International Journal of Computer Sciences
January '2017 Issue
ISSN Number: 4142 3453

ADAPTIVE PRIVACY POLICY PREDICTION FOR USER UPLOADED IMAGES ON CONTENT SHARING SITES

J.Mamasa[1] , A.K.Puneeth Kumar[2]

[1]M.Tech(CSE) P.G Scholar, Dept. of CSE, Siddartha Educational Academy Group of Institutions, C. gollapalle, Tirupathi,Ap.

[2]Professor, Dept of CSE, Siddartha Educational Academy Group of Institutions, C. gollapalle, Tirupathi,Ap.

ABSTRACT:

With the increasing volume of images users share through social sites, maintaining privacy has become a major problem. In light of these incidents, the need of tools to help users control access to their shared content is highly essential. Toward addressing this need, we propose an Adaptive Privacy Policy Prediction (A3P) system to help users compose privacy settings for their images. Social Network is an emerging E-service for content sharing sites (CSS). This proposes a privacy policy prediction and access restrictions along with blocking scheme for social sites using data mining techniques. To perform this, the system utilizes APP (Access Policy Prediction) and Access control mechanism. The solution relies on an image classification framework for image categories which may be associated with similar policies and on a policy prediction Mechanism to automatically generate a policy for each newly uploaded image, also according to user's social features. Image Sharing takes place both among previously established groups of known people or social circles 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, Sharing images within online content sharing sites, therefore, may quickly lead to unwanted disclosure. The aggregated information can result in unexpected exposure of one's social environment and lead to abuse of one's personal information

Index Terms- Adaptive Privacy Policy Prediction (A3P), A3P- Core, A3P- Social.

.

1. INTRODUCTION

An A3P system helps users automates 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. A3P-core: (I) Image classification and (ii) Adaptive policy prediction. User images are first classified based on content and metadata. Privacy policies of each category of images are analyzed for the policy prediction. A3P-social multi-criteria inference mechanism that generates representative policies by leveraging key information related to the user's social context. Images searching for content based and image based the result found for each image privacy policy set of user privacy in sharing site. Contentbased classification is based on an efficient and yet accurate image similarity approach. Classification mechanism compares image signatures defined based on quantified and sanitized version of transformation. The Image encodes frequency and spatial information related to image color, size, and texture. The small number of coefficients is selected to form the signature of the image.

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.

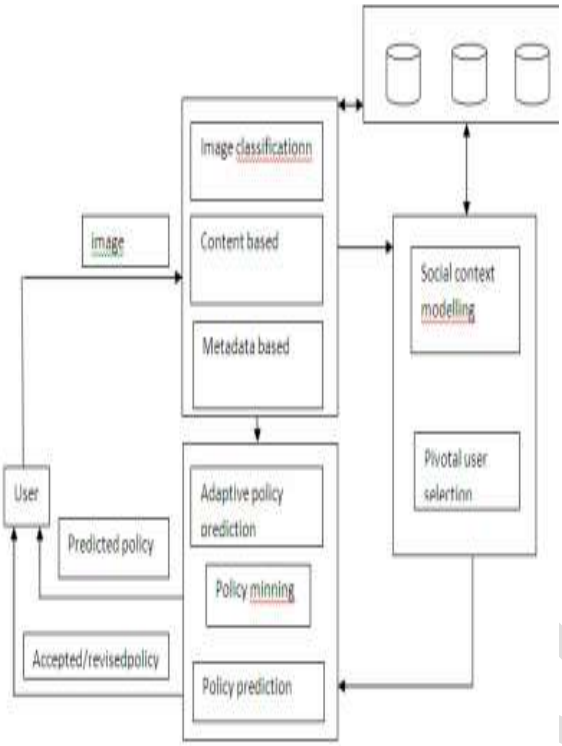
II. RELATED WORK

Content-based retrieval is ultimately dependent on the features used for the annotation of data and its efficiency is dependent on the invariance and robust properties. The Polar Fourier Transform (PFT) is similar to the Discrete Fourier Transform in two dimensions but uses transform parameters radius and angle rather than the Cartesian co-ordinates. To improve implications for content based retrieval of natural images where there will be a significantly higher number of textures [6]Local radial symmetry is to identify regions of interest within a scene. A facial feature detector and as a generic region of interest detector the new transform is seen to offer equal or superior performance to contemporary techniques. The method has been demonstrated on a series of face images and other scenes, and compared against a number of contemporary techniques from the literature. Equal or superior performance on the images tested while offering significant savings in both the computation required and the complexity of the implementation. [5]The refining process is formulated as an optimization framework based on the consistency between "visual similarity" and "semantic similarity" in social images. An image retagging scheme that aims at improving the quality of the tags associated with social images in terms of content relevance.

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

Privacy Setting Configuration

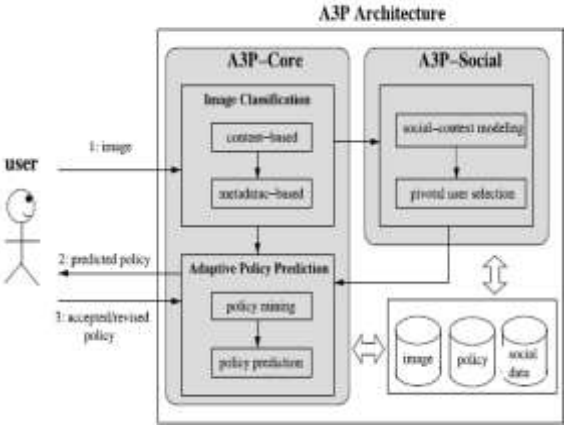
Several recent works have studied how to automate the task of privacy settings (e.g. [7], [16]). Bonneau et al. [7] proposed the concept of privacy suites which recommend to users a suite of privacy settings that "expert" users or other trusted friends have already set, so that normal users can either directly choose a setting or only need to do minor modification. Similarly, Danezis [8] proposed a machine-learning based approach to automatically extract privacysettings from the social context within which the data is produced. Parallel to the work of Danezis, Adu-Oppong et al. [16] 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. 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 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 mechanism. In addition, there is a large body of work on image content analysis, for classification and interpretation (e.g., [9], [15]), retrieval ([13], [14] are some examples), and photo ranking, also in the context of online photo sharing sites, such as Flickr [11] Of these works, Zerr's work is probably the closest to ours. Zerr explores privacy aware image classification using a mixed set of features, both content and meta-data. This is however a binary classification (private vs. public), so the classification task is very different than ours. Also, the authors do not deal with the issue of coldstart problem.

System Architecture

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 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.



III.PROPOSED SYSTEM

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, A3Pcore 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. 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

IV.RESULTS

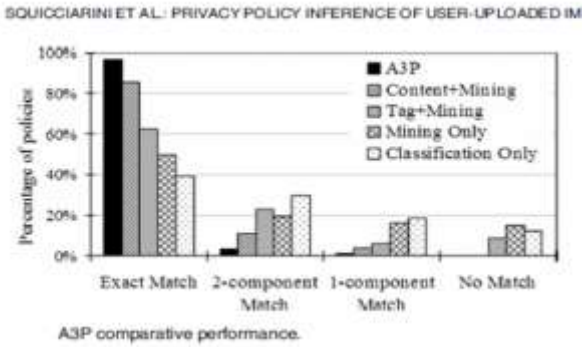
We observe that users with similar background tend to have similar privacy concerns, as seen in previous research studies and also confirmed by our collected data. This observation inspires us to develop a social context modeling mechanism that can capture the common social elements of users and identify 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 mechanism 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. 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. The distribution of contacts may shed light on the user's behavior of privacy settings. We assume that users who mainly share images among family members may not want to disclose personal information publicly, while users having a large group of friends may be willing to share more images with a larger audience [19]. Formally, we model the ratio of each type of relationship among all contacts of a user as social connection. Let R1, ..., Rn denote the n types of relationships observed among all users. Let NuRi denote the number of user U's contacts belonging to relationship type Ri. The connection distribution (denoted as Conn) is represented as below: For example, suppose that there are four types of relationships being used by users in the system: R1="family", R2="colleague", R3="friend", R4="others". Bob has 20 contacts, among which he has 10 family members, 5 colleagues, and 5 friends. His social connection is represented as { 10/20 , 5/ 20 , 5/ 20 , 0/ 20}. It is worth noting that, the number of social context attributes may grow when more rich information is collected by social networking sites in the future, and our mechanism is dynamic and capable of dealing with any number of attributes being considered. The second step is to identify groups of users who have similar social context and privacy preference. Regarding social context, it rarely happens that users share the same values of all social context attributes. More common cases are that a group of users have common values for a subset of social context attributes. Such subset can be different for different groups of users, which makes the user grouping a challenging task. We illustrate the scenario using the following example. For simplicity of illustration, we take a smaller set of attributes to be considered.

In this first round of tests, we used the two datasets collected through our survey to evaluate the accuracy of our recommended policies. 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 A3P-core 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 mechanism, denoted as "Content+Mining". A3P Social In the second round of experiments, we analyze the performance of the A3P-Social component using the first set of data collection. For each user, we use he A3P-Social to predict policies and compare it with three other alternative approaches:(i) prediction based on only similarity of privacy strictness levels; (ii) prediction based on Cosine similarity; (iii) prediction based on Pearson similarity. In particular, the first base-line approach does not consider social contexts but bases recommendation only on social groups that have similar privacy strictness level for same type of images. The second approach adopts Cosine similarity to measure the similarity of the social contexts between the new user and all the existing users, and then finds the top two users with the highest similarity score as the candidate users. The images of the candidate users are then sent to the A3P-core for the policy prediction. The approach using the Pearson similarity requires an additional assumption that the new user should have already provided privacy

preferences (levels) for several image categories other than the one waiting for the recommendation. These user specified privacy preferences are then treated as the "rating" in the Pearson similarity formula. The data we use for this assumption is the response to three privacy-related questions users provide on their pre-session survey during data collection (the questions are adapted from the wellknown privacy-index measures from Westin). Accordingly, we use the

Pearson similarity to find the candidate users who are similar to this new user. The experimental results show that the policy prediction accuracy (full matching) of our A3P-Social and the other three approaches are: A3P-Social(88.6%), Strictness level similarity (86.4%), Cosine similarity(82.5%) and Pearson similarity (81.4%)..More importantly, the A3Psocial is the most general approach and most efficient among the all. Although the prediction accuracy yielded by the approach using strictness level similarity is quite close to the A3P-social, it requires the new user to provide preferred privacy level because it needs this information to look for existing users with similar strictness levels. The same assumption is required by the approach using the Pearson similarity too. The A3P-social instead also works when the new user has no idea about what privacy level is appropriate. The A3P-social considers social contexts thus can take care of such new users who did not provide preferred privacy level and need some guidance on their initial privacy settings. Moreover, in terms of efficiency, all the three comparison approaches need to scan all the existing users whereas the A3P-Social just needs to check a subset of users attributed to the use of the inverted index.

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. shows the ercentage 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



A3P comparative performance.

VI.CONCLUSION AND FUTURE

This paper has 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 Privacy Enhancing Technologies Workshop, 2006.

[2] R. Agrawal and R. Srikant. Fast mechanisms for mining association rules in large databases. In J. B. Bocca, M. Jarke, and C. Zaniolo, editors, 20th International Conference on Very Large Data Bases, September 12-15, pages 487–499. Morgan Kaufmann, 1994.

[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 Conference on Human factors in computing systems, pages 357–366. ACM, 2007.

[4] M. Ames and M. Naaman. Why we tag: motivations for annotation in mobile and online media. In Conference on Human factors in computing systems, CHI' 07, pages 971– 980. ACM, 2007.

[5] A. Besmer and H. Lipford. Tagged photos: concerns, perceptions, and protections. In CHI '09: 27th international conference extended abstracts on Human factors in computing systems, pages 4585–4590. ACM, 2009.

[6] A. D. Bland JM. Multiple significance tests: the bonferroni method. BMJ, 310(6973), 1995.
[7] J. Bonneau, J. Anderson, and L. Church. Privacy suites:shared privacy for social networks. In Symposium on Usable Privacy and Security, 2009.
[8] J. Bonneau, J. Anderson, and G. Danezis. Prying data out of a social network. In ASONAM: International Conference on Advances in Social Network Analysis and Mining, pages 249– 254, 2009.
[9] O. Chapelle, P. Haffner, and V. Vapnik. Support vector machines for histogram-based image classification. Neural Networks, IEEE Transactions on, 10(5):1055–1064, 1999.
[10] 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 16th ACM international conference on multimedia, pages 737–740. ACM, 2008.
[11] 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 2009 IEEE International Conference on Multimedia and Expo, ICME 2009, pages 1238–1241. IEEE, 2009.
[12] L. Church, J. Anderson, J. Bonneau, and F. Stajano. Privacy stories: Confidence on privacy behaviors through end user programming. In Symposium on Usable Privacy and Security (SOUPS), 2009.
[13] R. da Silva Torres and A. Falc˜ao. Content-based image retrieval: Theory and applications. Revista de Inform´atica Te´orica e Aplicada, 2(13):161–185, 2006.
[14] R. Datta, D. Joshi, J. Li, and J. Wang. Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys (CSUR), 40(2):5, 2008.
[15] J. Deng, A. C. Berg, K. Li, and L. Fei-Fei. What does classifying more than 10,000 image categories tell us? In 11th European conference on Computer vision: Part V, ECCV'10, pages 71–84, Berlin, Heidelberg, 2010. Springer-Verlag.
[16] A. K. Fabeah Adu-Oppong, Casey Gardiner and P. Tsang. Social circles: Tackling privacy in social networks. In Symposium On Usable Privacy and Security, 2008.
[17] L. Geng and H. J. Hamilton. Interestingness measures for data mining: A survey. ACM Comput. Surv., 38(3):9, 2006.
[18] Image-net Dataset. www.image-net.org.
[19] S. Jones and E. O'Neill. Contextual dynamics of groupbased sharing decisions. In Conference on Human Factors in Computing Systems, CHI '11, pages 1777–1786. ACM, 2011.
[20] A. Kaw and E. Kalu. Numerical methods with applications. 2010.