International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 04 Issue 06
May 2017

Privacy Policy Of Users Images On Social Networking Sites

V. Rama Rao
Associate Professor, Gandhi Academy of Technical Education, Kodad, Telangana

Abstract— 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. Toward addressing this need, we propose an Adaptive Privacy Policy Prediction (A3P) system to help users compose privacy settings for their images. We examine the role of social context, image content, and metadata as possible indicators of users' privacy preferences. We propose 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. Our solution relies on an image classification framework for image categories which may be associated with similar policies, and on a policy prediction algorithm to automatically generate a policy for each newly uploaded image, also according to users' social features. Over time, the generated policies will follow the evolution of users' privacy attitude. We provide the results of our extensive evaluation over 5,000 policies, which demonstrate the effectiveness of our system, with prediction accuracies over 90 percent.

1INTRODUCTION

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. However, semantically rich images may reveal contentsensitive information []. Consider a photo of a students 2012 graduationceremony, for example. It could be shared within a Google+ circle or Flickr group, but may unnecessarily expose the studentsBApos familymembers and other friends. Sharing images within online content sharing

sites,therefore,may quickly leadto unwanted disclosure and privacy violations [3], [24]. Further, the persistent nature of online media makes it possible for other users to collect rich aggregated information about the owner of the published content and the subjects in the published content .The aggregated information can result in unexpected exposure of one's social environment and lead to abuse of one's personal information.

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. In this paper, we propose an Adaptive Privacy Policy Prediction (A3P) system which aims to provide users a hassle free privacy settings experience by automatically generating personalized policies. The A3P system handles user uploaded images, and factors in the following criteria that influence one's privacy settings of images: The impactof social environment and personalcharacteristics. Social context of users, such as their profile information and relationships with others may provide usefulinformationregardingusers' privacypreferences. For example, users interested in photography may like to share their photos with other amateur photographers. Users who have several family members among their social contacts may share with them pictures related to family events. However, using common policies across all users or across users with similar traits may be too simplistic and not satisfy individual preferences. Users may have drastically different opinions even on the same type of images. Forexample,aprivacyadversepersonmaybewi lling to share all his personal images while a

more conservative person may just want to share personal images

2 RELATED WORK

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

2.1 Privacy Setting Configuration

Several recent works have studied how to automate the task of privacy settings (e.g.,. Bonneau et al. proposed the concept of privacy suites which recommend to users a suite of privacy settings that "expert" users or other trusted friends have already set, so that normal users can either directly choose a setting or only need to do minor modification. Similarly, proposed a machine-learning based approach to automatically extract privacy settings from the social context within which the data is produced. Parallel to the work of Danezis, Adu-Oppong et al. develop privacy settings based on a concept of "Social Circles" which consist of clusters of friends formed by partitioning users' friend lists. Ravichandran et al. [30] studied how to predict a user's privacy preferences for location-based data (i.e., share her location or not) based on location and time of day. Fang et al. proposed a privacy wizard to help users grant privileges to their friends. The wizard asks users to first assign privacy labels to selected friends, and then uses this as input to construct a classifier which classifies friends based on their profiles and automatically assign privacy labels to the unlabeled friends. More recently, Klemperer et al. studied whether the keywords and captions with which users tag their photos can be used to help users more intuitively create and maintain access-control policies. Their findings are inline with our approach: tags created for organizational purposes can be repurposed to help create reasonably accurate access-control rules. The aforementioned approaches focus on deriving policy settings for only traits, so they mainly 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 [41] 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.

2.2 Recommendation Systems

Our work is related to some existing recommendation systems which employ machine learning techniques. Chen et al. [9] proposed a system named SheepDog to automatically insert photos into appropriate groups and recommend suitable tags for users on Flickr. They adopt concept detection to predict relevant concepts (tags) of a photo. Choudhury et al. [10] proposed a recommendation framework to connect image content with communities in online social media. They characterize images through three types of features: visual features, user generated text tags, and social interaction, from which they recommend the most likely groups for a given image. Similarly, Yu et al. proposed an automated recommendation system for a user's images to suggest suitable photo-sharing groups. There is also a large body of work on the customization and personalization of tag-based information retrieval which utilizes techniques such as association rule mining. For example, proposes an interesting experimental evaluation of several collaborative filtering algorithms to recommend groups for Flickr users. These approaches have a totally different goal to our approach as they focus on sharing rather than protecting the content.

3 A3P FRAMEWORK

3.1 Preliminary Notions

Users can express their privacy preferences about their content disclosure preferences with their socially connected users via privacy policies. We define privacy policies according to Definition 1. Our policies are inspired by popular content sharing sites (i.e., Facebook, Picasa, Flickr), although the actual implementation depends on the specific content-management site structure and 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): A boolean expression which must be

satisfied in order to perform the granted actions.

In the definition, users in S can be represented by their identities, roles (e.g., family, friend, coworkers), or organizations (e.g., non-profit organization, profit organization). D will be the set of images in the user's profile. Each image has a unique ID along with some associated metadata like tags "vacation", "birthday". Images can be further grouped into albums. As for A, we consider four common types of actions: {view, comment, tag, download}. Last, the condition component C specifies when the granted action is effective. C is a Boolean expression on the grantees' attributes like time, location, and age. For better understanding, an example policy is given below.

3.2 System Overview

The A3P system consists of two main components: A3P-core and A3P-social. The overall data flow is the following. When a user uploads an image, the image will be first sent to the A3P-core. The A3P-core classifies the image and determines whether there is a need to invoke the A3P-social. In most cases, the A3P-core predicts policies for the users directly based on their historical behavior. If one of the following two cases is verified true, A3P-core will invoke A3Psocial: (i) The user does not have enough data for the type of the uploaded image to conduct policy prediction; (ii) The A3P-core detects the recent major changes among the user's community about their privacy practices along with user's increase of social networking activities (addition of new friends, new posts on one's profile etc). In above cases, it would be beneficial to report to the user the latest privacy practice of social communities that have similar background as the user.

4 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.

Recall that when a user uploads a new image, the user is waiting for a recommended policy. The two-stage approach allows the system to employ the first stage to classify the new image and find the candidate sets of images for the subsequent policy recommendation.

4.1 Image Classification

To obtain 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. 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 creates two categories: "landscape" and "kid". Images C, D, E and F are included in both categories as they show kids playing outdoor which satisfy the two themes: "landscape" and "kid". These two categories are further divided into subcategories based on tags associated with the images. As a result, we obtain two subcategories under each theme respectively. Notice that image G is not shown in any subcategory as it does not have any tag; image A shows up in both subcategories because it has tags indicating both "beach" and "wood".

4.1.1 Content-Based Classification

Our approach to content-based classification is based on an efficient and yet accurate image similarity approach. Specifically, our classification 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. Our selected similarity criteria include texture, symmetry, shape (radial symmetry and phase congruency and We also account for color and size. We set the system to start from five generic image classes: (a) explicit(e.g., nudity, violence, drinking etc), (b) adults, (c) kids, (d) scenery (e.g., beach, mountains), (e) animals. As a preprocessing step, we populate the five baseline classes by manually assigning to each class a number of images crawled from Google images, resulting in about 1,000 images per class. Having a large image data set beforehand reduces the chance of misclassification. Then, we generate signatures of all the images and store them in the database. Upon adjusting the settings of our content classifier, we conducted some preliminary test to evaluate its accuracy. Precisely, we tested our classifier it against a ground-truth data set, Image-net.org . In Image-net, over 10 million images are collected and classified according to the wordnet structure. For each image class, we use the first half set of images as the training data set and classify the next 800 images. The classification result was recorded as correct if the synset's main search term or the direct hypernym is returned as a class. The average accuracy of our classifier is above 94 percent. Having verified the accuracy of the classifier, we now discuss how it is used in the context of the A3P core. When a user uploads an image, it is handled as an input query image. The signature of the newly uploaded image is compared with the signatures of images in the current image database. To determine the class of the uploaded image, we find its first m closest matches. The class of the uploaded image is then calculated as the class to which majority of the m images belong. If no predominant class is found, a new class is created for the image. Later on, if the predicted policy for this new image turns out correct, the image will be inserted into the corresponding image category in our image database, to help refine future policy prediction. In our current prototype, m is set to 25 which is obtained using a small training data set.

5 A3P-SOCIAL

The A3P-social employs a multi-criteria inference mechanism that generates representative policies by leveraging key information related to the user's social context and his general attitude toward

privacy. As mentioned earlier, A3Psocial will be invoked by the A3P-core in two scenarios. One is when the user is a newbie of a site, and does not have enough images stored for the A3P-core to infer meaningful and customized policies. The other is when the system notices significant changes of privacy trend in the Major Level Subject Action 0 family view 1 family comment 2 family tag 3 family download 4 friend view 5 friend comment 6 friend tag 7 friend download 8 coworker view 9 coworker comment 10 coworker tag 11 coworker download 12 stranger view 13 stranger comment 14 stranger tag 15 stranger download user's social circle, which may be of interest for the user to possibly adjust his/her privacy settings accordingly. In what follows, we first present the types of social context considered by A3P-Social, and then present the policy recommendation process.

5.1 Modeling Social Context

We observe that users with similar background tend to have similar privacy concerns, as seen in previous research studies and also confirmed by our collected data. This observation inspires us to develop a social context modeling algorithm that can capture the common social elements of users and identify communities formed by the users with similar privacy concerns. The identified communities who have a rich set of images can then serve as the base of subsequent policy recommendation. The social context modeling algorithm consists of two major steps. The first step is to identify and formalize potentially important factors that may be informative ofone'sprivacy settings.The second step isto group users based on the identified factors. First, we model each user's social context as a list of attributes: {sc1, sc2;...;scn}, where sci denote a social context attribute, and n is the total number of distinct attributes in the social networking site. These social context attributes are extracted from users' profiles. Besides basic elements in users' profiles, many social sites also allow users to group their contacts based on relationships (e.g., friends, family members). If such grouping functionality is available, we will consider its influence on privacy settings too. In a social site, some users may only have their family members as contacts, while some users may have contacts including different

kinds of people that they met offline or on the Internet.

6EXPERIMENTAL EVALUATION

We evaluate the effectiveness of our A3P system in terms of the policy prediction accuracy and user acceptability. The A3P was implemented as a Java file embedded in an open source content management site, deployed using an Apache server.

6.1 Experimental Settings

We conduct and collect data sets for two types of experiments: survey-based study and direct user evaluation. Survey-based study and data collection. We collected two sets of actual user-specified policies to be used as ground truth for our evaluation. Data collection 1. This study involved 88 participants (48 female and 40 males) who were recruited from a large US university community (staff, students, and the community at large). Their average age is 26.3 years old (Range: 18-39). The participants completed at least 90 percent of the questionnaire consisting of two parts. The first part contains questions related to one's background information and online privacy practices and the second part is to collect user-specified policies. In the first part of the questionnaire, the participants were asked to indicate any social networks they were a part of (98 percent indicated Facebook and 37 percent also indicated others like Myspace). In terms of usage frequency, 95 percent of the respondents accessed social network sites at least once a week, with 76 percent of reporting that they were daily users. We also asked 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.

6.2.1 A3P-

Core Our first experiment compares A3P-core with alternative prediction approaches. In particular, we use a straw man solution as the baseline approach, whereby we sample at random a small set of image settings from the same user and use them to determine a baseline setting (by counting the most frequent items). The baseline settings are applied to all images of the users. Further, we compare the A3Pcore with two variants of itself, in order to evaluate the contribution of each component in the A3P-core made for privacy prediction. The first variant uses

only content-based image classification followed by our policy mining algorithm, denoted as "Content+Mining". The second variant uses only tag classification followed by the policy mining, denoted as "Tag+Mining". All the algorithms were tested against the collected real user policies. Fig. 4 shows the percentage of predicted policies in four groups: "Exact Match" means a predicted policy is exactly the same as the real policy of the same image; "x-component Match" means a predicted policy and its corresponding real policy have x components (i.e., subject, action, condition) fully matched; "No match" simply means that the predicted policy is wrong for all components. As shown in the figure, each component of the A3P-core singularly contributes toward policy prediction, however, none of them individually equalizes the accuracy achieved by the A3P-core in its entirety.

6.2.2 Analysis of Users' Characteristics We are also interested in examining whether our algorithm performs better for users with certain characteristics. Therefore, we study possible factors relevant to the performance of our algorithm. We used a least squares multiple regression analysis, regressing performance of the A3P-core to the following possible predictors: Frequency of social network use was measured on a frequency rating scale (1 ¼ daily; 2 ¼ weekly; 3 ¼ monthly; 4 ¼ rarely; 5 ¼ never) with the item 'How often do you access Social Network Sites?' Privacy settings take time was measured on a Likert Scale (5-point rating scale, where 1 ¼ strongly agree and 5 ¼ strongly disagree) with the item 'Changing privacy settings for images uploaded on a social site can be very time consuming.' Frequency of sharing pictures was measured using three items (a ¼ 0:69) rated on a Likert scale.

7 CONCLUSION

We have proposed 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.

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.

[8] J. Bonneau, J. Anderson, and G. Danezis, "Prying data out of a social network," in Proc. Int. Conf. Adv. Soc. Netw. Anal. Mining., 2009, pp.249–254.

[9] H.-M. Chen, M.-H. Chang, P.-C. Chang, M.-C. Tien, W. H. Hsu, and J.-L. Wu, "Sheepdog: Group and tag recommendation for flickr photos by automatic search-based learning," in Proc. 16th ACM Int. Conf. Multimedia, 2008, pp. 737–740.

[10] 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.

[11] L. Church, J. Anderson, J. Bonneau, and F. Stajano, "Privacy stories: Confidence on privacy behaviors through end user programming," in Proc. 5th Symp. Usable Privacy Security, 2009.

[12] R. da Silva Torres and A. Falc~ao, "Content-based image retrieval: Theory and applications," Revista de Inform□atica Te□orica e Aplicada, vol. 2, no. 13, pp. 161–185, 2006.

[13] R. Datta, D. Joshi, J. Li, and J. Wang, "Image retrieval: Ideas, influences, and trends of the new age," ACM Comput. Surv., vol. 40, no. 2, p. 5, 2008.

[14] J. Deng, A. C. Berg, K. Li, and L. Fei-Fei, "What does classifying more than 10,000 image categories tell us?" in Proc. 11th Eur. Conf. Comput. Vis.: Part V, 2010, pp. 71–84. [Online]. Available: http://portal.acm.org/citation.cfm?id=1888150.1888157

[15] 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.

[16] L. Geng and H. J. Hamilton, "Interestingness measures for data mining: A survey," ACM Comput. Surv., vol. 38, no. 3, p. 9, 2006.

[17] Image-net data set. [Online]. Available: www.image-net.org, Dec. 2013