ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

Review: Privacy of User Uploaded Information through Prediction

Snehal Vilas Gadhave[1], Prof. Lomesh Ahire[2],

[1]Student, Dept. of M.E. Comp. Network, Nutan Maharashta Institute of Engineering and Technology, Pune, India

[2]Professor, Dept. of M.E.Comp.Network, Nutan Maharashta Institute of Engineering and Technology, Pune, India

ABSTRACT: User shares images through social sites, to maintain privacy has become a major problem, by a recent wave of publicized incidents where inadvertently user shares personal information. In focus of these incidents, there is need of tool to help control of users access to their shared content is apparent. To address this need, we propose a system which is an Adaptive Privacy Policy Prediction (A3P) system to help user's images to compose privacy settings. We observe the role of image content, social context, and metadata as possible indicators of user's privacy preferences. According to user available history on the site,we propose a two level framework which determines the best available privacy policy for the user's images being uploaded. Our solution relies for image categories on an image classification framework which may be associated with similar policies, and according to users' social features, a policy prediction algorithm it automatically generate a policy for each newly uploaded image.

KEYWORDS: Web-based services, online information services.

I. Introduction

Just consider social context like as one's friend list. While interesting, they may not be sufficient to address the challenges brought by image files for which privacy might vary substantially not just because of social context but also due to the actual image content. As long as images, authors in have presented an expressive language for images uploaded in social sites. This work is totally complementary to ours as we do not deal with policy expressiveness, but it 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, photo ranking and interpretation, retrieval, also in the context of online photo sharing sites of these work which is probably closest to ours. It explores privacy-aware image classification using a mixed set of the features, also both content and meta-data. However this is a binary classification (private versus public) problem.

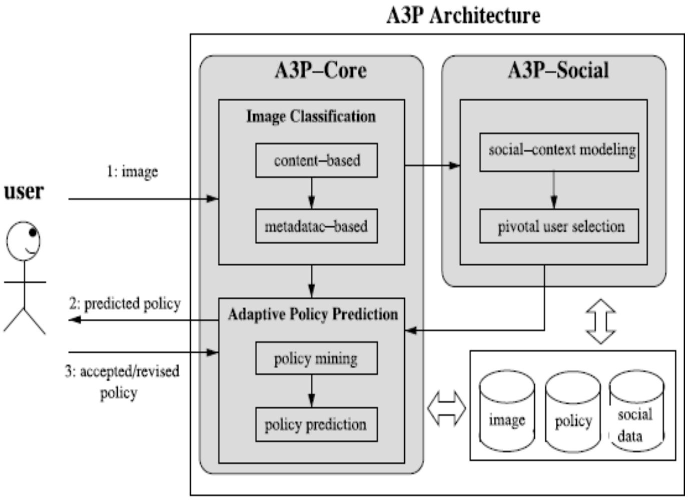

Figure 1: The Architecture ofAdaptive Privacy Policy Prediction (A3P)

DOI: 10.15680/IJIRCCE.2015. 0311187

So the classification task is very different than ours. Also, the authors do not deal with the issue of cold start the proposed System an Adaptive Privacy Policy Prediction (A3P) which helps user to automate the privacy policy settings for their uploaded images. To the infer privacy preferences based on the information available for a given user A3P system provides a comprehensive framework. We also effectively solved the issue of cold-start problem, leveraging social context information. A3P system is a practical tool that offers significantimprovements over the current approaches to privacy it is proved by our experimental study. The “social media” term refers to mobile services and wide range of Internet-based that allow users to participate in online exchanges, join online communities or contribute user-created content. Online social networks are websites that allows user to build connections and relationships to other Internet users. Social networks stores remotely information, rather than on a user’s personal computer. Social networking can be used to keep in touch with friends, find people with similar interests and ideas, and make new contacts. The relation between a person’s social network and privacy is multi-faceted. There is a need to develop more security mechanisms for different communication technologies, particularly online social networks. Privacy is essential to the design of security mechanisms. Most social networks providers have offered privacy settings to allow or deny others access to personal information details. In certain occasions we want information about ourselves to be known only by a small circle of close friends, and not by strangers. In other instances, we are willing to reveal personal information to anonymous strangers, but not to those who know us better. Social network theorists have discussed the relevance of relations of different strength and depth in a person’s social network and the importance of so-called weak ties in the flow of information across different nodes in a network.

II. RELATED WORK

Our work is related to works on privacy setting configuration in the social sites,privacy analysis of online images and recommendation systems.

A. Privacy Setting Configuration

Many recent works have been studied how to automate the task of privacy settings. Bonne au proposed the concept of privacy suites which recommends to users a privacy settings suites that “expert” users or other trusted friends have already set, sothe normal users can either directly choose a setting or only need to do minor modification. Similarly, Danezis have proposed a machine-learning based approach to automatically extract privacy settings from the social context within the data is produced. Parallel to the work of Danezis, Adu-Oppong develop privacy settings based on a concept of “Social Circles” which contains of friends clusters which is formed by partitioning users’ friend lists. Ravichandran studied how to predict a privacy preferences of user for location-based data (i.e., share her location or not) based on location and time of day. Fang proposed a privacy wizard to help users grant privileges to their friends. The wizard asks users to first to assign the privacy labels to selected friends, and later uses this as input to construct a classifier which classifies friends which is based on their profiles and automatically assign privacy labels to the friends whose unlabelled. More recently, Klemperer studied whether the keywords and captions with which users tag their photos can be used to help users more inspirately create and maintain access-control policies. Their findings are in line with our approach: tags created for purpose of organizational that can be repurposed to help create reasonably accurate access-control rules.

The aforementioned approaches focuses on deriving settings of policies for only traits, so they mainly considers social context such as one’s friend list. little interesting, they may not be enough to address challenges brought by image files for which privacy might vary substantially not just because of social context but also due to the actual image content. As long as images, authors have presented an expressive language for images uploaded in the social sites. This work iscomplementary to ours as we do not behave with policy expressiveness, which rely on common forms policy specification for our predictive algorithm.

B. Recommendation Systems

Our work is completed related to some existing recommendation systems which employ the machine learning techniques. Chen et proposed a system which is named Sheep Dog, it automatically inserts photo into appropriate groups and recommend suitable tags for users on the Flickr. They adopt a concept detection to predict relevant concepts (tags) of a photo. Choudhury proposed a recommendation framework for connecting image content with communities in online social media. Characterization of images through three types of features: visual features, social interaction user generated text tags and from which they recommend the most likely groups for a given image. Likely, Yu proposed an automated recommendation system for a user's image to suggest suitable, photo-sharing groups. There is also a large body of work on the personalization and customization of tag-based information retrieval, which utilizes techniques like that association rule mining. For example, to recommend groups for Flickr users it proposes an interesting experimental evaluation of several collaborative filtering algorithms. These approaches have a different goal to our approach as they focus on sharing instead of protecting the content.

III. A3P FRAMEWORK

A. *Preliminary Notions*

Users can be express their privacy preferences about their content disclosure preferences and with their socially connected users via privacy policies. We can define privacy policies according to Definition 1.Generally our policies are inspired by popular content sharing sites (i.e., Facebook, Flickr,picasa),the actual implementation mainly depends on the specific content management site structure as well as implementation.

Definition 1. A privacy policy P of user u consists of the following components:

_ Subject (S): A set of users socially connected to u.

_ Data (D): A set of data items shared by u.

_ Action (A): A set of actions granted by u to S on D.

_ Condition (C): In order to perform the granted actions boolean expression must be satisfied. In the definition, users in S can be represented by their roles and identities, role (e.g., family, friend, coworkers), or organizations (e.g. profit, non-profit organization).

In the user's profile D is the set of images. Along with some associated metadata each image has a unique ID such astags "birthday","vacation". Images can be grouped into albums further. For A, we assume four common types of actions: {view, tag, download, comment,}. Lastly, the condition component C specifies when the granted action is very effective. Component C is a Boolean expression on the grantee's attributes like time, age and location. For better understanding, an example policy is given here.

Example 1. Alice would like to permit his friends and co-workers to tag and comment images in the album named "vacation album" and the image named "summer.jpg" before year 2012. Her privacy preferences can be expressed by the bellowed policy:

P: {friend, coworker}, {vacation_album, summer.jpg}, {comment, tag}, (date< 2012).

B. *System Overview*

The A3P system includes of two main components: A3P-core As well as A3P-social. The overall data flow is mentioned below. When an image is uploaded by user, the image will be first sent to the A3P-core. The A3P-core first makes classifies the image and determines whether there is a need to invoke the A3P-social. In many cases, A3P-core predicts policies for the users directly which is based on their historical behaviour. If one of the following two cases are verified true, A3P-core will invoke A3Psocial:

(i) To conduct policy prediction the user does not have enough data for the type of the uploaded image;

(ii) The recent major changes among the user's community are detected by the A3P core about their privacy practices contiguous with user's increase of social networking activities (new posts on one's profile and addition of new friends etc.). In above case, it will be beneficial to report to the user of the latest privacy practice of social communities that have similar background as like the user. The A3P-social makes group of users into social communities with similar privacy preferences social context and continuously monitors the social groups. When the A3P-social is invoked or called, social group is automatically identified for the user and sends back the information to the A3P-core about the group for policy prediction. At the last, the policy which is predicated will be displayed to the user. If the user is totally satisfied by the predicted policy, he or she can accept it. Otherwise, the user can also choose to revise the policy. In the policy repository the actual policy will be stored of the system for the policy prediction of the future uploads.

IV. A3P CORE

Major two components in the A3P-core: (i) Image classification (ii) Adaptive policy prediction. For each user, his/her images are first classified which are based on metadata and content. For the policy prediction the privacy

DOI: 10.15680/IJIRCCE.2015. 0311187

policies of each category of images are analysed. To adopting a two-stage approach is more suitable for the recommendation of policy than applying the common one-stage data mining approaches to mine both policies together and image features. Recall that when a new image is uploaded by user, the user is waiting for a recommended policy. The approach two-stage allows the system to employ first stage to classify the new image and find the candidate sets of image for the recommendation of subsequent policy .For the one-stage mining approach, it unable to locate the right class of the new image because the classification criteria needs both image features and the policies policies of the new image are not available yet.Moreover,combining both image policies and features into a singleclassifier would lead to a system which is verydependent to the specific syntax of the policy, the wholelearning model would need to change, if a change in the supported policies were to be introduced.

A. Image Classification

We propose a hierarchical image classification which classifies images which based on their contents and then refine each category into subcategories based on their metadata to attain groups of images that may be associated with similar privacy preferences. Images which do not have metadata will be grouped only by content. Like a hierarchical classification gives a higher priority to minimize the influence of missing tags and image contents.

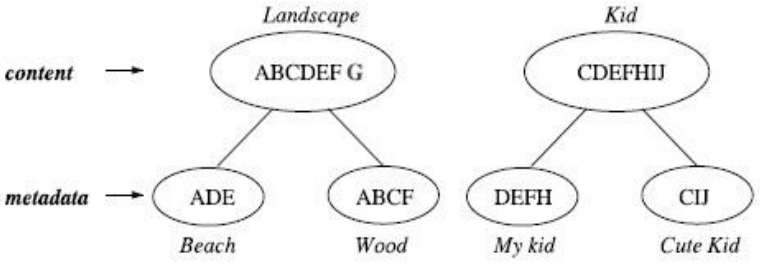

fig1.Two-level Image classification.

Important is, it is possible that some images are included in multiple categories as far like they obtain the typical content metadata or features of those categories. Moreover, Fig. 2 shows an example of image classification for 10 images named as A, B, C, D, E, F, G, H, I, J,respectively. The content-based classification makes two categories: “landscape” and “kid”. Images C, D, E and F are contained in both categories as they show kids playing outdoor which satisfy the two themes: “landscape” and “kid”. These two types are further divided into subcategories based on the tags associated with the images. As in result, we obtain two subtypes under each theme respectively. Notice that image G is not shown in any subtype as it does not have any tag; image a shows up in both subtypes because it has tags indicating both “beach” and “wood”.

a. Content-Based Classification

Our appeal for content-based classification is based on an efficient and yet accurate image similarity approach. Emphatically, our classification algorithm compares image signatures defined based on theoretical and sanitized version of Haar wavelet transformation. For every image, the wavelet transform encodes frequency and spatial information related to image colour, invariant transform, size, symmetry, shape, texture, etc. After that a small number of coefficients are selected to form the signature of the image. Likeness among the image contents is then resolute by the distance among their image signatures.

Our selected likeness criteria include symmetry, texture, shape (radial symmetry and phase congruency), and SIFT. We also account for size and colour. We sets the system to start from five generic classes of image: (a) explicit (e.g., nudity, drinking, violence etc.), (b) adults, (c) kids, (d) scenery (e.g., beach, mountains), (e) animals. So a pre-processing step, five baseline classes are populated by manually assigning to each class a number of images crawled from Google images, resulting in about 1,000 images per class. Having a huge image data set beforehand decreases the chance of misclassification. Thereafter, we generate signatures of all the images and store them in the database. Onto adjusting the settings of our content classifier, we conducted some prelim test to evaluate its accuracy. Alongside, we tested our classifier it against a ground-truth data set, Image-net.org. In Image-net, over 10 million images are classified and collected according to the WordNet structure. For every image class, we use the first half set of the images as the training data set and classify the next 800 images. Correct classification result was recorded as if the sunset's main search term or the direct hyponym is returned as a class. Above 94 percent is the average accuracy of our classifier. Have verified the accuracy of the classifier, we now discuss how it is used in the context of the A3P core.

The image is handled as an input query image, when it is uploaded. The signature of the freshly uploaded image is compared with the signatures of images in the current image database. To decide the class of the uploaded image, we find first m closest matches of it. The class of uploaded image is then calculated as the class to which mainly of the m images belong. If no of the predominant class is found, then a new class is created for the image. Then, if the predicted policy for this new image turns out correct, so the image will be inserted into the compatible image category in our image database, to help cultivated future policy prediction. In the current prototype, m is set to 25 which gain is using a small training data set.

b. Metadata-Based Classification

The metadata-based classification makes group images into subtypes underneath aforementioned baseline categories. The process have of three main steps. To extract keywords from the metadata associated with an image is the first step. The tags are considered as a metadata in our work, captions, and comments. We perceive all the verbs, nouns, and adjectives in the metadata and store them as metadata vectors.

B. Adaptive Policy Prediction

A predicted policy of a newly uploaded image to the user for his/her reference is provided by Adaptive Policy Prediction Algorithm. Most importantly, the predicted policy will reflect the feasible changes of a user's privacy concerns. The prediction process contains three main phases: (i) policy normalization; (ii) policy mining; and (iii) policy prediction. The policy normalization is an intelligible decomposition process to transmute a user policy into a set of molecular rules in which the data (D) component is a set of single-element.

a. Policy Mining

For policy mining we propose a hierarchical mining approach. To discover popular patterns in policies approach leverages association rule mining techniques is used. Policy mining is carried out within the similar category of the new image because images in the same category are more probably under the same level of privacy protection. The total basic idea of the hierarchical mining is to imitate a natural order in which a user determines a policy. Given an image, most of the time user first decides who can access the image, then thinks about what specific access rights (e.g.,download or view only to find out popular patterns in policies) should be given, and lastly refine the access conditions like a setting the expiration date.The hierarchical mining first look for popular subjects which is given by the user, after that see the famous actions in the policies containing the popular subjects, and finally for popular conditions in the policies containing both popular subjects and conditions.

b. Policy Prediction

The policy mining phase might be generate several candidate policies while the objective of our system is to return the most ensuring one to the user. So, we show an approach to choose the ever best candidate policy that imitates the user's privacy trend.

V. A3P SOCIAL

The A3P-social exercise a multi-criteria peculation mechanism that generates delegate policies by leveraging valuable information related to the user's social context and his general aspect toward privacy. As mentioned earlier, A3P-core in two scenarios is invoked by A3P social. First one is when the user is a newbie of a site, and do not have enough images stored for the A3P-core to signify customized and meaningful policies. The second is when the system notifies significant changes of privacy tendency in the social circle of user, which might be of interest for the user to probably adjust his/her privacy settings respectively. In that follows, we first shows the types of social context regard by A3P-Social, after that present the policy recommendation process.

A. Modelling Social Context

We examine that users with same background having tendency of similar privacy concerns, as seen in existing research studies and also assured by our collected data. This observation motivates us to build a social context modelling algorithm that can catch the commonage social factors of users and identify perception formed by the users with same privacy concerns. The identified perceptions who have a prosperous set of images can after on serve as the base of respectively policy recommendation. Two major steps are contained in the social context modelling algorithm. The first important step is to identify as well as formalize potentially main factors that might be informative of one's privacy settings. Users are grouped based on the identified factors this is the second step. First of all, we model each user's social context as a list of attributes like: {sc1, sc2; . . . ; scn}, where sci is denoted to a social context attribute, and total number of distinct attributes in the social networking site is denoted by **n**. These social context attributes are extracted from the profile of user.

B. Identifying Social Group

We now inset the policy recommendation process which is based on the social groups obtained from the earlier step. Assume that a new image is uploaded by user U and the A3P-core called the A3P-social for policy recommendation. The social group is found by the A3P-social which is most same to user U and then select the representative user in the social group with that his images which is sent to the A3P-Core policy prediction module to generate the recommended policy for user U. The number of users in social network may be tremendous and a large number of social groups may be join by a users.it is very time consuming to compare the new user's social context attributes against the frequent pattern of each social group. In case of to speed up the group identification process and ensure required response time, we leverage the inverted file structure to describe the social group information.

VI. EXPERIMENTAL EVALUATION

Policy prediction accuracy and user acceptability is evaluated by us in terms of effectiveness of our A3P system. The A3P was described as a Java file embedded in an open source content management site, deployed using an Apache server.

VII. CONCLUSION

The proposed system an Adaptive Privacy Policy Prediction (A3P) that is always ready to help users automate the privacy policy setting for their uploaded images. The A3P system provides a relative framework to conclude privacy preferences which based on the information available for a given user. We also collectively solve the issue of leveraging social context information, cold-start. Our experimental study proves that our A3P is a practical tool that gives significant improvements over current approaches to privacy.

REFERENCES

1. 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.
2. R. Agrawal and R. Srikant,"Fast algorithms for mining association rules in large databases," in Proc. 20th Int. Conf. Very Large Data Bases, 1994, pp. 487–499.
3. S. Ahern, D. Eckles, N. S. Good, S. King, M. Naaman, and R. Nair, "Over-exposed?: Privacy patterns and considerations in online and mobile photo sharing," in Proc. Conf. Human Factors Comput.
Syst., 2007, pp. 357–366.
4. 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.
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. D. G. Altman and J. M. Bland ,"Multiple significance tests: The bonferroni method," Brit. Med. J., vol. 310, no. 6973, 1995.
7. J. Bonneau, J. Anderson, and L. Church, "Privacy suites: Shared privacy for social networks," in Proc. Symp. Usable Privacy Security,2009.

DOI: 10.15680/IJIRCCE.2015. 0311187