International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

An Improved Privacy Policy Inference over the Socially Shared Images

*A.MAMATHA

**M.SHIRISHA

*M.TECH student, Dept of CSE, VAAGDEVI COLLEGE OF ENGINEERING
**Assistant Professor, Dept of CSE, VAAGDEVI COLLEGE OF ENGINEERING

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.

Index terms:Online information services,Web-based services.

1. INTRODUCTION

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 ofsocial 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 students2012 graduationceremony, for example. It could beshared within a Google+ circle or Flickr group, but mayunnecessarily expose the studentsBApos familymembersand other friends. Sharing images within online contentsharing sites,therefore,may quickly leadto unwanted disclosureand privacy violations,Further, the persistentnature of online media makes it possible for other users to collect rich aggregated information aboutthe owner of the published content and the subjects in the published content.The aggregated information can result in un expected 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

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

following criteria that influence one's privacy settings of images:

- 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.

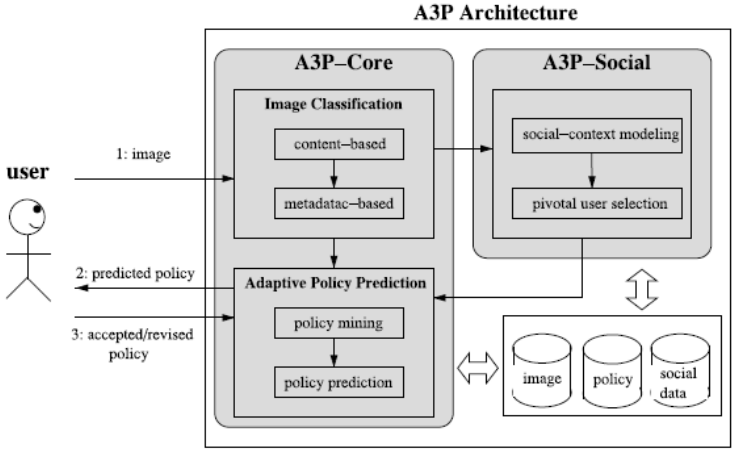

Fig.1. System Overview

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.

- 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.

Corresponding to the aforementioned two criteria, the proposed A3P system is comprised of two main building blocks (as shown in Fig. 1): A3P-Social and A3P-Core. The A3P-core focuses on analyzing each individual user's own images and metadata, while the A3P-Social offers a community perspective of privacy setting recommendations for a user's potential privacy improvement. We design the interaction flows between the two building blocks to balance the benefits from meeting personal characteristics and obtaining community advice.

To assess the practical value of our approach, we built a system prototype and performed an extensive experimental evaluation. We collected and tested over 5,500 real policies generated by more than 160 users. Our experimental results demonstrate both efficiency and high prediction accuracy of our system.

A preliminary discussion of the A3P-core was presented in [32]. In this work, we present an overhauled version of A3P, which includes an extended policy prediction algorithm in A3P-core (that is now parameterized based on user groups and also factors in possible outliers), and a new A3P-social module that develops the notion of social context to refine and extend the prediction power of our system. We also conduct additional experiments

with a new data set collecting over 1,400 images and corresponding policies, and we extend our analysis of the empirical results to unveil more insights of our system's performance.

2. LITERATURE SURVEY

1) A3p: Adaptive policy prediction for shared images over popular content sharing sites

AUTHORS: A. C. Squicciarini, S. Sundareswaran, D. Lin, and J. Wede

More and more people go online today and share their personal images using popular web services like Picasa. While enjoying the convenience brought by advanced technology, people also become aware of the privacy issues of data being shared. Recent studies have highlighted that people expect more tools to allow them to regain control over their privacy. In this work, we propose an Adaptive Privacy Policy Prediction (A3P) system to help users compose privacy settings for their images. In particular, we examine the role of image content and metadata as possible indicators of users' privacy preferences. We propose a two-level image classification framework to obtain image categories which may be associated with similar policies. Then, we develop a policy prediction algorithm toautomatically generate a policy for each newly uploaded image. Most importantly, the generated policy will follow the trend of the user's privacy concerns evolved with time. We have conducted an extensive user study and the results demonstrate effectiveness of our system with the prediction accuracy around 90%.

2) Non-parametric kernel ranking approach for social image retrieval

AUTHORS: J. Zhuang and S. C. H. Hoi

Social image retrieval has become an emerging research challenge in web rich media search. In this paper, we address the research problem of text-based social image retrieval, which aims to identify and return a set of relevant social images that are related to a text-based query from a corpus of social images. Regular approaches for social image retrieval simply adopt typical text-based image retrieval techniques to search for the relevant social images based on the associated tags, which may suffer from noisy tags. In this paper, we present a novel

framework for social image re-ranking based on a non-parametric kernel learning technique, which explores both textual and visual contents of social images for improving the ranking performance in social image retrieval tasks. Unlike existing methods that often adopt some fixed parametric kernel function, our framework learns a non-parametric kernel matrix that can effectively encode the information from both visual and textual domains. Although the proposed learning scheme is transductive, we suggest some solution to handle unseen data by warping the non-parametric kernel space to some input kernel function. Encouraging experimental results on a real-world social image testbed exhibit the effectiveness of the proposed method.

3) Privacy-aware image classification and search

AUTHORS: S. Zerr, S. Siersdorfer, J. Hare, and E. Demidova

Modern content sharing environments such as Flickr or YouTube contain a large amount of private resources such as photos showing weddings, family holidays, and private parties. These resources can be of a highly sensitive nature, disclosing many details of the users' private sphere. In order to support users in making privacy decisions in the context of image sharing and to provide them with a better overview on privacy related visual content available on the Web, we propose techniques to automatically detect private images, and to enable privacy-oriented image search. To this end, we learn privacy classifiers trained on a large set of manually assessed Flickr photos, combining textual metadata of images with a variety of visual features. We employ the resulting classification models for specifically searching for private photos, and for diversifying query results to provide users with a better coverage of private and public content. Large-scale classification experiments reveal insights into the predictive performance of different visual and textual features, and a user evaluation of query result rankings demonstrates the viability of our approach.

4) Personalized photograph ranking and selection system

AUTHORS: C.-H. Yeh, Y.-C. Ho, B. A. Barsky, and M. Ouhyoung

In this paper, we propose a novel personalized ranking system for amateur photographs. Although some of the features used in our system are similar to previous work, new features, such as texture, RGB color, portrait (through face detection), and black-and-white, are included for individual preferences. Our goal of automatically ranking photographs is not intended for award-wining professional photographs but for photographs taken by amateurs, especially when individual preference is taken into account. The performance of our system in terms of precision-recall diagram and binary classification accuracy (93%) is close to the best results to date for both overall system and individual features. Two personalized ranking user interfaces are provided: one is feature-based and the other is example-based. Although both interfaces are effective in providing personalized preferences, our user study showed that example-based was preferred by twice as many people as feature-based.

5) Strategies and struggles with privacy in an online social networking community

AUTHORS: K. Strater and H. Lipford

Online social networking communities such as Facebook and MySpace are extremely popular. These sites have changed how many people develop and maintain relationships through posting and sharing personal information. The amount and depth of these personal disclosures have raised concerns regarding online privacy. We expand upon previous research on users' under-utilization of available privacy options by examining users' current strategies for maintaining their privacy, and where those strategies fail, on the online social network site Facebook. Our results demonstrate the need for mechanisms that provide awareness of the privacy impact of users' daily interactions.

3.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. As
for the one-stage mining approach, it would not be able locate the right class of the new image because its classification criteria needs both image features and policies whereas the policies of the new image are not available yet. Moreover,
combining both image features and policies into a single classifier would lead to a system which is very dependent to the specific syntax of the policy. If a change in
the supported policies were to be introduced, the whole learning model would need to change.

3.1 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 subcategoriesbased 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.

3.1.1 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 [26]), and SIFT [25].

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 [17]. 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 ifthe 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.

3.1.2 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 metadata considered in our work are tags, captions, and comments.

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. Then, we compute the distance between representative hypernyms of a new incoming image and each existing subcategory.

3.2 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 predictionprocess consists of three main phases: (i) policy normalization; (ii) policy mining; and (iii) policy prediction.

The policy normalization is a simple decomposition process to convert a user policy into a set of atomic rules in which the data (D) component is a single-element set.

3.2.1 Policy Mining

We propose a hierarchical mining approach for policy mining. Our approach leverages association rule mining techniques to discover popular patterns in policies. Policy mining is carried out within the same category of the new image because images in the same category are more likely under the similar level of privacy protection. The basic idea of the hierarchical mining is to follow a natural order in which a user defines a policy. Given an image, a user usually first decides who can access the image, then thinks about what specific access rights (e.g., view only or download)

should be given, and finally refine the access conditions such as setting the expiration date. Correspondingly, the hierarchical mining first lookfor popular subjects defined by the user, then look for popular actions in the policies containing the popular subjects, and finally for popular conditions in the policies containing both popular subjects and conditions.

3.2.2 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.

To model the user's privacy tendency, we define a notion of strictness level. The strictness level is a quantitative metric that describes how "strict" a policy is. In particular, a strictness level L is an integer with minimum value in zero, wherein the lower the value, the higher the strictness level. It is generated by two metrics: major level (denoted as l) and coverage rate (a), where l is determined by the combination of subject and action in a policy, and a is determined by the system using the condition component. l is obtained via Table 4. In Table 4, all combinations of common subject and common actions are enumerated and assigned an integer value according to the strictness of the corresponding subjects and actions. For example, "view" action is considered more restricted than "tag"
action. Given a policy, its l value can be looked up from the table by matching its subject and action. If the policy has multiple subjects or actions and results in multiple l values, we will consider the lowest one. It is worth noting
that the table is automatically generated by the system but can be modified by users according to their needs.

Then, we introduce the computation of the coverage rate a which is designed to provide fine-grained strictness level. a is a value ranging from 0 to 1 and it will just adjust but not dominate the previously obtained major level. In particular, we define a as the percentage of people in the specified subject category who satisfy the condition in the policy.

4 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 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.

4.1 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.

4.2 Identifying Social Group

We now introduce the policy recommendation process based on the social groups obtained from the previous step.

Suppose that a user U 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 U 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 U. Given that the number 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 [31] 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 information.

5 .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 provesthat 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 Informatica Teorica 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.

[18] S. Jones and E. O'Neill, "Contextual dynamics of group-based sharing decisions," in Proc. Conf. Human Factors Comput. Syst., 2011, pp. 1777–1786. [Online]. Available: http://doi.acm.org/10.1145/1978942.1979200

[19] A. Kaw and E. Kalu, Numerical Methods with Applications: Abridged., Raleigh, North Carolina, USA: Lulu.com, 2010.

[20] P. Klemperer, Y. Liang, M. Mazurek, M. Sleeper, B. Ur, L. Bauer, L. F. Cranor, N. Gupta, and M. Reiter, "Tag, you can see it!: Using tags for access control in photo sharing," in Proc. ACM Annu. Conf. Human Factors Comput. Syst., 2012, pp. 377–386.

AUTHOR 1 :-

* A. Mamatha completed her B tech in Balaji Institute of Technology & Sciences in 2014 and pursuing M-Tech in Vaagdevi College of Engineering

AUTHOR 2:-

**M. Shirisha is working as Assistant Professor in Dept of CSE, Vaagdevi College of Engineering