PRIVACY POLICY INFERENCE OF USER-UPLOADED IMAGES ON CONTENT SHARING SITES

A.Sisindra Raj[1], N. Swapna[2], Dr. G. Vishnu Murthy[3]

[1]*M.Tech, Dept of CNIS, Anurag Group of Institutions*

[2]*Associate Professor, Dept of CS, Anurag Group of Institutions*

[3]*Professor and HOD, Dept of CS, Anurag Group of Institutions*

ABSTRACT:

Images which are shared by users through social sites has been increasing drastically, 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.

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 student's 2012 graduation ceremony, for example. It could be shared within a Google+ circle or Flickr group, but may unnecessarily expose the student's, family members and other friends. Sharing images within online content sharing sites, therefore, may quickly lead to unwanted disclosure and privacy violations. Further, the persistent nature of online media makes it possible for other users to collect rich aggregated information about the owner of the published content and the subjects in the published content. The aggregated information can result in unexpected exposure of one's social environment and lead to abuse of one's personal information.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 setting. 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. 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. 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.

LITERATURE SURVEY:

Users' Awareness of Privacy on Online Social Networking sites

Online social networking offers a new, easy and inexpensive way to maintain already existing relationships and present oneself to others. However, the increasing number of actions in online services also gives a rise to privacy concerns and risks. In an attempt to understand the factors, especially privacy awareness, that influence users to disclose or protect information in online environment, we view privacy behavior from the perspectives of privacy protection and information disclosing.

We have reviewed earlier research on privacy issues related to social networking sites, and presented the results of our empirical study among users of a particular SNS, Facebook. We have viewed privacy behavior from two perspectives: privacy protection and information disclosing. Both of these aspects were analysed and used in attempt to understand the factors, especially privacy awareness, that influence users to disclose or protect information on Facebook.

In our empirical study, we surveyed users of Facebook, and acquired 210 usable responses. Our results indicate, that most of respondents, who seem to be active users of Facebook, do disclose a considerable amount of private information of themselves, and contrary to their own belief, are not too well aware of the visibility of their information to people they do not necessarily know. Furthermore, the privacy policy and terms of use of Facebook were largely not known or understood by our respondents. This was particularly true as regard to Facebook's policy of allowing third party application providers access to the users‟ information. Encouragingly, however, many of the respondents were awakened by the survey, and resolved to pay more attention to their privacy settings in the future.

As the whole online environment and social networks in particular are fairly new phenomena, a number of issues are not fully understood by the users, who might even appear to behave irrationally. Privacy is a complex construct and, as such, difficult to understand. Accordingly, there are many different factors that affect privacy behavior. Hence, more research into privacy awareness and related behavior on social networking sites.

Fast Algorithms for Mining Association Rules

We consider the problem of discovering association rules between items in a large database of sales transactions. We present two new algorithms for solving this problem that are funda- mentally different from the known algorithms. Experiments with synthetic as well as real-life data show that these algorithms outperform the known algorithms by factors ranging from three for small problems to more than an order of magnitude for large problems. We also show how the best features of the two proposed algorithms can be combined into a hybrid algorithm, called AprioriHybrid. Scale-up experiments show that AprioriHybrid scales linearly with the number of transactions. AprioriHybrid also has excellent scale-up properties with respect to the transaction size and the number of items in the database.

We presented two new algorithms, Apriori and AprioriTid, for discovering all significant association rules between items in a large database of transactions. We compared these algorithms to the previously known algorithms, the AIS [AIS93b] and SETM [HS93] algorithms. We presented experimental results, using both synthetic and real-life data, showing that the proposed algorithms always outperform AIS and SETM. The performance gap increased with the problem size, and ranged from a factor of three for small problems to more than an order of magnitude for large problems. We showed how the best features of the two proposed algorithms can be combined into a hybrid algorithm, called AprioriHybrid, which then becomes the algorithm of choice for this problem. Scale-up experiments showed that AprioriHybrid scales linearly with the number of transactions. In addition, the execution time decreases a little as the number of items in the database increases. As the average transaction size increases (while keeping the database size constant), the execution time increases only gradually. These experiments demonstrate the feasibility of using AprioriHybrid in real applications involving very large databases. The algorithms presented in this paper have been implemented on several data repositories, including the AIX le system, DB2/MVS, and DB2/6000.

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 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.

Issues of online privacy have long been of concern in the HCI community, and are of growing concern for the general public as an increasing amount of personal content is becoming available online. We have conducted a qualitative and quantitative analysis of privacy in a real-world photo-sharing mobile and online application. Using context-aware camera phones as capture devices allowed us to explore patterns of privacy in a way that was previously unavailable. Our findings, and design implications, are relevant to researchers and designers of content-sharing systems as well as mobile capture devices. Are users over-exposed? For now, it is a matter of taste; but while the potential for disaster exists, some users remain unconcerned. We are hoping to keep investigating the topic to get a more detailed look at patterns across a longer time period, and perhaps in different cultures.

Motivations for Image Publishing and Tagging On Flickr

Changes in photographic and internet technology have revolutionized the way people create, process and share digital images. This paper investigates people's motivations for image publishing and tagging on the web 2.0 site Flickr. Using an online pilot survey, 33 participants answered questions about their uploading and tagging practices, and whether or not they hope to make a commercial gain from their images. The results show that most people have two main motivational reasons both for using Flickr, and for the tagging of their images. However, whilst a person may be motivated to use Flickr for both personal and social reasons, tagging motivation tends to focus more exclusively on either one or the other of these two factors. Overall it was found that social organization and social communication are the most popular motivational factors for both using Flickr and for tagging images, suggesting that Flickr is enjoyed for the community environment it provides rather than as a place to store images. However despite people's desire to share their images, most users are not hoping to make a commercial gain from the items they upload.

Whilst motivations for using Flickr and uploading images can be for a number of different reasons at the same time, motivations for tagging images tends to have a more predominant role. People may use Flickr as both a personal archive and as a place to share images with friends and family, but their reasons behind choice of tags will tend to be very distinctly either a 'self' or a 'social' action, with less hesitation in the mind of the tagger as to who will ultimately benefit from their choice of tag. People don't appear to want to use a mixture of highly personal and social tags; they will adopt one strategy or the other, regardless of if they are tagging for archive and storage or communicative purposes. However in support of much of the previous work carried out on Flickr, the respondents who took part in this investigation seem to use Flickr for the social aspects and the community environment which it provides with social organization and social communication being the two most popular motivational factors overall. Despite people's desire to have their images found and commented upon, as a general rule, people aren't interested in making a commercial gain from the images they upload – the community spirit of Flickr and its ability to connect people both known and unknown to the image uploaded is its most appealing feature. The responses from the pilot questionnaire have given a valuable first insight into why people publish and tag their images on Flickr, and also into the changing nature of self-publishing in the world of user-generated content.

Privacy Perceptions of Photo Sharing in Facebook

Online photo sharing applications are increasingly popular, offering users new and innovative ways to share photos with a variety of people. Many social networking sites are also incorporating photo sharing features, allowing users to very easily upload and post photos for their friends and families. For example, Facebook is the largest photo sharing site on the Internet with 14 million photos uploaded daily. Integrating photo sharing within social networking communities has also provided the opportunity for user-tagging, annotating and linking images to the identities of the people in them. This feature further increases the opportunities to share photos among people with established offline relationships and has been largely successful. However, this increased access to an individual's photos has led to these images being used for purposes that were not intended. For example, photos on Facebook profiles have been used by employers and law enforcement to investigate the behavior of individuals. We are focusing on these privacy concerns and needs, as well as exploring ideas for privacy protection mechanisms, for users of social networking sites such as Facebook. In understanding user's current concerns and behaviors, we can design tools they desire, adopt, and ones they will be motivated to use.

Photo sharing through online social networking sites are allowing huge numbers of people to upload and socially communicate around photos. However, users have lost control over their identity and disclosures as other users can upload and tag undesired photos. Additionally, users are struggling to manage their identity through the contents of photos across multiple audiences and the many people in their social networks. Users desire and need more tools to allow them to regain control over their privacy, and manage their privacy decisions over time. Users want these tools to respect everyone's rights and provide a fair playing ground for all. However, in being driven by impression management concerns, users may also not be motivated to alter their behavior or use tools that do not fit in with their existing activities or that protect against other potential threats such as strangers using photos to infer location or personal information. While Facebook does already have extensive privacy controls, users did express desire for more fine-grained controls over the accessibility of individual photos linked to them. While this study focused on Facebook in particular, other social networking sites are adding similar features. For example, MySpace recently added the ability to tag other users in photos. Thus, the concerns and issues we discovered will likely be applicable to other general social networking sites with photo sharing. As these sites continue to grow in popularity and users add more and more photos, meeting users privacy needs is important to allow safe and comfortable participation on these online communities.

Online Social Networks Privacy Threats and Defences

With over 1 billion users connected through online social networks, user privacy is becoming ever more important and is widely discussed in the media and researched in academia. In this chapter we provide a brief overview of some threats to users' privacy. We classify these threats as: users' limitations, design flaws and limitations, implicit flows of information, and clash of incentives. We also discuss two defense mechanisms which deploy usable privacy through a visual and interactive flow of information and a rational privacy vulnerability scanner.

In a nutshell, in this chapter we discussed four different causes of privacy leaks in online social networks. These four causes include: users' limitations; design flaws and limitations; implicit flow of information; and clash of interests. Then we discussed two defense mechanisms involving the visual and interactive control of flow information, and economic modeling of the privacy threat, thus providing users with the means to make a more rational choice. Interested readers may follow the given references for a detailed review of the provided threats and defenses.

Privacy Settings of User Data and Images on Content Sharing Sites

Social media's become one of the most important parts of our daily life as it enables us to communicate with a lot of people. Creation of social networking sites such as MySpace, LinkedIn, and Facebook, individuals are given opportunities to meet new people and friends in their own and also in the other diverse communities across the world. Users of social-networking services share an abundance of personal information with a large number of "friends." This improved technology leads to privacy violation where the users are sharing the large volumes of images across more number of peoples. This privacy need to be taken care in order to improve the user satisfaction level. The goal of this survey is to provide a comprehensive review of various privacy policy approaches to improve the security of information shared in the social media sites.

This paper describes various privacy policy techniques for user uploaded data and images in various content sharing sites. The privacy policy can be applied based on the user social behavior and the user uploaded image content. Table I presents the overview of various privacy policy techniques among the existing systems. Future research lead towards improving the performance by a novel semantic retrieval of images is done based on Hidden Markov model based annotated images.

Personalized Geo-Specific Tag Recommendation for Photos on Social

Now a day's social tagging is an important in social websites to provide a good tagging for photos uploaded to the websites to access high quality social tags. Tag recommendation by automatically assigning related tags to photos to find out particular interesting area. In this paper we concentrate on the personalized recommendation work and try to choose user preferred geo-location specific as well as relevant tags for photos on social website. For users and geo-locations we assume they have various preferred tags assigned to a photo and purpose a subspace learning method to individually uncover the user preference and geo-location preference. The goal of our work is to combine a visual and textual space into a unified subspace. According to unified subspace is mapped from the intermediate subspace and textual subspace respectively. We create formula for above learning problems into united form and present the minimization with its convergence rule. For a given an untagged photo with its geo-location to a user we used the nearest neighbor search in the relating unified space. The user preferred and geo-location specified tags. Experiments on big scaled data sets collected from flicker examine the effectively of the proposed system. In this work, we suggest to mine the personalized tag for newly updated photos using user profile based information such as their tagging histories, Geographic information, Geographic location information like the latitude and longitude values.

Modeling Social Strength in Social Media Community via Kernel-based Learning

Modeling continuous social strength rather than conventional binary social ties in the social network can lead to a more precise and informative description of social relationship among people. In this paper, we study the problem of social strength modeling (SSM) for the users in a social media community, who are typically associated with diverse form of data. In particular, we take Flickr—the most popular online photo sharing community—as an example, in which users are sharing their experiences through substantial amounts of multimodal contents (e.g., photos, tags, geo-locations, friend lists) and social behaviors (e.g., commenting and joining interest groups). Such heterogeneous data in Flickr bring opportunities yet challenges to the research community for SSM. One of the key issues in SSM is how to effectively explore the heterogeneous data and how to optimally combine them to measure the social strength. In this paper, we present a kernel-based learning to rank framework for inferring the social strength of Flickr users, which involves two learning stages. The first stage employs a kernel target alignment algorithm to integrate the heterogeneous data into a holistic similarity space. With the learned kernel, the second stage rectifies the pair-wise learning to rank approach to estimating the social strength. By learning the social strength graph, we are able to conduct collaborative recommendation and collective classification. The promising results show that the learning-based approach is effective for SSM. Despite being focused on Flickr, our technique can be applied to model social strength of users in any other social media community.

Objective of the Problem

Adaptive Privacy Policy Prediction (A3P) system that 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. It 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 improvement over current approaches to privacy.

Existing System

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.

Disadvantages

- Sharing images within online content sharing sites therefore, may quickly lead to unwanted disclosure and privacy violations.
- 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.
- Lead to abuse of one’s personal information

Proposed System

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.

The role of image’s content and metadata.

Advantages

- The A3P system provides a comprehensive framework to infer privacy preferences based on the information available for a given user.
- 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.

EXPERIMENTAL DESIGN OR METHODOLOGY

Modules

- System Construction Module
- Content-Based Classification
- Metadata-Based Classification

- Adaptive Policy Prediction

Module Description

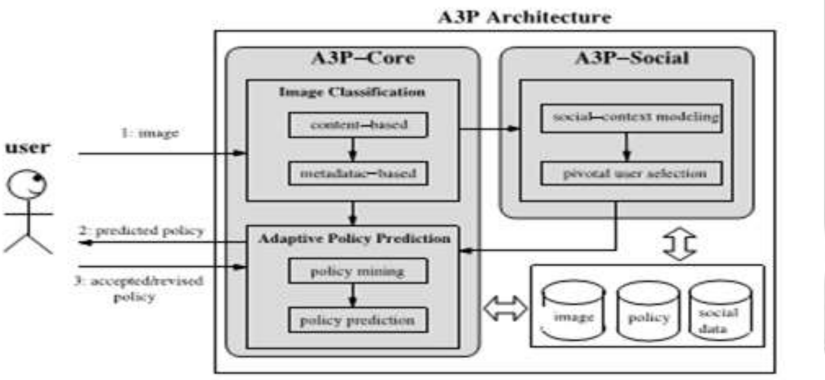
System Construction Module

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.



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 (iii) policy prediction.

ARCHITECTURE

The system architecture is the overall organization of the system components called subsystems. The architecture provides the context in which more detailed decisions are made in later design stages. System Architecture is a response to the conceptual and practical difficulties of the description and the design of complex systems.

CONCLUSION:

We have proposed an Adaptive Privacy Policy Prediction (A3P) system that helps users automate the privacy policy settings for their uploaded images. The A3P system provides a comprehensive framework to infer privacy preferences based on the information available for a given user. We also effectively tackled the issue of cold-start, leveraging social context information. Our experimental study proves that our A3P is a practical tool that offers significant improvements over current approaches to privacy.

REFERENCES

[1] A. Acquisti and R. Gross, “Imagined communities: Awareness, information sharing, and privacy on the Facebook,” in Proc. 6th Int. Conf. Privacy Enhancing Technol. Workshop, 2006, pp. 36–58.
[2] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules in large databases,” in Proc. 20th Int. Conf. Very Large Data Bases, 1994, pp. 487–499.
[3] S. Ahern, D. Eckles, N. S. Good, S. King, M. Naaman, and R. Nair, “Over-exposed? Privacy patterns and considerations in online and mobile photo sharing,” in Proc. Conf. Human Factors Comput. Syst., 2007, pp. 357–366.
[4] M. Ames and M. Naaman, “Why we tag: Motivations for annotation in mobile and online media,” in Proc. Conf. Human Factors Comput. Syst., 2007, pp. 971–980.
[5] A. Besmer and H. Lipford, “Tagged photos: Concerns, perceptions, and protections,” in Proc. 27th Int. Conf. Extended Abstracts Human Factors Comput. Syst., 2009, pp. 4585–4590.
[6] D. G. Altman and J. M. Bland, “Multiple significance tests: The bonferroni method,” Brit. Med. J., vol. 310, no. 6973, 1995.
[7] J. Bonneau, J. Anderson, and L. Church, “Privacy suites: Shared privacy for social networks,” in Proc. Symp. Usable Privacy Security, 2009.
[8] J. Bonneau, J. Anderson, and G. Danezis, “Prying data out of a social network,” in Proc. Int. Conf. Adv. Soc. Netw. Anal. Mining. 2009, pp.249–254.
[9] H.-M. Chen, M.-H. Chang, P.-C. Chang, M.-C. Tien, W. H. Hsu, and J.-L. Wu, “Sheepdog: Group and tag recommendation for flickr photos by automatic search-based learning,” in Proc. 16th ACM Int. Conf. Multimedia, 2008, pp. 737–740.
[10] M. D. Choudhury, H. Sundaram, Y.-R. Lin, A. John, and D. D. Seligmann, “Connecting content to community in social media via image content, user tags and user communication,” in Proc. IEEE Int. Conf. Multimedia Expo, 2009, pp.1238–1241.
[11] L. Church, J. Anderson, J. Bonneau, and F. Stajano, “Privacy stories: Confidence on privacy behaviors through end user programming,” in Proc. 5th Symp. Usable Privacy Security, 2009.
[12] R. da Silva Torres and A. Falc~ao, “Content-based image retrieval: Theory and applications,” Revista de Inform_atica Te_orica e Aplicada, vol. 2, no. 13, pp. 161–185, 2006.
[13] R. Datta, D. Joshi, J. Li, and J. Wang, “Image retrieval: Ideas, influences, and trends of the new age,” ACM Comput. Surv. vol. 40, no. 2, p. 5, 2008.
[14] J. Deng, A. C. Berg, K. Li, and L. Fei-Fei, “What does classifying more than 10,000 image categories tell us?” in Proc. 11th Eur. Conf. Comput. Vis.: Part V, 2010, pp. 71–84. [Online]. Available: http:// portal.acm.org/citation.cfm?id=1888150.1888157
[15] A. Kapadia, F. Adu-Oppong, C. K. Gardiner, and P. P. Tsang, “Social circles: Tackling privacy in social networks,” in Proc. Symp. Usable Privacy Security, 2008.