theory, the particular content of such resources.

4. User Profile Model

In the situation of social bookmarking, a user browses the net bookmarks pages and assigns tags to them in line with his/her profile of interests. As in our previous work on tag suppression [6], we tend to contemplate n tag classes, indexed by 1; ...; n, and model the profile of a user as a likelihood Mass perform (PMF)q=(q1,….,qn), that is, a bar chart of relative frequencies of tags across these classes. Our model of user profile is akin to the tag clouds that varied cooperative tagging services use to envision the tags denote by users, a tag cloud could be a visual illustration during which tags are weighted in line with their frequency of use. In equivalence wherever we tend to show associate example of user profile sculptural as a tag cloud [7].

Under this model, a privacy offender, probably the social bookmarking supplier itself, purportedly observes a rattled version of this profile, per a tag suppression strategy, and is unaware or ignores the actual fact that the ascertained user profile, additionally within the style of a bar graph, doesn't mirror the particular profile of interests of the user in question. By refraining from tagging on sure classes, the particular profile of interest's letter of the alphabet is perceived from the surface because the apparent PMFs $= \frac{q-r}{1-\sigma}$. during this expression, σ∈[0,1) denotes the suppression rate, that is, the full fraction of tags a user is willing to eliminate, and r the suppression strategy, that models the particular frequencies of classes suppressed, with $0 \leq r_i \leq q_i$ and ∑ri=σ in a very shell, the apparent profile could also be taken intuitively because the results of the suppression of some tags from the particular profile, and also the posterior normalization by 1/(1-σ) in order that $\frac{1}{1-\sigma}$ so that $\sum_{i=1}^{n} s_i = 1$.

5. Measuring the Privacy of a User Profile

To make the presentation of our privacy criterion suited to a wider audience, next we tend to shall review 2 basic quantities of data theory, particularly Shannon's entropy and Kullback-Leibler (KL) divergence. Recall that the Claude E. Shannon entropy H(s)of a PMF s is outlined as H(s)= $\sum_{i=1}^{n} s_i \ \log_2 s_i$ that's, as a live of the uncertainty of the result of a variant distributed in line with PMF [8].

Recently, these two information-theoretic quantities are used as measures of the privacy of user profiles. during this work, we tend to rule out the employment of divergence as a privacy criterion attributable to the unreasonable assumption that the population's tag distribution p is on the market to users. underneath this assumption, note that each measures of privacy are primarily equivalent. impelled by all this, during this work we tend to live the extent of privacy earned by the apparent profile s as its applied scientist entropy [9].

The use of entropy as a live of privacy, within the widest sense of the term, is by no suggests that new. As a matter of reality, Shannon's add the fifties introduced the construct of equivocation because the conditional entropy of a personal message given Associate in Nursing determined secret writing, later employed in the formulation of the matter of the wiretap channel as a live of confidentiality. newer studies rescue the appropriate pertinence of the construct of entropy as a live of privacy, by proposing to live the degree of obscurity evident by Associate in Nursing aggressor because the entropy of the chance distribution of potential senders of a given message.

6. Experimental Results

The impact that tag suppression could wear Associate in Nursing increased cooperative tagging system supported Delicious. In our experiments, we tend to used the Delicious information set retrieved by the Distributed computing Laboratory (DAILabor). This information set includes those bookmarks and tags marked as public by roughly 950,000 users. The data is organized within the style of triples (username, bookmark, tag), all modeling the action of a user associating a bookmarker with a tag. supported this suppression rate and also the user profile across the n= two hundred subcategories, our approach numerically solves the improvement downside [10].

The results of this improvement could be a suppression strategy r, that is, AN n-tuple containing the proportion of tags that a user ought to eliminate in every subcategory. With this data, our suppression rule yield to pick out that explicit tags ought to be born. As AN example, suppose that a selected user announce the tags “sex,” “erotica,” and “women,” all of them akin to the subcategory “entertainment for adults,” and assume that the suppression strategy r recommends eliminating one in every of those 3 tags. Since the tag “sex” is nearer to the centre of mass than the opposite 2 tags, as shown in Figure. 1, our rule would recommend that the user ought to eliminate this tag.

6.1 False acceptance rate

The false acceptance ratio is a unit used to measure the average number of false acceptances within a biometric security system. It measures and evaluates the efficiency and accuracy of a biometric system by determining the rate at which unauthorized or illegitimate users are verified on a particular system.

$$\text{FAR}(\%) = \frac{\text{Number of false acceptances}}{\text{number of total imposter attempts}}$$

6.2 False Rejection Ratio

The False Rejection Rate is the probability that the system fails to detect a match between the input pattern and a matching template in the database. It measures the percent of valid inputs which are incorrectly rejected and it is calculated by below equation

$$\text{FRR} = \frac{\text{Number of false rejections}}{\text{number of total authentic attempts}}$$

R. Yuvarani *et al*, International Journal of Computer Science and Mobile Applications,
Vol.2 Issue. 9, September- 2014, pg. 45-50 ISSN: 2321-8363

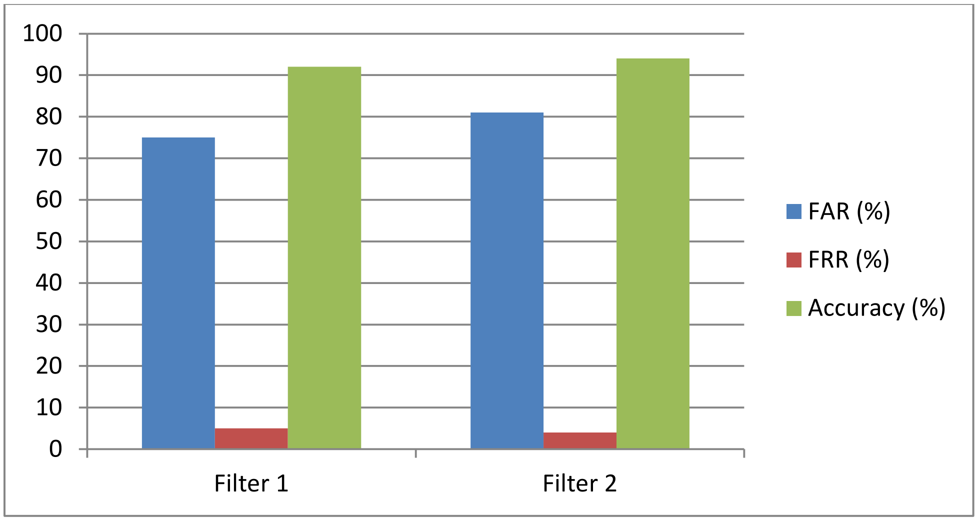

Figure 1: Comparison of FAR, FRR and Accuracy

The above figure shows the FAR, FRR and accuracy of the proposed system, where the error is less in both filter.

6.3 Accuracy

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{(\text{TP}+\text{TN}+\text{FP}+\text{FN})}$$

It is calculated by the above equation, where True positive is correctly identified, False positive is incorrectly identified, True negative is correctly rejected and False negative is incorrectly rejected.

	FAR (%)	FRR (%)	Accuracy (%)
Filter 1	75	5	92
Filter 2	81	4	94

7. Conclusion & Future Work

One of these potential applications is that the provision of internet access functionalities likes content filtering and discovery. The event of a privacy-preserving cooperative tagging service, by showing however a selected privacy-enhancing technology, particularly tag suppression, is accustomed shield end-user privacy. Then opinion primarily based tag filtering technique is introduced that helps to filtrate the tags supported users alternative. Moreover, this may have an effect on the effectiveness of a policy-based cooperative tagging system that supports increased internet access functionalities, like content filtering and discovery, supported preferences such by finish users. Results are evaluated for the filters mistreatment false positive rate, false rejection rate and accuracy.

For on-line Social Networks (OSN) this may be more extended. Here the users will management the messages denote on their own non-public area to avoid unwanted messages displayed and that they can even block their friend from friends list.

References

[1] B. Carminati, E. Ferrari, and A. Perego, 2008, “Combining Social Networks and Semantic Web Technologies for Personalizing Web Access,” *Proc. Fourth Int’l Conf. Collaborative Computing: Networking, Applications and Worksharing*, pp. 126-144.

[2] A. Nanopoulos, July 2011, "Item Recommendation in Collaborative Tagging Systems," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol.41, no.4, pp.760,771.

[3] A. Parra-Arnau, A. Perego, E. Ferrari, J. Forne, D. Rebollo-Monedero , Jan 2014 "Privacy-Preserving Enhanced Collaborative Tagging System," *Knowledge and Data Engineering, IEEE Transactions on* , vol.26, no.1, pp.180,193.

[4] D. Godoy, A. Amandi, Oct 2008, "Hybrid Content and Tag-based Profiles for Recommendation in Collaborative Tagging Systems," *Latin American Web Conference, 2008. LA-WEB '08*., vol., no., pp.58,65, 28-30.

[5] K. Zauder, J.L. Lazic, M.B. Zorica, June 2007, "Collaborative Tagging Supported Knowledge Discovery," Information Technology Interfaces, 2007. ITI 2007. *29th International Conference on* , vol., no., pp.437,442, 25-28.

[6] Sheung-On Choy; AK. Lui, Dec 2006,"Web Information Retrieval in Collaborative Tagging Systems," *Web Intelligence, 2006. WI 2006. IEEE/WIC/ACM International Conference on*, vol., no., pp.352,355, 18-22 .

[7] K.W. Przytula, G.B. Isdale, Tsai-Ching Lu, Sept 2006, "Collaborative Development of Large Bayesian Networks," *Autotestcon, 2006 IEEE* , vol., no., pp.515,522, 18-21.

[8] J. Parra-Arnau, A. Perego, E. Ferrari, J. Forne, D. Rebollo-Monedero, Jan 2014, "Privacy- Preserving Enhanced Collaborative Tagging," *Knowledge and Data Engineering, IEEE Transactions on* , vol.26, no.1, pp.180,193.

[9] R.L. Shuler, A. Balasubramanian, B. Narasimham, B.L. Bhuva, P.M.O. Neill, C. Kouba, Dec 2006, "The Effectiveness of TAG or Guard-Gates in SET Suppression Using Delay and Dual-Rail Configurations at 0.35 μ m," *Nuclear Science, IEEE Transactions on* , vol.53, no.6, pp.3428,3431.

[10] P. Mika, 2005, “Ontologies Are Us: A Unified Model of Social Networks and Semantics,” *Proc. Int’l Semantic Web Conf. (ISWC ’05)*, Y. Gil, E. Motta, V. Benjamins, and M. Musen, eds., pp. 522-536.