PRIVACY POLICY RECOMMENDATION FOR USER UPLOADED IMAGES ON SOCIAL SITES

Jayashree Walunj[1], Bharat Burghate[2]

[1] Student, Department of Computer Engineering, JSPM's, BSIOTR, Maharashtra, India
[2] Assistant Professor Department of Computer Engineering, JSPM's, BSIOTR, Maharashtra, India

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 advertised incidents where users unknowingly shared personal information. In light of these incidents, the need of tools to help users control access to their shared content is obvious. 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.

Keyword: *Privacy Policy, Recommendation system, Social Networking, feature extraction.*

1. INTRODUCTION

Pictures are presently one of the key empowering influences of clients' network. Sharing happens both among already settled gatherings of known individuals or social circles (e. g., Google+, Flickr or Picasa), furthermore progressively with individuals outside the client's social circles, for purposes of social revelation to assist them with recognizing new associates and find out about companion's hobbies and social environment. Be that as it may, semantically rich pictures may uncover content sensitive data. Consider a photograph of an understudies 2012 graduation ceremony, for instance. It could be shared inside of a Google+ circle or Flickr bunch, yet might superfluously uncover the students to the family members and different companions. Sharing pictures inside online substance sharing sites, therefore, may rapidly lead to undesirable exposure and protection violations, Further, the determined way of online media makes it workable for different clients to gather rich totaled data about the proprietor of the distributed substance and the subjects in the distributed substance. The totaled data can bring about unforeseen introduction of one's social surroundings and lead to manhandle of one's close to home data.

Most substance sharing sites permit clients to enter their protection inclinations. Shockingly, late studies have demonstrated that clients battle to set up and keep up such protection settings. One of the primary reasons gave is that given the measure of shared data this procedure can be dreary and slip inclined. In this way, numerous have recognized the need of arrangement proposal frameworks which can help clients to effortlessly and appropriately design security settings. In any case, existing proposition for robotizing security settings give off an impression of being deficient to address the exceptional protection needs of pictures because of the measure of data certainly conveyed inside of pictures, and their association with the online environment wherein they are uncovered.

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:

1.1 The impact of social environment and personal characteristics

Social context of users, such as their profile information and relationships with others may provide useful information regarding users' privacy preferences. 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. For example, a privacy adverse person may be willing to share all his personal images while a more conservative person may just want to share personal images with his family members. In light of these considerations, it is important to find the balancing point between the impact of social environment and users' individual characteristics in order to predict the policies that match each individual's needs.

Moreover, individuals may change their overall attitude toward privacy as time passes. In order to develop a personalized policy recommendation system, such changes on privacy opinions should be carefully considered.

1.2 The role of image's content and metadata

In general, similar images often incur similar privacy preferences, especially when people appear in the images. For example, one may upload several photos of his kids and specify that only his family members are allowed to see these photos. He may upload some other photos of landscapes which he took as a hobby and for these photos, he may set privacy preference allowing anyone to view and comment the photos.

Analyzing the visual content may not be sufficient to capture users' privacy preferences. Tags and other metadata are indicative of the social context of the image, including where it was taken and why [4], and also provide a synthetic description of images, complementing the information obtained from visual content analysis.

.

2.LITERATURE SURVEY

2.1 Privacy Setting Configuration

Online social networks such as Friendster, MySpace, or the Facebook have experienced exponential growth in membership in recent years. These networks offer attractive means for interaction and communication, but also raise privacy and security concerns. In this study we survey a representative sample of the members of the Facebook (a social network for colleges and high schools) at a US academic institution, and compare the survey data to information retrieved from the network itself. We look for underlying demographic or behavioral differences between the communities of the network's members and non-members; we analyze the impact of privacy concerns on members' behavior; we compare members' stated attitudes with actual behavior and we document the changes in behavior subsequent to privacy-related information exposure. We find that an individual's privacy concerns are only a weak predictor of his membership to the network. Also privacy concerned individuals join the network and reveal great amounts of personal information. Some manage their privacy concerns by trusting their ability to control the information they provide and the external access to it. However, we also find evidence of members' misconceptions about the online community's actual size and composition, and about the visibility of members' profiles. [1]

As sharing personal media online becomes easier and widely spread, new privacy concerns emerge - especially when the persistent nature of the media and associated context reveals details about the physical and social context in which the media items were created. In a first-of-its-kind study, we use context-aware camera phone devices to examine privacy decisions in mobile and online photo sharing. Through data analysis on a corpus of privacy decisions and associated context data from a real-world system, we identify relationships between location of photo capture and photo privacy settings. Our data analysis leads to further questions which we investigate through a set of interviews with 15 users. The interviews reveal common themes in privacy considerations: security, social disclosure, identity and convenience. Finally, we highlight several implications and opportunities for design of media sharing applications, including using past privacy patterns to prevent oversights and errors. [2]

Why do people tag? Users have mostly avoided annotating media such as photos both in desktop and mobile environments despite the many potential uses for annotations, including recall and retrieval. We investigate the incentives for annotation in Flickr, a popular web-based photo-sharing system, and ZoneTag, a camera phone photo capture and annotation tool that uploads images to Flickr. In Flickr, annotation (as textual tags) serves both personal and social purposes, increasing incentives for tagging and resulting in a relatively high number of annotations. ZoneTag, in turn, makes it easier to tag camera phone photos that are uploaded to Flickr by allowing annotation and suggesting relevant tags immediately after capture. A qualitative study of ZoneTag/Flickr users exposed various

tagging patterns and emerging motivations for photo annotation. We offer a taxonomy of motivations for anno tation in this system along two dimensions (sociality and function), and explore the various factors that people consider when tagging their photos. Our findings suggest implications for the design of digital photo organization and sharing applications, as well as other applications that incorporate user-based annotation. [4]

Photo sharing has become a popular feature of many online social networking sites. Many of the photo sharing applications on these sites, allow users to annotate photos with those who are in them. A number of researchers have examined the social uses and privacy issues of online photo sharing sites, but few have explored the privacy issues of photo sharing in social networks. In this paper, we begin by examining some of our findings from a series of focus groups on photo privacy in the social networking domain. We then devise a new mechanism to enhance photo privacy based on these findings. [5]

Preventing adversaries from compiling significant amounts of user data is a major challenge for social network operators. We examine the difficulty of collecting profile and graph information from the popular social networking Website Facebook and report two major findings. First, we describe several novel ways in which data can be extracted by third parties. Second, we demonstrate the efficiency of these methods on crawled data. Our findings highlight how the current protection of personal data is inconsistent with user's expectations of privacy. [8]

2.2 Recommendation Systems

Our work is related to some existing recommendation systems which employ machine learning techniques. [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. [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, 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 ARCHITECTURE

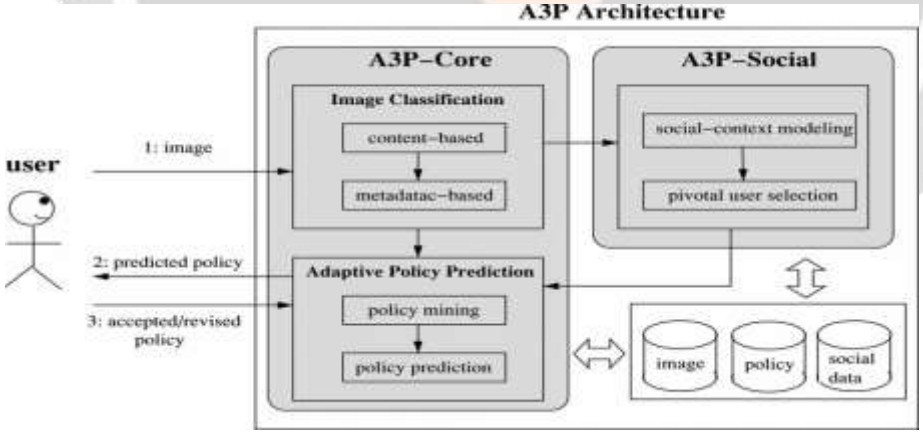

Fig -1 System Framework

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 A3P-social: (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. The A3P-social groups users into social communities with

similar social context and privacy preferences, and continuously monitors the social groups. When the A3P-social is invoked, it automatically identifies the social group for the user and sends back the information about the group to the A3P-core for policy prediction. At the end, the predicted policy will be displayed to the user. If the user is fully satisfied by the predicted policy, he or she can just accept it. Otherwise, the 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 future uploads.

4. A3P-CORE

There are two major segments in A3P-center: (i) Image classification and (ii) Adaptive policy prediction. For every client, his/her pictures are initially grouped in view of substance and metadata. At that point, security arrangements of every class of pictures are broke down for the approach expectation. Receiving a two-stage methodology is more suitable for arrangement suggestion than applying the basic one-stage information mining ways to deal with mine both picture components and strategies together. Review that when a client transfers another picture, the client is sitting tight for a prescribed arrangement. The two-stage methodology permits the framework to utilize the first stage to group the new picture and discover the applicant sets of pictures for the consequent strategy proposal. With respect to the one-stage mining methodology, it would not have the capacity to find the right class of the new picture in light of the fact that its characterization criteria need both picture components and arrangements though the approaches of the new picture are not accessible yet. Besides, consolidating both picture components and approaches into a solitary classifier would prompt a framework which is exceptionally subordinate to the particular sentence structure of the arrangement. On the off chance that an adjustment in the upheld approaches were to be presented, the entire learning model would need to change.

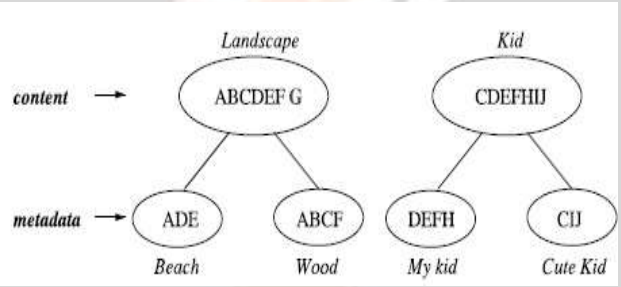

Fig 2. Image classification

4.1 Image Classification

To get groups of pictures that may be connected with comparative privacy preferences, we propose a progressive picture grouping which arranges pictures initially in view of their substance and afterward refine every classification into subcategories taking into account their metadata. Pictures that don't have metadata will be gathered just by substance. Such a various leveled grouping gives a higher need to picture content and minimizes the impact of missing labels. Note that it is conceivable that a few pictures are incorporated into various classifications the length of they contain the run of the mill substance elements or metadata of those classes. The substance based characterization makes two classifications: "scene" and "child". Pictures C, D, E and F are incorporated into both classes as they show children playing open air which fulfill the two topics: "scene" and "child". These two classifications are further separated into subcategories in view of labels connected with the pictures. Subsequently, we get two subcategories under every topic separately. Notice that picture G is not appeared in any subcategory as it doesn't have any label; picture an appears in both subcategories on the grounds that it has labels demonstrating both "shoreline" 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.

4.1.2 Metadata-Based 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 second step is to derive a representative hypernym (denoted as h) from each metadata vector. The third

step is to find a subcategory that an image belongs to. This is an incremental procedure. At the beginning, the first image forms a subcategory as itself and the representative hypernyms of the image becomes the subcategory's representative hypernyms.

4.2 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.

4.2.1 Policy Mining

We propose a various leveled digging methodology for arrangement mining. Our methodology influences affiliation guideline mining strategies to find well known examples in arrangements. Arrangement mining is done inside of the same class of the new picture in light of the fact that pictures in the same classification are more probable under the comparable level of security assurance. The essential thought of the progressive mining is to take after a characteristic request in which a client characterizes a strategy. Given a picture, a client typically first chooses who can get to the picture, then contemplates what particular access rights (e.g. See just or download) ought to be given, lastly refine the entrance conditions, for example, setting the lapse date. Correspondingly, the progressive digging first search for well-known subjects characterized by the client, then search for famous activities in the approaches containing the prominent subjects, lastly for prevalent conditions in the arrangements containing both mainstream subjects and conditions.

4.2.2 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.

5. A3P-SOCIAL

The A3P-social utilizes a multi-criteria inference mechanism that produces agent arrangements key data identified with the client's social setting and his general disposition toward security. As specified prior, A3Psocial will be summoned by the A3P-center in two situations. One is the point at which the client is an amateur of a site, and does not have enough pictures put away for the A3P-center to deduce significant and redid approaches. The other is the point at which the framework sees noteworthy changes of protection pattern in the client's social circle, which may be of enthusiasm for the client to potentially conform his/her security settings in like manner. In what tails, we first present the sorts of social setting considered by A3P-Social, and after that present the arrangement proposal 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 of one's privacy settings. The second step is to group users based on the identified factors.

5.2 Identifying Social Group

We now introduce the policy recommendation process based on the social groups obtained from the previous step. Suppose that a user Formula uploaded a new image and the A3P-core invoked the A3P-social for policy recommendation. The A3P-social will find the social group which is most similar to user Formula and then choose the representative user in the social group along with his images to be sent to the A3P-Core policy prediction module to generate the recommended policy for user Formula. Given that the number of users in social network may be huge and that users may join a large number of social groups; it would be very time consuming to compare the new user's social context attributes against the frequent pattern of each social group. In order to speed up the group identification process and ensure reasonable response time, we leverage the inverted file structure to organize the social group information. The inverted file maps keywords (values of social context attribute) occurring in the frequent patterns to the social groups that contain the keywords. Specifically, we first sort the keywords (except the

social connection) in the frequent patterns in an alphabetical order. Each keyword is associated with a link list which stores social group ID and pointers to the detailed information of the social group.

6. DECISION VOTING SYSTEM

This facilitates the privacy policy recommendation at individual level as well. If any exclusion at individual level is taken, then that is considered for further policy prediction. This helps to provide more meaningful prediction. Here DV is decision voting value and Evaluation (p) represents the policy p decision.

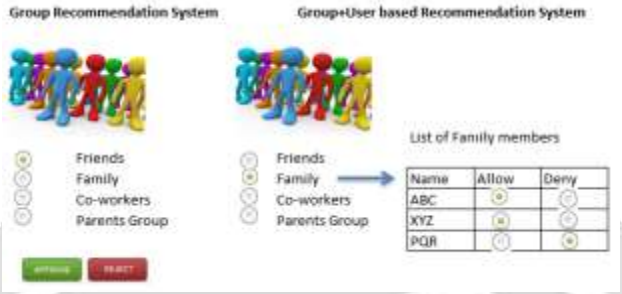

Fig. 3 Decision Voting Mechanism

7. MATHEMATICAL MODEL

Let S is the Whole System Consist of S = {I, P, O} I = Input. I = {U, Q, D, IMG} U = User U = {u1, u2 …. un} Q = Query Entered by user Q = {q1, q2, q3…qn} D = Dataset. IMG = Images IMG = {img1, img2....img n} P = Process: P= {PPR-CORE,PPR Social,CBC,MBC,APP,PM,PP,SCM,PUS} CBC = Content-Based Classification MBC = Metadata-Based Classification APP = Adaptive Policy Prediction PM= Policy Mining PP=Policy Prediction SCM= Social Context Modelling PUS=Pivotal User Selection	[Step1:] User enters the Query(Image). [Step2:] Privacy Policy Recommendation Primary (Classification and policy prediction) [Step3:] Content Based Classification. [Step4:] Metadata Based Classification. [Step5:] Policy mining [Step6:] Policy prediction [Step7:] Social Context modelling. [Step8:] Pivotal user selection.

8. CONCLUSION

Privacy Policy Recommendation enables users to automate privacy policies for images that users upload on content sharing sites. This system gives a comprehensive structure to infer privacy preferences based on historical information available for the users. This system handles the cold-start issue by utilizing the social context information. Existing system provides the recommendation to social groups like friends, family, co-workers, etc. Whereas the proposed system with Decision Voting scheme facilitates privacy recommendation for individual users. This works on conflict resolution as well. Also, to this, we are encrypting images while saving to ensure security to contents of the images. As a future scope, we can integrate the existing system with business intelligence and data warehousing solution which can provide strategic as well as operational analysis for further refinement of privacy policies or strategies.

9. 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] Ricardo da Silva Torres Alexandre Xavier Falcão , "Content-based image retrieval: Theory and applications," Ricardo da Silva Torres Alexandre Xavier Falcão , vol. 2, no. 13, pp. 161–185, 2006.13.

[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.