Computer Standards & Interfaces 41 (2015) 17–27

Contents lists available at ScienceDirect

Computer Standards & Interfaces

journal homepage: www.elsevier.com/locate/csi

On content-based recommendation and user privacy in social-tagging systems

Silvia Puglisi, Javier Parra-Arnau, Jordi Forné *, David Rebollo-Monedero

Department of Telematics Engineering, Universitat Politècnica de Catalunya (UPC), C. Jordi Girona 1-3, 08034 Barcelona, Spain

ARTICLE INFO

Article history:
Received 30 May 2014
Received in revised form 25 December 2014
Accepted 8 January 2015
Available online 14 February 2015

Keywords:
Information privacy
Privacy-enhancing technology
Privacy risk
Recommendation system
Tag forgery

ABSTRACT

Recommendation systems and content-filtering approaches based on annotations and ratings essentially rely on users expressing their preferences and interests through their actions, in order to provide personalised content. This activity, in which users engage collectively, has been named social tagging, and it is one of the most popular opportunities for users to engage online, and although it has opened new possibilities for application interoperability on the semantic web, it is also posing new privacy threats. In fact, it consists in describing online or offline resources by using free-text labels, i.e., tags, thereby exposing a user's profile and activity to privacy attacks. As a result, users may wish to adopt a privacy-enhancing strategy in order not to reveal their interests completely. Tag forgery is a privacy-enhancing technology consisting in generating tags for categories or resources that do not reflect the user's actual preferences too accurately. By modifying their profile, tag forgery may have a negative impact on the quality of the recommendation system, thus protecting user privacy to a certain extent but at the expenses of utility loss. The impact of tag forgery on content-based recommendation isconsequently investigated in a real-world application scenario where different forgery strategies are evaluated, and the resulting loss in utility is measured and compared.

1. Introduction

Recommendation and information-filtering systems have been developed to predict users' preferences, and to eventually use the resulting predictions for a variety of services, from search engines to resources suggestions and advertisement. The system functionality relies on users implicitly or explicitly revealing their activity and personal preferences, which are ultimately used to generate personalised recommendations.

Such annotation activity has been called *social tagging* and it consists of users collectively assigning keywords (i.e., *tags*) to real life objects and web-based resources that they find interesting. Social tagging is currently one of the most popular online activities. Therefore, different functionalities have been implemented in various online services, such as Twitter, Facebook, YouTube, and Instagram, to encourage their users to tag resources collectively.

Tagging involves classifying resources according to one's experience. Unlike traditional methods where classification happens by choosing labels from a controlled vocabulary, in social tagging systems users freely choose and combine terms. This is usually referred to as free-form tag annotation, and the resulting emergent information organisation has been called *folksonomy*.

This scenario has opened new possibilities for semantic interoperability in web applications. Tags, in fact, allow autonomous agents to categorise web resources easily, obtaining some form of semantic representation of their content. However, annotating online resources poses potential privacy risks, since users reveal their preferences, interests and activities. They may then wish to adopt privacy-enhancing strategies, masquerading their real interests to a certain extent, by applying tags to categories or resources that do not reflect their actual preferences. Specifically, *Tag forgery* is a privacy-enhancing technology (PET) designed to protect user privacy, by creating bogus tags in order to disguise real user's interests. As a perturbation-based mechanism, tag forgery poses an inherent trade-off between privacy and usability. Users are able to obtain a high level of protection by increasing their forgery activity, but this can substantially affect the quality of the recommendation.

The primary goal of this work is to investigate the effects of tag forgery to content-based recommendation in a real-world application scenario, studying the interplay between the degree of privacy and the potential degradation of the quality of the recommendation. An experimental evaluation is performed on a dataset extracted from Delicious [1], a social bookmarking platform for web resources. In particular, three different tag forgery strategies have been evaluated, namely: *optimised tag forgery* [2], *uniform tag forgery* and *TrackMeNot* (TMN) [3], the last consists of simulating a possible TMN like agent, periodically issuing randomised tags according to popular categories.

Using the dataset and a measure of utility for the recommendation system, a threefold experiment is conducted to evaluate how the

* Corresponding author. Tel.: +34 93 401 1871.
E-mail addresses: silvia.puglisi@upc.edu (S. Puglisi), javier.parra@entel.upc.edu (J. Parra-Arnau), jforne@entel.upc.edu (J. Forné), david.rebollo@entel.upc.edu (D. Rebollo-Monedero).

http://dx.doi.org/10.1016/j.csi.2015.01.004
0920-5489/© 2015 Elsevier B.V. All rights reserved.

application of tag forgery may affect the quality of the recommender. Hence, we simulate a scenario in which users only apply one of the different tag forgery strategies considered. Measures of the recommender performances are computed before and after the application of each PET, obtaining an experimental study of the compromise between privacy and utility.

To the best of our knowledge, this is the first systematic evaluation of the impact of applying perturbation-based privacy technologies on the usability of content-based recommendation systems. For this evaluation, both suitable privacy and usability metrics are required. In particular, as suggested by Parra et al. [4], the KL divergence is used as privacy metric of the user profile; while the quality of the recommendation is computed following the methodology proposed by Cantador el al. [5].

This paper is organised as follows: Section 2 introduces the state of the art. Section 3 describes the adversary model considered. Section 4 explains a possible practical application of the proposed PET through the implementation of a communication module. Section 5 discusses the evaluation methodology and obtained results. Section 6 presents the conclusions that can be derived from the presented results, while also introducing future research lines.

2. State of the art

In recommendation systems employing tags or in any system allowing resource annotation, users decide to disclose personal data in order to receive, in exchange, a certain benefit. This earned value can be quantified in terms of the customised experience of a certain product [6]. For such a recommendation system to work, and successfully propose items of interest, user preferences need to be revealed and made accessible partially or in full, and thus exposed to possible privacy attacks.

When a user expresses and shares their interests by annotating a set of items, these resources and their categorisation will be part of their activity. The recorded users' activities will allow the used platform to "know more" about each of them, and therefore suggesting over time useful resources. These could be items similar to others tagged in the past, or simply close to the set of preferences expressed in their profile. In order to protect their privacy, a user could refrain from expressing their preferences altogether. While in this case an attacker would not be able to build a profile of the user in question, it would also become impossible for the service provider to deliver a personalised experience: the user would then achieve the maximum level of privacy protection, but also the worst level of utility.

Various and numerous approaches have been proposed to protect user privacy by also preserving the recommendation utility in the context of social tagging platform. These approaches can be grouped around four main strategies [7]: encryption-based methods, approaches based on trusted third parties (TTPs), collaborative mechanisms and data-perturbative techniques. In traditional approaches to privacy, users or application designers decide whether certain sensitive information is to be disclosed or not. While the unavailability of this data, traditionally attained by means of access control or encryption, produces the highest level of privacy, it would also limit access to particular content or functionalities. This would be the case of a user freely annotating items on a social tagging platform. By adopting traditional PETs, the profile of this user could be made available only to the service providers, but kept completely or partially hidden from their network of social connections on the platform. This approach would indeed limit the chances of an attacker profiling the user, but would, unfortunately, prevent them from receiving content suggested by their community.

A conceptually simple approach to protecting user privacy consists in a TTP acting as an intermediary or *anonymiser* between the user and an untrusted information system. In this scenario, the system cannot know the user ID, but merely the identity of the TTP involved in the communication. Alternatively, the TTP may act as a *pseudonymiser* by supplying a pseudonym ID′ to the service provider, but only the TTP knows the correspondence between the pseudonym ID′ and the actual user ID. In online social networks, the use of either approach would not be entirely feasible as users of these networks are required to authenticate to login. Although the adoption of TTPs in the manner described must, therefore, be ruled out, the users could provide a pseudonym at the sign-up process. In this regard, some sites have started offering social-networking services where users are not required to reveal their real identifiers. Social Number [8] is an example of such networks, where users must choose a unique number as their ID.

Unfortunately, none of these approaches effectively prevents an attacker from profiling a user based on the annotated items content, and ultimately inferring their real identity. This could be accomplished in the case of a user posting related content across different platforms, making them vulnerable to techniques based on the ideas of reidentification. As an example, suppose that an observer has access to certain behavioural patterns of online activity associated with a user, who occasionally discloses their ID, possibly during interactions not involving sensitive data. The same user could attempt to hide under a pseudonym ID′ to exchange information of confidential nature. Nevertheless, if the user exhibited similar behavioural patterns, the unlinkability between ID and ID′ could be compromised through the exploitable similarity between these patterns. In this case, any past profiling inferences carried out by the pseudonym ID′ would be linked to the actual user ID.

A particularly rich group of PETs resort to users collaborating to protect their privacy. One of the most popular is *Crowds* [9], which assumes that a set of users wanting to browse the Web may collaborate to submit their requests. Precisely, a user wishing to send a request to a Web server selects first a member of the group at random, and then forwards the request to them. When this member receives the request, it flips a biased coin to determine whether to forward this request to another member or to submit it directly to the Web server. This process is repeated until the request is finally relayed to the intended destination. As a result of this probabilistic protocol, the Web server and any of the members forwarding the request cannot ascertain the identity of the actual sender, that is, the member who initiated the request.

We consider collaborative protocols [10–12] like Crowds, not suitable for the application addressed in this work although they may be effective in applications such as information retrieval and Web search. The main reason is that users are required to be logged into online social tagging platforms. That is, users participating in a collaborative protocol would need the credentials of their peers to log in, and post on their behalf, which in practice would be unacceptable. Besides, even if users were willing to share their credentials, this would not entirely avoid profiling based on the observation of the resources annotated.

In the case of perturbative methods for recommendation systems, [13] proposes that users add random values to their ratings and then submit these perturbed ratings to the recommender. When the system has received these ratings, it executes an algorithm and sends the users some information that allows them to compute the final prediction themselves. When the number of participating users is sufficiently large, the authors find that user privacy is protected to some degree, and the system reaches an acceptable level of accuracy. However, even though a user may disguise all their ratings, merely showing interest in an individual item may be just as revealing as the score assigned to that item. For instance, a user rating a book called "How to Overcome Depression" indicates a clear interest in depression, regardless of the score assigned to this book. Apart from this critique, other works [14,15] stress that the use of certain *randomised* data-distortion techniques might not be able to preserve privacy completely in the long run.

In line with these two latter works, [16] applies the same perturbative technique to collaborative filtering algorithms based on singular-value decomposition, focusing on the impact that their technique has on privacy. For this purpose, they use the privacy metric proposed by Agrawal, and Aggarwal, [17], effectively a normalised version of the mutual information between the original and the perturbed