Privacy Policy Inference of User-Uploaded Images on Content Sharing Sites

R. MANOHAR REDDY
MANOHARREDDY140@GMAIL.COM
DEPARTMENT OF CSE SRI KRISHNADEVARAYA ENGINEERING COLLEGE.
GOOTY, AP. INDIA

S. G NAWAZ
SNGNAWAZ@GMAIL.COM
ASSOCIATE PROFESSOR
DEPARTMENT OF CSE
SRI KRISHNADEVARAYA ENGINEERING COLLEGE.
GOOTY, AP. INDIA

DR. R. RAMACHANDRA
PRINCIPAL OF
SRI KRISHNADEVARAYA ENGINEERING COLLEGE.
GOOTY, AP. 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 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.

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 of social discovery-to help them identify new peers and learn about peers interests and social surroundings. However, semantically rich images may reveal content sensitive information []. Consider a photo of a students 2012 graduation ceremony, for example. It could be shared within a Google+ circle or Flickr group, but may unnecessarily expose the studentsBApos family members 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 [3], [20], [24]. 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 [1], [11], [22], [33]. 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 [7], [22], [28], [30]. However, existing proposals for automating privacy settings appear to be inadequate to address the unique privacy needs of images [3], [5], [41], 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 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.

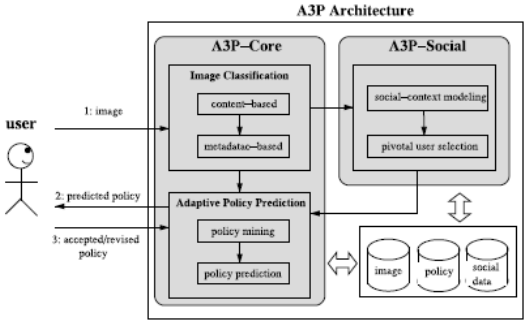

Fig. 1. System overview.

Literature Survey

#1 Over-Exposed? Privacy Patterns and Considerations in Online and Mobile Photo Sharing

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 camerephone 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. Privacy Suites: Shared Privacy for Social Networks

Creating privacy controls for social networks that are both expressive and usable is a major challenge. Lack of user un- derstanding of privacy settings can lead to unwanted disclosure of private information and, in some cases, to material harm. We propose a new paradigm which allows users to easily choose \suites" of privacy settings which have been speci_ed by friends or trusted experts, only modifying them if they wish. Given that most users currently stick with their default, operator-chosen settings, such a system could dramatically increase the privacy protection that most users experience with minimal time investment.

#3 SheepDog – Group and Tag Recommendation for Flickr Photos by Automatic Search-based Learning

Online photo albums have been prevalent in recent years and have resulted in more and more applications developed to provide convenient functionalities for photo sharing. In this paper, we propose a system named *SheepDog* to automatically add photos into appropriate groups and recommend suitable tags for users on Flickr. We adopt concept detection to predict relevant concepts of a photo and probe into the issue about training data collection for concept classification. From the perspective of gathering training data by web searching, we introduce two mechanisms and investigate their performances of concept detection. Based on some existing information from Flickr, a ranking-based method is applied not only to obtain reliable training data, but also to provide reasonable group/tag recommendations for input photos. We evaluate this system with a rich set of photos and the results demonstrate the effectiveness of our work.

SYSTEM STUDY

2.1 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

MODULES:

- System Construction Module
- Content-Based Classification
- Metadata-Based Classification
- Adaptive Policy Prediction

MODULES DESCSRIPTION:

System Construction Module

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).

Content-Based 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.

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.

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.

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.

SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the

Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

TYPES OF TESTS

Unit testing
Integration testing
Black Box Testing
System Test
White Box Testing
Functional test

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. 16^{th} 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