JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

SEQURITY INFERENCE ON USER UPLOADED IMAGES OVER SOCIAL NETWORK

[1] **Supriya Jadhav**
[1]ME Student
[1,] Computer Department,
[1]CSMSS College of Engineering, Aurangabad

Abstract

In recent years online social networking communities have undergone massive explosion. The number of sites as well as kinds of sites have grown and it allows us to communicate with a lot of people across the world. Social networking sites such as Facebook , Flickr, MySpace and LinkedIn, give opportunities to share large amount of personal information. People upload their photos to these sites to gain public attention for social purposes, and thus many public consumer photographs are available online. The proliferation of personal data leads to privacy violation .Risks such as identify theft, embarrassment, and blackmail are faced by user's .In order to overcome these risks flexible privacy mechanisms need to be considered. An Adaptive Privacy Policy Prediction (A3P) system helps users to compose privacy settings for their images. A two-level framework which according to the user's available history on the site, determines the best available privacy policy for the user's images being uploaded. A3P system aims to provide users a hassle free privacy settings experience by automatically generating personalized policies. The A3P system provides a comprehensive framework to infer privacy preferences based on the information available for a given user. .When meta data information is unavailable it is difficult to generate accurate privacy policy. Privacy violation as well as inaccurate classification will be the after effect of manual creation of meta data log informatio .

1. Introduction

Social media is the two way communication in Web 2.0 and it means to communicate, share, and interact with an individual or with a large audience. Social networking websites are the most famous websites on the Internet and millions of people use them every day to engage and connect with other people. Twitter, Facebook, LinkedIn and Google Plus seems to be the most popular Social networking websites on the Internet. Today, for every single piece of content shared on sites like Facebook—every wall post, photo, status update, and video—the up loader must decide which of his friends, group members, and other Facebook users should be able to access the content. As a result, the issue of privacy on sites like Facebook has received significant attention in both the research community [1] and the mainstream media [2]. Our goal is to improve the set of privacy controls and defaults, but we are limited by the fact that there has been no in-depth study of users' privacy settings on sites like Facebook. While significant privacy violations and mismatched user expectations are likely to exist, the extent to which such privacy violations occur has yet to be quantified. Images are now one of the key enablers of users' connectivity. Sharing takes place both among previously established groups of known people or social circles (e.g. Google+, Flickr or Picasa), and also increasingly with people outside the users social circles, for purposes of social discovery to help them identify new peers and learn about peers interests and social surroundings. With the increasing volume of images users share through social sites, maintaining privacy has become a major problem, as demonstrated by a recent wave of publicized incidents where users inadvertently shared personal information. In light of these incidents, the need of tools to help users control access to their shared content is apparent. An image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images. Most traditional and common methods of image retrieval utilize some method of adding metadata such as captioning, keywords or descriptions to the image retrieval can be performed over the annotation words. Manual image annotation is time consuming, laborious and expensive to address this, there has been a large amount of research done on automatic image annotation. Additionally, the increase social web applications and the semantic web have inspired the development of several web-based image annotation tools. Automatic image annotation [6] is the process by which a computer system automatically assigns metadata in the form of captioning or keywords to a digital image. This application of computer vision techniques is used in image retrieval systems to organize and locate images of interest from a database. This method can be regarded as a type of multi-image classification with a very large number of classes large as the vocabulary size. Typically, image analysis in the form of extracted feature vectors and training annotation words are used by machine learning techniques to attempt to automatically apply annotations to new images.

2. Related Work

Privacy Suites [1] is proposed by Jonathan Anderson which allows users to easily choose ―suites" of privacy settings. Using privacy programming a privacy suite can be created by an expert. Privacy Suites could also be created directly through existing configuration UIs or exporting them to the abstract format. To the members of the social sites the privacy suite is distributed through existing distribution channels. Transparency is the main goal, which is essential for convincing influential users that it is safe to use. The disadvantage of a rich programming language is less understandability for end users. To verify a Privacy Suite sufficiently high-level language and good coding practice, motivated users are able. Privacy-Aware Image Classification and Search [2] is a technique to automatically detect private images, and to enable privacy-oriented image search introduced by Sergej Zerr. To provide security policies technique combines textual meta data images with variety of visual features. It uses various classification models trained on a large scale dataset with privacy assignments obtained through a social annotation game. In this the selected image features (edges, faces, color histograms) which can help discriminate between natural and man-made objects/scenes (the EDCV feature) that can indicate the presence or absence of particular objects (SIFT). A tag based access control of data [3] is developed by Peter F. Klemperer. It is a system that creates access-control policies from photo management tags. Every photo is incorporated with an access grid for mapping the photo with the participant's friends. A suitable preference can be selected by participants and access the information. Based on the user needs photo tags can be categorized as organizational or communicative. There are several important limitations .First, our results are limited by the participants recruited and the photos provided by them. Machine generated access-control rules are the second limitation. Algorithm used here has no access to the context and meaning of tags and no insight into the policy the participant intended when tagging for access control. Hence, some rules appeared strange to the participants who makes them to tag explicitly like ―private‖ and ―public A decentralised authentication protocol [4], is a access control system proposed by Ching-man Au Yeung based on a descriptive tags and linked data of social networks in the Semantic websites. Here users can specify access control rules based on open linked data provided by other parties and it allows users to create expressive policies for their photos stored in one or more photo sharing.

Adaptive Privacy Policy Prediction (A3P) [5] system is introduced by Anna Cinzia Squicciarini. Personalized policies can be automatically generated by this system. It makes use of the uploaded images by users and a hierarchical image classification is done. Images content and metadata is handled by the A3P system .It consists of two components: A3P Core and A3P Social. The image will be first sent to the A3P-core, when the user uploads the image. The A3P-core classifies the image and determines whether there is a need to invoke the IPPS. When meta data information is unavailable it is difficult to generate accurate privacy policy. This is the disadvantage of this system. Privacy violation as well as inaccurate classification will be the after effect of manual creation of meta data log information.

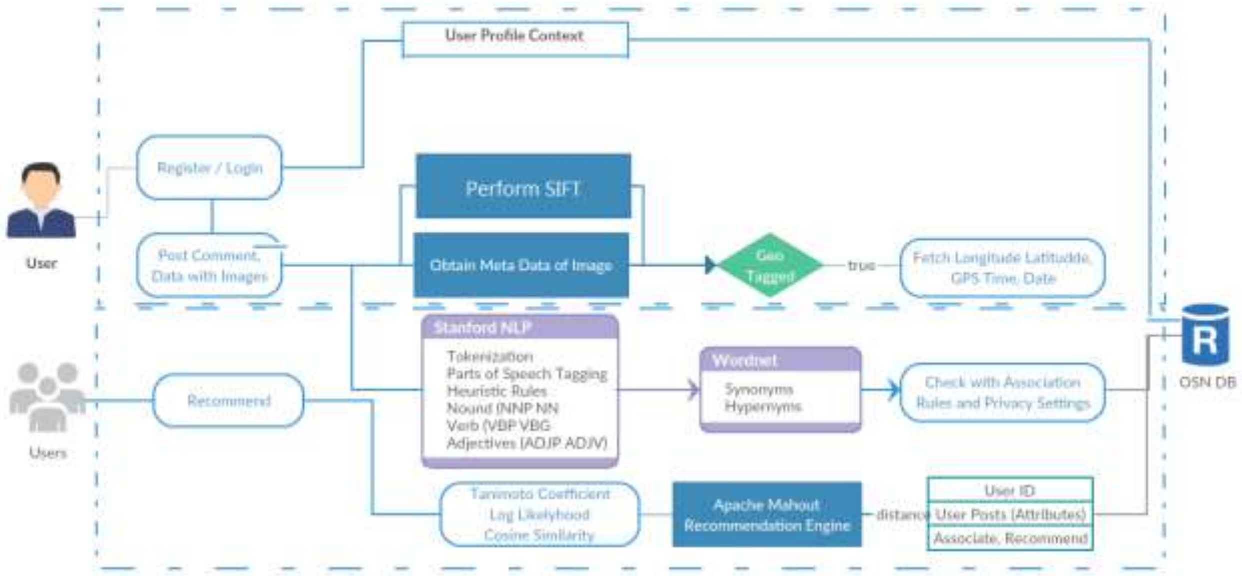
Consider social context such as one’s friend list. While interesting, they may not be sufficient to address challenges brought by image files for which privacy may vary substantially not just because of social context but also due to the actual image content. As far as images, authors in have presented an expressive language for images uploaded in social sites. This work is complementary to ours as we do not deal with policy expressiveness, but rely on common forms policy specification for our predictive algorithm. In addition, there is a large body of work on image content analysis, for classification and interpretation, retrieval, and photo ranking, also in the context of online photo sharing sites. Of these works, probably the closest to ours. explores privacy-aware image classification using a mixed set of features, both content and meta-data. This is however a binary classification (private versus public), so the classification task is very different than ours. Also, the authors do not deal with the issue of cold-start problem.

Most content sharing websites allow users to enter their privacy preferences. Unfortunately, recent studies have shown that users struggle to set up and maintain such privacy settings. One of the main reasons provided is that given the amount of shared information this process can be tedious and error-prone. Therefore, many have acknowledged the need of policy recommendation systems which can assist users to easily and properly configure privacy settings. However, existing proposals for automating privacy settings appear to be inadequate to address the unique privacy needs of images, due to the amount of information implicitly carried within images, and their relationship with the online environment wherein they are exposed

3. Proposed System

We intend to create a adaptive Privacy Policy Prediction (A3P) system that aims to produce users a trouble free privacy settings expertise by mechanically generating customized policies. The A3P system handles user uploaded pictures, and factors within the following criteria that influence one’s privacy settings of images: The impact of social surroundings and private characteristics. Social context of users, like their profile data and relationships with others could offer helpful data relating to users’ privacy preferences the role of image’s content and data. In general, similar pictures typically incur similar privacy preferences, particularly once individuals seem within the pictures. as an example, one could transfer many photos of his children and specify that solely his relations square measure allowed to check these photos

In proposed System an Adaptive Privacy Policy Prediction (A3P) system that helps users automate the privacy policy settings for their uploaded images. The A3P system provides a comprehensive framework to infer privacy preferences based on the information available for a given user. We also effectively tackled the issue of cold-start, leveraging social context information. Our experimental study proves that our A3P is a practical tool that offers significant improvements over current approaches to privacy.



A3P-CORE**:** There are two major components in A3P-core: (i) Image classification and (ii) Adaptive policy prediction. For each user, his/her images are first classified based on content and metadata. Then, privacy policies of each category of images are analyzed for the policy prediction. Adopting a two-stage approach is more suitable for policy recommendation than applying the common one-stage data mining approaches to mine both image features and policies together Image classification: Groups of images that may be associated with similar privacy preferences; we propose a hierarchical image classification which classifies images first based on their contents and then refine each category into subcategories based on their metadata. Images that do not have metadata will be grouped only by content. Such a hierarchical classification gives a higher priority to image content and minimizes the influence of missing tags. Note that it is possible that some images are included in multiple categories as long as they contain the typical content features or metadata of those categories. Adaptive policy prediction: The policy prediction algorithm provides a predicted policy of a newly uploaded image to the user for his/her reference. More importantly, the predicted policy will reflect the possible changes of a user's privacy concerns. The prediction process consists of three main phases: (i) policy normalization; (ii) policy mining; and (iii) policy prediction. 1) Policy normalization: The policy normalization is a simple decomposition process to convert a user policy into a set of atomic rules in which the data (D) component is a single-element set. 2) Policy mining: hierarchical mining first look for popular subjects defined by the user, then look for popular actions in the policies containing the popular subjects, and finally for popular conditions in the policies containing both popular subjects and conditions. 3) Policy prediction: The policy mining phase may generate several candidate policies while the goal of our system is to return the most promising one to the user. Thus, we present an approach to choose the best candidate policy that follows the user's privacy tendency. To model the user's privacy tendency, we define a notion of strictness level. The strictness level is a quantitative metric that describes how "strict" a policy is.

Definition: A Privacy policy „Pp" can be described for user „Us‟ by Subject(S) : A Set of users socially connected to user Us. Data (D) : A set of data items shared by Us. Action (Ac) : A set of actions granted by Us to S on D. Condition (Co) : A Boolean expression which must be satisfied in order to perform the appointed actions.

In the above definition, Subject(S) can be socially connected people on websites like , relations such as family, friend, coworkers, etc. and organizations. Data (D) is the collection of image uploaded by user till date. Action (Ac) consists of four factors: View, Comment, Tags and Share. Condition (Co) specifies whether the actions are effective or not.

A3P (Adaptive Privacy Policy Prediction) may be a framework used for outlining new privacy preferences policies for users and to form the expertise versatile and secure at the time. The A3P design consists of followings blocks:

The A3P-core focuses on analyzing each individual user’s own images and metadata, while the IPPS offers a community perspective of privacy setting recommendations. ¬ Design the interaction flows between the two building blocks to balance the benefits from meeting personal characteristics and obtaining community advice.

Meta-based Image classification: The metadatabased classification groups images into subcategories under aforementioned baseline categories. The process consists of three main steps. The first step is to extract keywords from the metadata associated with an image. The meta-data considered in our work are tags, captions, and comments ,this tags are compared with the already uploaded images.

SIFT Based Content Extraction

Content-based Image classification: Approach to content-based classification is based on an efficient and yet accurate image similarity approach. Specifically, our classication algorithm compares image signatures defined based on quantified and sanitized version of Haar wavelet transformation. For each image, the wavelet transform encodes frequency and spatial information related to image color, size, invariant transform, shape, texture, symmetry, etc. Then, a small number of coefficients are selected to form the signature of the image. The content similarity among images is then determined by the distance among their image signatures. SIFT Algorithm is used to extract the features of image.

Metadata Based Extraction

Metadata Based Image Classification As mentioned, the metadata based Image classification are divided into sub-categories with the help of following three steps.

Step 1 : of this method permits to extract keywords from the information of the image. Tags, Comments and Captions area unit the categories of information through that the keywords area unit obtained. After the keywords area unit obtained, our task is to spot completely different properties like nouns, verbs and adjectives and store them into a information vector like Tn=, Tv=, Ta= wherever k, j and l area unit the whole range of nouns, verbs and adjectives severally.

Step 2 : of this method is to possess an identical word from every vector. The word is denoted by „h‟ and 1st retrieved for each „ti‟. This word is delineate as “h=”.Here „v‟ area unit hypernyms and „f‟ is for frequency. as an example, think about a information vector T=.By this set {we can|we will|we area unit able to} learn that Joband Promotion are with same word „work‟ however Party encompasses a hypernym „Activity‟. Hence, this show the hypernm list as h=. This list tend to choose the word with the most frequency.

Step 3: of this method is to indicate and learn the subcategory within which the image fits in. The progressive procedure within which the primary image forms a subcategory and therefore the hypernyms of the image are assigned to their individual subcategory. The closeness between these hypernyms and every class is computed to outline a subcategory for that image.

Algorithm

Flow of Image Uploading System
1. START
2. Select an image to upload by the login user.
3. Enter appropriate title for the selected image.
4. Process to upload the image in the system.
5. Call method to get image id which is having most similar heading and suitable names. (Algorithm of Privacy Policy)
6. Get privacy policies already set for the result image unique identity.
7. Shows policies to user.
8. If user is satisfied with policies then continue to upload image.
9. If user is not satisfied with policies then allow user to set privacy policy for the image and continue to upload.
10. STOP.

4. Performance Analysis

This section covers the experimental setup of the project and the different experiment carried out for the proper results.

Experimental setup:

Proposed system is developed under Windows 8 Operating System the hardware we used is LENOVO workstation, with an Intel core (TM) 2.50 GHz CPU and 4GB RAM. NetBeans IDE 8.2 environment is used for development with MySQL database.

Data collection:

The dataset consist of different types of images. Some images are real time images with caption , some images with location and caption, and some of them are taken from the social networking sites.

From these data collection we conduct different experiments. When we conduct experiment we also asked for participants if they have had concerns about their privacy due to shared images. Over 51 percent of the participants indicated that they had privacy concerns.

Users also reported that image content is an important factor when determining privacy settings for an image with 87 percent of people agreeing or strongly agreeing with the statement

“When I set privacy settings for a certain image I usually think about the content of the image”, and over 91 percent of users agreeing or strongly agreeing with the statement “The content of an image determines whether I upload the image to a social network site.” Surprisingly, however, many users indicated that they never changed privacy settings for images (38 percent) or changed their settings only 1 or 2 times (36 percent) since joining the social network. There seems to be a clear disconnect between users privacy inclinations and their practice of setting privacy policies.

The possible reason could be “Changing privacy settings for every image uploaded on a social site can be very time consuming”, as strong agreed or agreed by 70 percent of users.

Experimental Results:

Method	**Overall Accuracy**
A3P Core	92.53%
Propagation	66.84%
Tag-Only	87.01%
Proposed Hybrid Method	94.82%

Comparison of Overall Accuracy

We have calculated Precision, Recall, Accuracy and F-measure for 50 posts which is demonstrated on following fig .

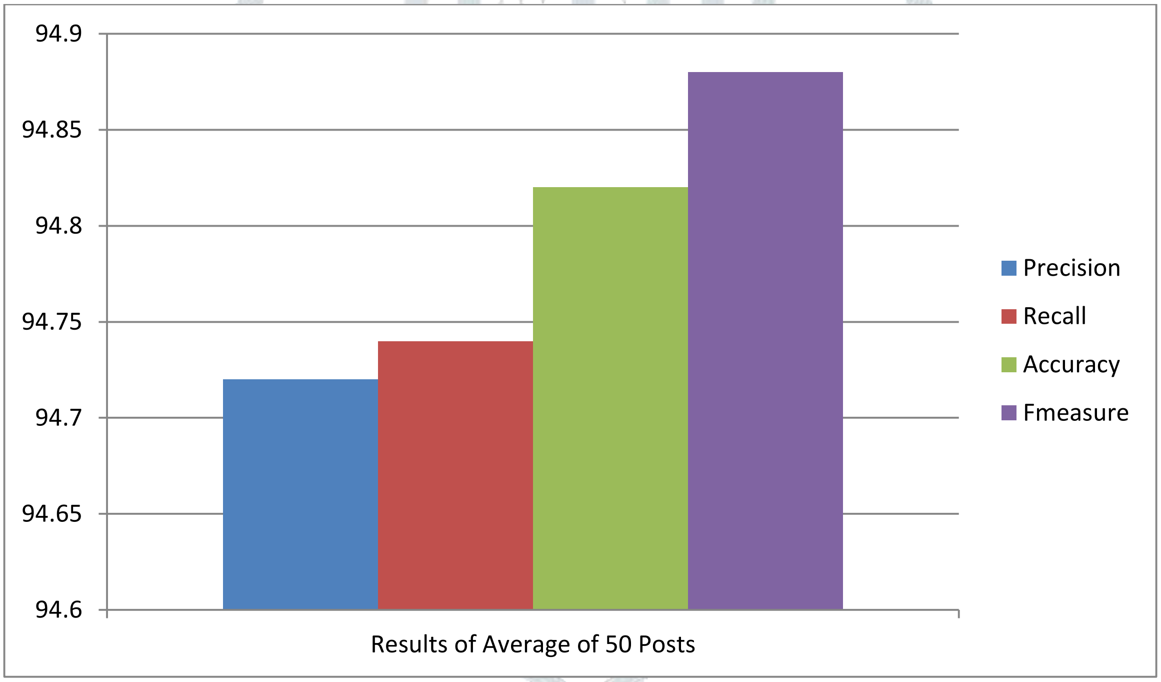


Fig: Precision, Recall, Accuracy and F-measure results of Average of 50 Posts

5. Conclusion

1. We have effectively made use of Association among the user's subjects like family, friend etc. to improve the privacy prediction performance in image-uploads or posts in social media websites

2. By considering textual information and modeling it with NER has improved the performance, which text information can be used as feature to make accurate privacy predictions.

3. In our proposed method, Image classification with geo-tagging information give better results compared with only image classification approach.

4. We have addressed the problem of privacy prediction in social media websites, where we are able to recommend user his/her privacy preference.

REFERENCES

1. M. D. Choudhury, H. Sundaram, Y.-R. Lin, A. John, and D. D. Seligmann, “Connecting content to community in social media via image content, user tags and user communication,” in Proc. IEEE Int. Conf. Multimedia Expo, 2009, pp.1238–1241. Connecting content to community in social media via image content, user tags and user communication
2. A. Acquisti and R. Gross, “Imagined communities: Awareness, information sharing, and privacy on the facebook,” in Proc. 6th Int. Conf. Privacy Enhancing Technol. Workshop, 2006, pp. 36–58.
3. R. Agrawal and R. Srikant,“Fast algorithms for mining association rules in large databases,” in Proc. 20th Int. Conf. Very Large DataBases, 1994, pp. 487–499.
4. J. Bonneau, J. Anderson, and L. Church, “Privacy suites: Shared privacy for social networks,” in Proc. Symp. Usable Privacy Security, 2009.
5. A. Besmer and H. Lipford, “Tagged photos: Concerns, perceptions, and protections,” in Proc. 27th Int. Conf. Extended Abstracts Human Factors Comput. Syst., 2009, pp. 4585–4590.
6. M. Ames and M. Naaman, “Why we tag: Motivations for annotation in mobile and online media,” in Proc. Conf. Human Factors Comput. Syst., 2007, pp. 971–980.
7. Y. Liu, K. P. Gumm
8. adi, B. Krishnamurthy, and A. Mislove. (2011). “Analyzing Facebook privacy settings: user expectations vs. reality,” in Proc. IMC.
9. J. Bonneau, J. Anderson, and G. Danezis, (2009). “Prying data out of a social network,” in Proc. Int. Conf. Adv. Soc. Netw. Anal. Mining, pp.249–254.
10. D. G. Lowe, (2004, Nov.). Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. [Online]. 60(2), pp. 91–110. Available: http://dx.doi.org/10.1023/B: VISI.0000029664.99615.94
11. G. Loy and A. Zelinsky, “Fast radial symmetry for detecting points of interest,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 8, pp. 959–973, Aug. 2003.
12. K. Strater and H. Lipford, “Strategies and struggles with privacy in an online social networking community,” in Proc. Brit. Comput. Soc. Conf. Human-Comput. Interact., 2008, pp.111–119.
13. N. Zheng, Q. Li, S. Liao, and L. Zhang, “Which photo groups should I choose? A comparative study of recommendation algorithms in flickr,” J. Inform. Sci., vol. 36, pp. 733–750, Dec. 2010.
14. A. Kapadia, F. Adu-Oppong, C. K. Gardiner, and P. P. Tsang, “Social circles: Tackling privacy in social networks,” in Proc. Symp. Usable Privacy Security, 2008.
15. H. Lipford, A. Besmer, and J. Watson, “Understanding privacy settings in facebook with an audience view,” in Proc. Conf. Usability, Psychol., Security, 2008.
16. A Survey on Deep Learning for Named Entity Recognition Jing Li, Aixin Sun, Jianglei Han, Chenliang Li Submitted on 22 Dec 2018 (v1), last revised 18 Mar 2020 (this version, v3)]

17. A Survey on Recent Advances in Named Entity Recognition from Deep Learning models Vikas Yadav, Steven Bethard
18. Named Entity Recognition: A Literature Survey Rahul Sharnagat 11305R013 June 30, 2014
19. Techniques for Named Entity Recognition: A Survey Girish Keshav Palshikar Tata Research Development and Design Centre, India
20. A Brief Survey on Name Entity Recognition in Natural Language Processing For Indian Languages N.Vasunthira Devi Ph.D., Research Scholar in Computer Science, Mother Teresa Women's University, Kodaikanal, India vasunthira@gmail.com Dr.R.Ponnusamy Professor, Department of Computer Science & Engineering, Sri Lakshmi Ammal Engineering College, Chennai, Tamilnadu, India. r_ponnusamy@hotmail.com