International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 05 | May-2016 www.irjet.net p-ISSN: 2395-0072

Privacy Policy Inference of User Uploaded Images on Content Sharing Sites with Automatic Image Annotation

Ashita[1], Ambily Balaram[2]

[1] *Student, Department of Computer Science and Engineering, KMCTCE ,Calicut*
[2]*Assistant Professor, Department of Computer Science and Engineering, KMCTCE ,Calicut*

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 information .To provide security for the information, automated annotation of images are introduced which aims to create the meta data information about the images by using K-means clustering, KNN and SIFT descriptors. It results in better security, scalability, efficiency and accuracy.*

Key Words: Meta data, Online Social networking communities, Privacy Policy, Security, Automatic Image Annotation.

1. INTRODUCTION

The online social networking sites are the websites that enable users to join online communities, make new contacts, find old friends, and share common interests and ideas with large number of people across the world. It allows us to communicate with other internet users and build connections. The kinds and numbers of these content sharing sites have grown and participation of users also increased. As part of their participation lot amount of personal information are shared.Particularly young internet users share private images about themselves, their friends and classmates without being aware of the consequences. Photo sharing users often lack awareness of privacy issues. Many photos publicly shared by young people are of such a private nature that they would not show these images to their parents and teachers. A variety of risks are faced by individuals, such as identify theft, stalking, embarrassment, and blackmail as a result of proliferation of personal data .Despite these risks, many privacy mechanisms of content sharing sites are very weak.

There is a need to develop more security features in online social networks. Privacy is critical feature among the security mechanisms. In some situations, we like to share information only to best friends, family members and in other instances we like to share with strangers also. Existing sharing platforms do not support users in making adequate privacy decisions in multimedia resource sharing. On the contrary, these platforms quite often employ rather lax default configurations, and mostly require users to manually decide on privacy settings for each single resource. Given the amount of shared information this process can be tedious and error-prone [1].

To address the unique privacy needs of images existing proposals for automating privacy settings are

inadequate. A definition of internet privacy is it involves the right or mandate of personal privacy concerning the storing, repurposing, provision to third parties, and displaying of information pertaining to oneself via the internet. Internet privacy is a subset of data privacy. Privacy concerns have been articulated from the beginnings of large scale computer sharing. The privacy of user data can be given in two ways. 1. The user can enter the privacy preferences alone 2.Usage of recommendation systems which assist users for setting the privacy preferences.

The privacy policy of user uploaded data can be provided based on the personal characteristics. The privacy preferences of a user can be obtained from their profile information and relationships with others. The privacy policy of user uploaded image can be provided based on the content and meta data of user uploaded images. A hierarchical classification of images gives a higher priority to image content.

Privacy concerns with social networking services is a subset of data privacy, involving the binding personal privacy concerning storing, re-purposing, provision to third parties, and displaying of information through the Internet. Each day these sites process large amount of information. In order to gain access of other user's private information features like messages, invitations, photos, open platform application other applications are helpful. In the case of Facebook privacy features are weak .Various level of privacy are offered by these sites. There are even sites in which user doesn't reveal their actual names. It is also possible for users to block other users. Most users do not realize that while they may make use of the security features on Facebook the default setting is restored after each update.

The privacy strategies introduced by our participants may have initially achieved desired privacy protection and matched their initial mental models of audience and accessibility, but these strategies often failed now due to excessive use.

When making decisions regarding the disclosure of information and privacy, users who are new to Facebook do appear to consider the possibility of a broad and public audience and take into consideration the range of people who might access their profiles. The perception of online audience appears to shrink, as users continue to explore the Facebook interface, enlarge their social networks, and interact with their friends through these sites.

For sensitive and risky information a solution to over-disclosures is to enforce, or at least default to, more restrictive settings. This may help new users by providing immediate protection, and it may also protect even experienced users while by allowing them to customize their settings to share information when desired. Sensitive information can appear in many profile areas, so new defaults may do not match the desires of users. Privacy controls also need to be more visible, making them accessible while users are modifying their profile instead of located on separate pages. If the user ignores these privacy pages, they will never see their options for modifying the privacy settings.

2. RELATED WORKS

Many researches has been done in the area of privacy related with online social networking sites. In last few years various efficient methods have been proposed for privacy protection. Some noticeable work in area of privacy protection is as follows:

Based on the concept of **social circles [2]** privacy settings were introduced by Fabeah Adu-Oppong. To protect personal information web based solution is provided. The technique named Social Circles Finder automatically generates the friend's list. It is a technique that analyses the social circle of a person and identifies the intensity of relationship and therefore social circles provide a meaningful categorization of friends for setting privacy policies. This technique will allow the subject identify the social circles but not show them to the subject. The willingness of subject to share a piece of their personal information will be asked. The application finds the visual graph of users based on the answers.

PViz Comprehension Tool [3], an interface and system that corresponds more directly with how users model groups and privacy policies applied to their networks was developed by Alessandra Mazzia . According to automatically-constructed, natural sub-groupings of friends, and at different levels of granularity PViz allows the user to understand the

visibility of her profile. PViz is better than other current policy comprehension tools Facebook's Audience View and Custom Settings page. It also addresses the important sub-problem of producing effective group labels since the user must be able to identify and distinguish automatically-constructed groups.

Privacy Suites [4] 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 [1] 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 [5] 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.

YourPrivacyProtector [6] is a recommender system proposed by Kambiz Ghazinour that understands the social internet behavior of their privacy settings and recommending reasonable privacy options. The parameters used are user's personal profile, User's interests and User's privacy settings on photo albums .With the help of these parameters the system constructs the personal profile of the user. For a given profile of users it will automatically learn and assign the privacy options. It detects the possible privacy risks and allows users to see their current privacy settings on their social network profile, namely Facebook, and monitors frequently. Necessary privacy settings are adopted based on these risks.

A decentralised authentication protocol [7], 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) [8] 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 A3P-social.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. **Automatic Image Annotation (AIA)** helps to overcome the problem with meta data information.

3. A3P SYSTEM COMBINED WITH AIA

There is a need of tools to help users control access to their shared content is necessary. Toward addressing this, propose an Adaptive Privacy Policy Prediction (A3P) system (Figure 1) to help users to compose privacy settings for their images. In this framework a two level framework is introduced called as Adaptive Privacy Policy Prediction (A3P) system which aims to provide users a hassle free privacy settings by automatically generating personalized privacy policies.

3.1 System Architecture

A3P stands for Adaptive Privacy Policy Prediction system which helps users to derive the privacy settings for their images The A3P Architecture consists of followings blocks :Image classification – Meta based image classification and Content based image classification.

The overall data flow is the following. When user uploads an image, the image will be directly sent to the A3P-core. The A3P-core classifies the image and determines whether there is a need to involve the A3P-social. The A3P-social devides users into social communities with similar social context and privacy preferences, and continuously monitors the social groups. When theA3P-social is invoked, it automatically find outs the social group for the user and sends back the information about the group to the A3P-core for policy prediction. At the last, the predicted policy will be displayed to the user. If the user is fully satisfied by the predicted policy, user can just accept it. Otherwise, user can choose to revise the policy. The actual policy will be stored in the policy repository of the system for the policy prediction of the future uploads by user

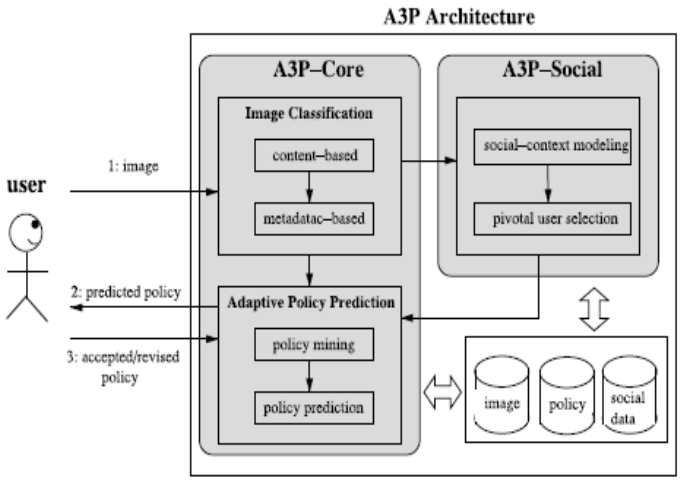


Fig -1: A3P system

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.

3.2 Image classification

Meta-based Image classification: The metadata-based 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.

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 waveletttransformation. 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. Using SHA1 algorithm hash code is generated for uploaded image.

3.3 Adaptive Policy Prediction

The Adaptive Policy Prediction consists of two following sub-parts:
1. Policy Mining
2. Policy Prediction

Policy Mining: A hierarchical mining approach for policy mining is used. Policy mining is carried out within the same category of the new image. The basic idea of this is to follow a natural order in which a user defines a policy. The 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.

Policy Prediction: It is an approach to choose the best candidate policy that follows the user's privacy tendency. To model the user's privacy tendency, define a notion of strictness level. The strictness level is a quantitative metric that describes how "strict" a policy is. a strictness level L is an integer with minimum value in zero, wherein the lower the value, the higher the strictness level.

.

3.4 Automatic Image Annotation

Automatic image annotation is a challenging problem in multimedia content analysis and computer vision. To annotate images a hierarchical framework is used. An image-filtering algorithm to remove most of the irrelevant images for an unlabeled image is presented first. For the unlabeled image, an image cluster is allocated using a discriminative model as the primary relevant image set in the algorithm. In the second stage, a hybrid annotation model is proposed to annotate images. K-means Algorithm is used to cluster the images in the training set and KNN Algorithm is used to determine the label of the cluster. SIFT Algorithm is used for feature extraction. Experiments have proved this method will provide better results. Figure 2 represents the proposed system.

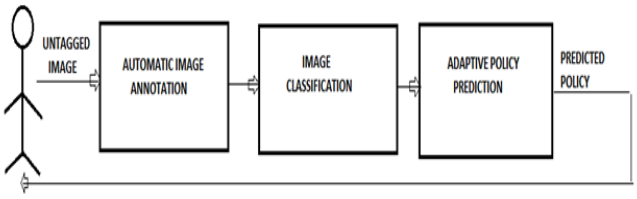


Fig -2: Proposed system

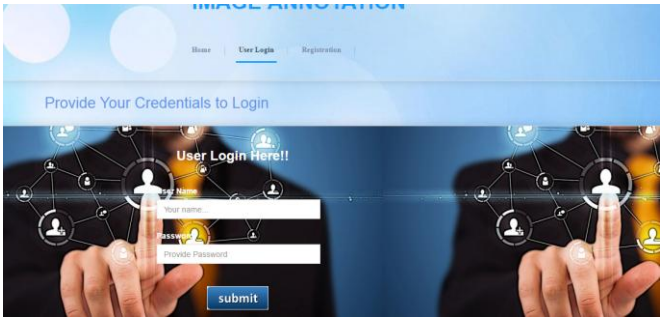
4. IMPLEMENTATION AND ANALYSIS

The A3P system combined with AIA is implemented using Java. The proposed method is tested on our own image set. A New user registration and Login Page is created. Based on user, he can upload and tag the images. The meta data based classification compares the tags with already uploaded images. The system will predict the policy accordingly. In Content-based classification features of image is extracted using SIFT Algorithm. AIA is done using K-Means and KNN Algorithm.

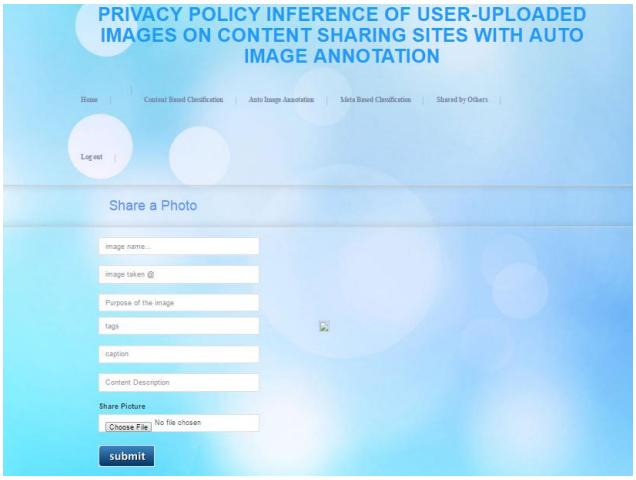
STEPS

Step 1:User Registration

Step 2: Login Page



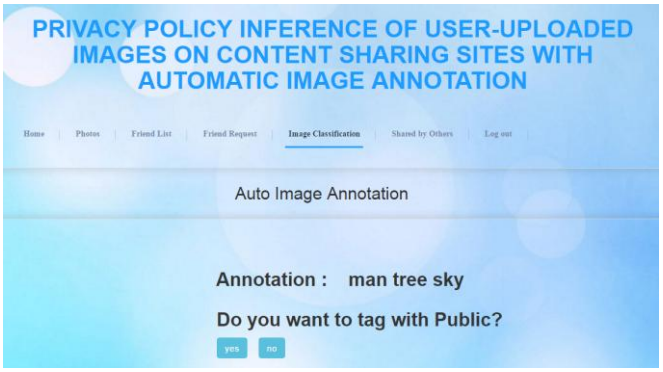
Step 3:Meta data based classification



Step 4: Content-based classification

Step 5: Policy Prediction

Step 6: Automatic Image Annotation



5. CONCLUSION

We have projected an Adaptive Privacy Policy Prediction (A3P) scheme that helps users computerize the privacy policy settings for their uploaded images. The A3P structure provides a wide-ranging structure to suppose privacy preferences based on the in order available for a given user. We also successfully tackled the subject of cold-start, leveraging social circumstance information. Automatic Image Annotation helps to overcome the issue of meta-data information of images being uploaded.

REFERENCES

[1] Sergej Zerr, Stefan Siersdorfer, Jonathon Hare, Elena Demidova ,"I Know What You Did Last Summer !:Priv acy-Aware Image Classification search"Proceedings of the 35th International ACM SIGIR conference on Research and development in information retrieval, 2012.

[2] A. Kapadia, F. Adu-Oppong, C. K. Gardiner, and P. P.Tsang, "Social circles: Tackling privacy in social network," in Proc. Symp. Sable Privacy Security, 2008.

[3] Alessandra Mazzia Kristen LeFevre and Eytan Adar,The PViz Comprehension Tool for Social Network Privacy Settings, Tech. rep., University of Michigan, 2011.

[4] J. Bonneau, J. Anderson, and L. Church, "Privacy suites: Shared privacy for social networks," in Proc. Symp. UsablePrivacy Security,2009

[5] Peter F. Klemperer, Yuan Liang, Michelle L. Mazurek, "Tag, You Can See It! Using Tags for Access Control in

Photo Sharing", Conference on Human factors in Computing Systems, May 2012.

[6] Kambiz Ghazinour, Stan Matwin and Marina Sokolova, Social "Yourprivacyprotector: A Recommender System For Privacy Settings In Social Networks", International Journal of Security, Privacy and Trust Management (IJSPTM) Vol 2, No 4, August 2013.

[7] C. A. Yeung, L. Kagal, N. Gibbins, and N. Shadbolt, "Providing access control to online photo albums based on tags and linkeddata,"in Proc. Soc. Semantic Web: Where Web 2.0 Meets Web 3.0 at the AAAI Symp., 2009, pp. 9–14.

[8] Anna Cinzia Squicciarini, Member, IEEE, Dan Lin,Sundareswaran, and Joshua Wede, "Privacy Policy Inference of User-Uploaded Images on Content Sharing Sites", IEEE Transactions on Knowledge and Data Engineering, Vol. 27,NO. 1, January 2015.

[9] R. Ravichandran, M. Benisch, P. Kelley, and N. Sadeh, "Capturing social networking privacy preferences," in Proc. Symp. Usable Privacy Security, 2009.

[10] Yuan-yuan ca., Zhi-chun mu, Yan-fei ren ,and Guo-qing xu "A Hybrid Hierarchical Framework For Automatic Image Annotation"in Proc of the 2014 International Conference on Machine Learning and Cybernetics, Lanzhou, 13-16 July, 2014.

AUTHOR PROFILE

Ashita is pursuing her M.Tech degree in Computer Science and Engineering from KMCT College of Engineering, Calicut University. She obtained her B.Tech Degree in Information Technology and Engineering from MES College of Engineering, in 2014.

Ambily Balaram, is Assistant Professor, Department of Computer Science and Engineering, KMCT College of Engineering, Calicut University.She obtained her B.Tech degree in Computer Science and Engineering from Government College of Engineering,Wayanad in 2007.