Journal of Applied and Advanced Research 2017, 2(3): 189–199
http://dx.doi.org/10.21839/jaar.2017.v2i3.93
http://www.phoenixpub.org/journals/index.php/jaar
ISSN 2519-9412 / © 2017 Phoenix Research Publishers

Research Article – Computer Science Engineering

Image sharing privacy policy on social networks using A3P

S. Saranya*, M. Ranjith Kumar, K. Madheswaran

Department of Computer Science and Engineering, SRG Engineering College, Aniyapuram, Namakkal – 637017, Tamil Nadu, India

Abstract

User Image sharing social site 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 an Adaptive Privacy Policy Prediction (A3P) system to help users compose privacy settings for their images. The 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 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.

Key words: Adaptive Privacy Policy Prediction, Image sharing, social networks

Introduction

Social media is the two way communication in Web 2.0 and it means to communicate, share, and interact with an individual or with a large audience. Social networking websites are the most famous websites on the Internet and millions of people use them every day to engage and connect with other people. Twitter, Facebook, LinkedIn and Google Plus seems to be the most popular Social networking websites on the Internet. Today, for every single piece of content shared on sites like Facebook - every wall post, photo, status update, and video - the up loader must decide which of his friends, group members, and other Facebook users should be able to access the content. As a result, the issue of privacy on sites like Facebook has received significant attention in both the research community and the mainstream media. My goal is to improve the set of privacy controls and defaults, but I limited by the fact that there has been no in- depth study of users' privacy settings on sites like Facebook. While significant privacy violations and mismatched user expectations are likely to exist, the extent to which such privacy violations occur has yet to be quantified.

In this paper, we propose an Adaptive Privacy Policy Prediction (A3P) system which aims to provide users a hassle free privacy settings experience by automatically generating personalized policies. The A3P system handles user uploaded images, and factors in the following criteria that influence one's privacy settings of images:

1. The impact of social environment and personal characteristics. Social context of users, such as their profile information and relationships with

Received: 20-05-2017; Accepted 18-06-2017; Published Online 19-06-2017

*Corresponding Author

S. Saranya, *Department of Computer Science and Engineering, SRG Engineering College, Aniyapuram, Namakkal – 637017, Tamil Nadu, India*

others may provide useful information regarding users' privacy preferences. For example, users interested in photography may like to share their photos with other amateur photographers. Users who have several family members among their social contacts may share with them pictures related to family events. However, using common policies across all users or across users with similar traits may be too simplistic and not satisfy individual preferences. Users may have drastically different opinions even on the same type of images. For example, a privacy adverse person may be willing to share all his personal images while a more conservative person may just want to share personal images with his family members. In light of these considerations, it is important to find the balancing point between the impact of social environment and users' individual characteristics in order to predict the policies that match each individual's needs. Moreover, individuals may change their overall attitude toward privacy as time passes. In order to develop a personalized policy recommendation system, such changes on privacy opinions should be carefully considered.

2. The role of image's content and metadata. In general, similar images often incur similar privacy preferences, especially when people appear in the images. For example, one may upload several photos of his kids and specify that only his family members are allowed to see these photos. He may upload some other photos of landscapes which he took as a hobby and for these photos, he may set privacy preference allowing anyone to view and comment the photos. Analyzing the visual content may not be sufficient to capture users' privacy preferences. Tags and other metadata are indicative of the social context of the image, including where it was taken and why, and also provide a synthetic description of images, complementing the information obtained from visual content analysis.

Literature Survey

Using Tags for Access Controlling Photo Sharing

Users often have rich and complex photo-sharing [19] preferences, but properly configuring access control can be difficult and time-consuming. In an 18-participant laboratory study, I explore whether the keywords and captions with which users tag their photos [8] can be used to help users more intuitively create and maintain access-control policies [18]. I find that (a) tags created for organizational purposes can be repurposed to create efficient and reasonably accurate access-control rules; (b) users tagging with access control in mind develop coherent strategies that lead to significantly more accurate rules than those associated with organizational tags alone; and (c) participants can understand and actively engage with the concept of tag-based access control.

Understanding Privacy Settings in Facebook with an Audience View

Users of online social networking communities are disclosing large amounts of personal information, putting themselves at a variety of risks. My ongoing research investigates mechanisms for socially appropriate privacy management in online social networking communities [2]. As a first step, I examine the role of interface usability in current privacy settings. In this paper I report on my first iterative prototype, where presenting an audience-oriented view of profile information significantly improved the understanding of privacy settings.

The PViz Comprehension Tool for Social Network Privacy Settings

Users' mental models of privacy and visibility in social networks [7] often involve subgroups within their local networks of friends. Many social networking sites have begun building interfaces to support grouping, like Facebook's lists and "Smart Lists," and Google+'s "Circles." However, existing policy comprehension tools, such as Facebook's Audience View, are not aligned with this mental model.

In this paper, I introduce PViz, an interface and system that corresponds more directly with how users model groups and privacy policies applied to their networks. PViz [12] allows the user to understand the visibility of her profile according to automatically-constructed, natural sub-groupings of friends, and at different levels of granularity. Because the user must be able to identify and distinguish automatically-constructed groups, I address the important sub-problem of producing effective group labels. I conducted an extensive user study comparing PViz to current policy comprehension tools (Facebook's Audience View and Custom Settings page). Our study revealed that PViz was comparable to Audience View for simple tasks, and provided a significant improvement for complex, group-based tasks, despite requiring users to adapt to a new tool.

Pying Data out of a Social Network

Preventing adversaries from compiling significant amounts of user data is a major challenge for social network operators. I examine the difficulty of collecting profile and graph information from the popular social networking Website Facebook and report two major findings. First, I describe several novel ways in which data can be extracted by third parties. Second, I demonstrate the efficiency of these methods on crawled data.

On Image Classification: City vs. Landscape

Grouping images into semantically meaningful categories using low-level visual features is a challenging and important problem in content-based image retrieval [6]. Based on these groupings, effective indices can be built for an image database. The authors show how a specific high-level classification problem (city vs. landscape classification) [17] can be solved from relatively simple low-level features suited for the particular classes. They have developed a procedure to qualitatively measure the saliency of a feature for image classification [20] problem based on the plot of the intra-class and inter-class distance distributions. They use this approach to determine the discriminative power of the following features: color histogram, color coherence vector DCT coefficient, edge direction histogram, and edge direction coherence vector. They determine that the edge direction-based features have the most discriminative power for the classification problem of interest. A weighted k-NN classifier is used for the classification.

Non-parametric kernel ranking approach for social image retrieval

Social image retrieval has become an emerging research challenge in web rich media search. In this paper, I address the research problem of text-based social image retrieval [5], 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, I 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; my 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, I suggest some solution to handle unseen data by warping the non-parametric kernel space to some input kernel function.

Connecting content to community in social media via image content, user tags and user communication

This paper will develop a recommendation framework to connect image content with communities in online social media [4]. The problem is important because users are looking for useful feedback on their uploaded content, but finding the right community for feedback is challenging for the end user. Social media are characterized by both content and community. Hence, in our approach, I characterize images through three types of features: visual features, user generated text tags, and social interaction (user communication history in the form of comments). A recommendation framework based on learning a latent space representation of the groups is developed to recommend the most likely groups for a given image. The model was tested on a large corpus of Flickr [9] images comprising 15,689 images.

Analysing Facebook features to support event detection for photo-based Facebook applications

Facebook witnesses an explosion of the number of shared photos: With 100 million photo uploads a day it creates as much as a whole Flickr each two months in terms of volume. Facebook has also one of the healthiest platforms to support third party applications, many of which deal with photos and related events. While it is essential for many Facebook applications, until now there is no easy way to detect and link photos that are related to the same events, which are usually distributed between friends and albums. In this work, I introduce an approach that exploits Facebook features to link photos related to the same event [13]. In the current situation where the EXIF header of photos is missing in Facebook, I extract visual-based, tagged areas-based, friendship-based and structure-based features. I evaluate each of these features and use the results in my approach. I introduce and evaluate a semi-supervised probabilistic approach that takes into account the evaluation of these features. In this approach I create a lookup table of the initialization values of our model variables and make it available for other Facebook applications or researchers to use. The evaluation of our approach showed promising results and it outperformed the other the baseline method of using the unsupervised EM algorithm in estimating the parameters of a Gaussian mixture model.

Multimedia Semantics: Interactions Between Content and Community

This paper reviews the state of the art and some emerging issues in research areas related to pattern analysis and monitoring of web-based social communities. This research area is important for several reasons. First, the presence of near-ubiquitous low-cost computing and communication technologies has enabled people to access and share information at an unprecedented scale. The scale of the data necessitates new research for making sense of such content. Furthermore, popular websites with sophisticated media sharing and notification features allow users to stay in touch with friends and loved ones; these sites also help to form explicit and implicit social groups. These social groups are an important source of information to organize and to manage multimedia data. In this article, I study how media-rich social networks provide additional insight into familiar multimedia research problems, including tagging and video ranking. In particular, I advance the idea that the contextual and social aspects of media are as important for successful multimedia applications as is the media content. I examine the inter-relationship between content and social context through the prism of three key questions. First, how do I extract the context in which social interactions occur? Second, does social interaction provide value to the media object? Finally, how do social media facilitate the repurposing of shared content and engender cultural memes? I present three case studies to examine these questions in detail. In the first case study, I show how to discover structure latent in the social media data, and use the discovered structure to organize Flickr photo streams. In the second case study, I discuss how to determine the interestingness of conversations - and of participants—around videos uploaded to YouTube. Finally, I show how the analysis of visual content, in particular tracing of content remixes, can help me understand the relationship among YouTube participants.

Content-Based Image Retrieval: Theory and Applications

Advances in data storage and image acquisition technologies have enabled the creation of large image datasets. In this scenario, it is necessary to develop appropriate information systems to efficiently manage these collections. The commonest approaches use the so-called Content-Based Image Retrieval (CBIR) systems. Basically, these systems try to retrieve images similar to a user-defined specification or pattern (e.g., shape sketch, image example). Their goal is to support image retrieval based on content properties (e.g., shape, color, texture), usually encoded into feature vectors. One of the main advantages of the CBIR approach is the possibility of an automatic retrieval process, instead of the traditional keyword-based approach, which usually requires very laborious and time-consuming previous annotation of database images.

Image retrieval: Ideas, Influences, and Trends of the new age

I have witnessed great interest and a wealth of promise in content-based image retrieval as an emerging technology. While the last decade laid foundation to such promise, it also paved the way for a large number of new techniques and systems, got many new people involved, and triggered stronger association of weakly related fields. In this article, I survey almost 300 key theoretical and empirical contributions in the current decade related to image retrieval and automatic image annotation, and in the process discuss the spawning of related subfields. I involved in the adaptation of existing image retrieval techniques to build systems that can be useful in the real world. In retrospect of what has been achieved so far, I conjecture what the future may hold for image retrieval research.

Existing System

Bonneau et al. 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 pro-posed a machine-learning based approach to automatically extract privacy settings from the social context within which the data is produced. Parallel to the work of Danezis, Adu-Oppong et al. 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.

Disadvantages

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.

Proposed System

I 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. 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.

Advantages

My proposed work helps users automate the privacy policy settings for their uploaded images efficiently. Time validity is also one of the important advantages in our project.

System Overview

The A3P system shown in Fig1 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). 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.

Fig1: System Overview

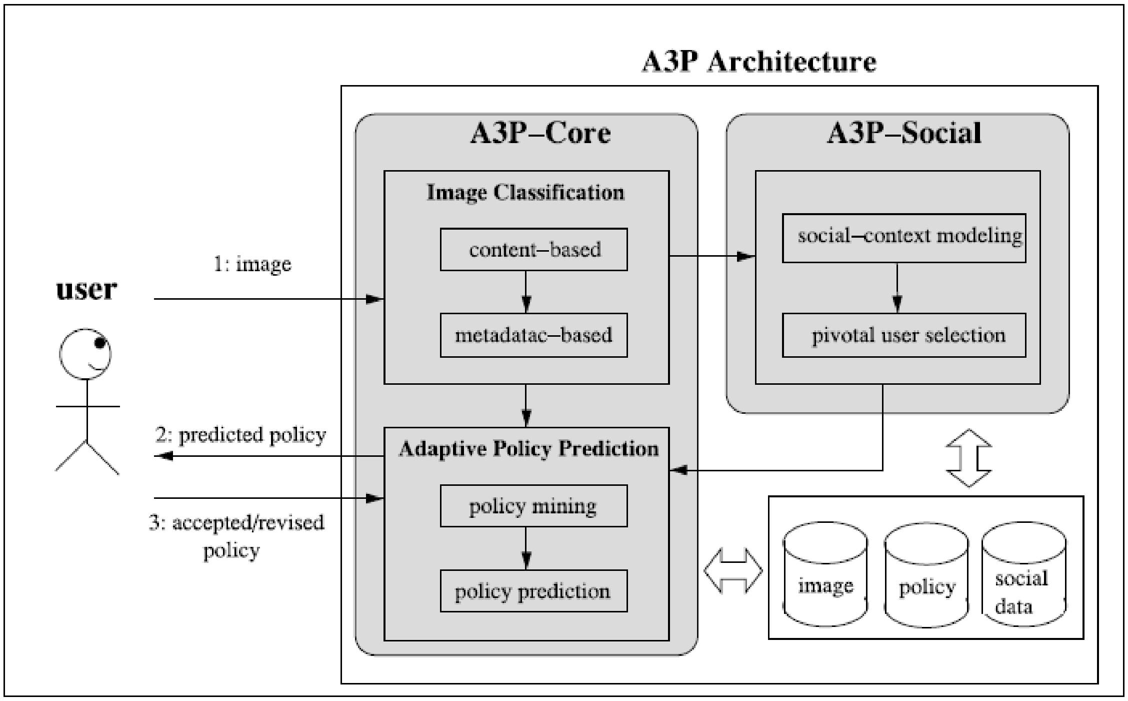
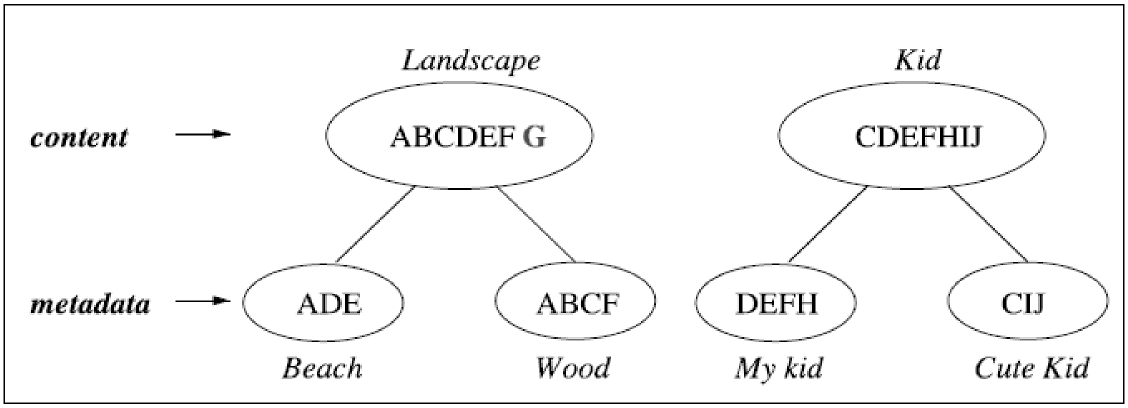


Fig. 2. Two-level Image classification.



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.

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.

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 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.

Moreover, Fig. 2 shows an example of image classification for 10 images named as A, B, C, D, E, F, G, H, I, J respectively. The content-based classification creates two categories: “landscape” and “kid”. Images C, D, E and F are included in both categories as they show kids playing outdoor which satisfy the two themes: “landscape” and “kid”. These two categories are further divided into subcategories based on tags associated with the images. As a result, we obtain two subcategories under each theme respectively. Notice that image G is not shown in any subcategory as it does not have any tag; image A shows up in both subcategories because it has tags indicating both “beach” and “wood”.

Adaptive Policy Prediction

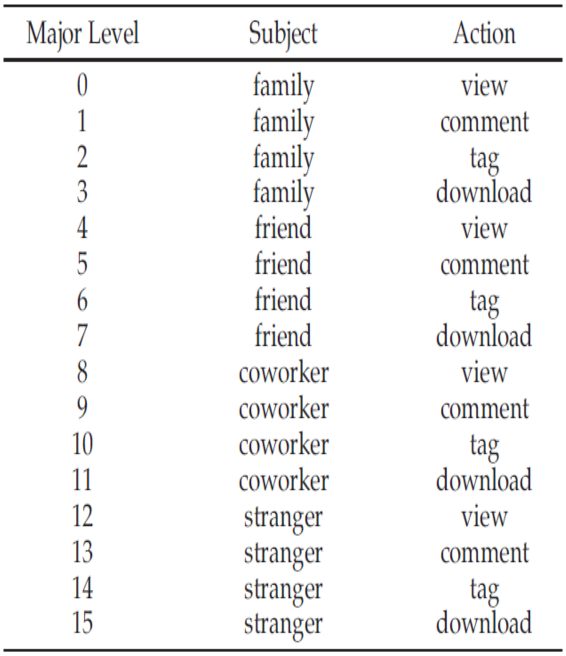
The policy prediction algorithm provides a predicted policy of a newly uploaded image to the user for his/her reference. More importantly, the predicted policy will reflect the possible changes of a user’s privacy concerns. The prediction process consists of three main phases: (i) policy normalization; (ii) policy mining; and (iii) policy prediction. 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.

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 (α), where l is determined by the combination of subject and action in a policy, and α is determined by the system using the condition component. l is obtained via Table 1.

Table 1. Major Level Look-Up Table



Major Level	Subject	Action
0	family	view
1	family	comment
2	family	tag
3	family	download
4	friend	view
5	friend	comment
6	friend	tag
7	friend	download
8	coworker	view
9	coworker	comment
10	coworker	tag
11	coworker	download
12	stranger	view
13	stranger	comment
14	stranger	tag
15	stranger	download

In Table 1, 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 α which is designed to provide fine-grained strictness level. α 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. For example, a user has five family members documented in the system and two of them are kids. When he specifies a policy with the condition age > 18, only three family members will satisfy this condition. The corresponding α is then 3=5 ¼ 0:6.

The larger the value of α, the more people are allowed to access the image and the policy is less restricted. Therefore, we subtract (1- α) from l to obtain the final strictness level as shown in Equation:

$$L = l - (1 - \alpha)$$

Example. Consider the candidate policy Pcan .It has two subjects {family, friend} as well as two actions {comment, tag}. By looking up Table 4, we find that the combination of “friend–tag” yields the lowest major level, i.e., 6. Suppose that the obtained a is 0.3 after evaluating the condition against available user profiles. The final strictness level for Pcan is Pcan 6-(1-0.3)=5.3. After we compute the strictness levels of all candidate policies, we now need to determine which strictness level fits best to the user’s privacy trend. For this purpose, we propose the following approach. We keep monitoring the average strictness level of existing policies in each category of images. The average strictness level is defined as follows:

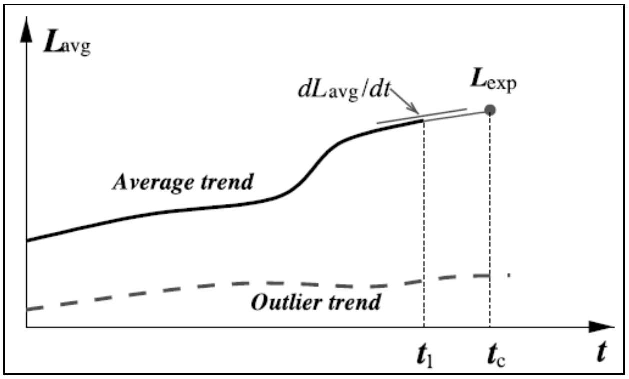
$$L_{avg} = \frac{\sum_{i=1}^{N_p} L_{p_i}}{N_p}$$

where Lpi denote the strictness level of policy Pi, and Np is the total number of policies that satisfy │ Lpi - Lavg │≤ ξ. Notice that the average strictness level is computed by outlier policies. This is because in some situations, users may define special policies which have a very different strictness level from most of others, either much more strict or much more loose. Considering such outliers into the average strictness level calculation would not represent the average case properly. Therefore, when a policy is inserted, we first compare its strictness level with current average strictness level. If the difference is more than a threshold (ξ), we put the policy in the outlier group.

In the experiments, we set ξ to 4 because each role of the policy subject has four different strictness levels as shown in Table 1. The change on the policy preferences being more than four is considered prominent as it exceeds one quarter of the maximum strictness level.

As time evolves, the average strictness levels in each category form a curve as shown in Fig. 3, where values of strictness levels are interpolated in-between any consecutive policy updates. Similarly, the outlier policies may form their own curves as denoted in the figure.

Fig. 3. Average strictness level curve



Let tl denote the last timestamp at which a policy is input to the system, and tc denote the current timestamp when a new image is uploaded. We estimate the expected strictness level Lexp for the new image based on the derivative of the curve of the average strictness level at tl. The derivative can be computed using Secant method. In a summary, Lexp is computed by Equation:

$$L_{exp} = L_{avg}(t_l) + (t_c - t_l) \cdot \frac{dL_{avg}}{dt}(t_l).$$

We compare the strictness level of the candidate policies with Lexp of the average trend, and select the policy which has the closet value to Lexp. When there is more than one policy with strictness levels within the same distance to Lexp, we will conservatively choose the one with the lowest value, i.e., the more restrictive one. Once the user accepted or revised the recommended policy, the new policy will be added to the user’s policy repository.

It is worth noting that the outlier trend may become the average trend at certain point as time passes, and during the transitional period, the policy prediction may not be very accurate. If a user suddenly changes his/her privacy strictness level to a much higher or much lower level, the prediction error of our approach for this single change will be high since we will treat this change as an outlier. If this change is the new preference of this user, this change will be identified when the number of images associated with this new privacy preference is larger than the number of the images associated with the average trend. At that point, the two curves will switch their roles.

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, I first present the types of social context considered by A3P-Social, and then present the policy recommendation process.

Large Scale Evaluation and Analysis

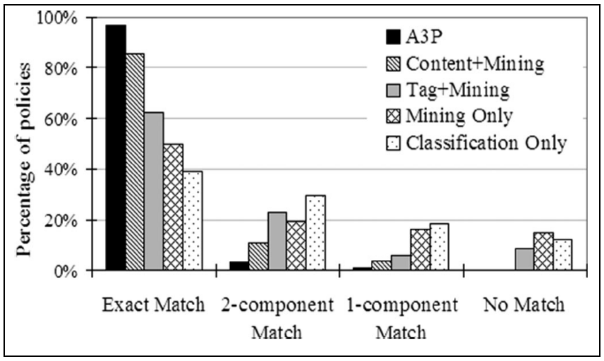
In this first round of tests, we used the two data sets collected through our survey to evaluate the accuracy of our recommended policies.

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 A3Pcore 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 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. 4 shows the percentage 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. Specifically, A3P-core has 90 percent exact match and 0 no match. Moreover, pairwise comparisons were made between A3P-core, "Content+Mining, "Tag+Mining" and the baseline algorithm, corrected using a Bonferroni method.

Fig. 4. A3P comparative performance.



Analysis of Users' Characteristics

We are also interested in examining whether our algorithm performs better for users with certain characteristics. Therefore, we study possible factors relevant to the performance of our algorithm. We used a least squares multiple regression analysis, regressing performance of the A3P-core to the following possible predictors:

- Frequency of social network use was measured on a frequency rating scale (1 ¼ daily; 2 ¼ weekly; 3 ¼ monthly; 4 ¼ rarely; 5 ¼ never) with the item 'How often do you access Social Network Sites?'
- Privacy settings take time was measured on a Likert Scale (5-point rating scale, where 1 ¼ strongly agree and 5 ¼ strongly disagree) with the item 'Changing privacy settings for images uploaded on a social site can be very time consuming.'
- Frequency of sharing pictures was measured using three items (a ¼ 0:69) rated on a Likert scale.

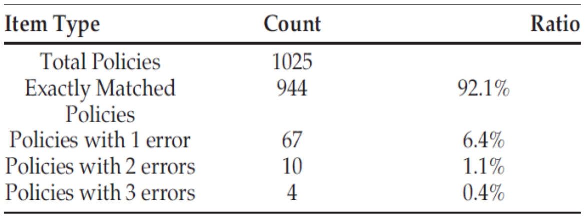
- Frequency of changing privacy settings was measured using four items (a ¼ 0:86) rated on a Likert scale. An example item is ‘I have changed privacy settings for individual pictures.’
- Content of concern was measured using three items (a ¼ 0:81) rated on a Likert scale. An example item is ‘The content of an image is of concern when determining the privacy level for an image.’
- Privacy concern was measured using four items (a ¼ 0:76) rated on a Likert scale. An example item is ‘I have had concerns about my privacy due to shared images on social network sites.’

A3P Social

In the second round of experiments, we analyze the performance of the A3P-Social component by using the first set of data collection. For each user, we use the A3PSocial to predict policies and compare it with a base-line algorithm which does not consider social contexts but bases recommendation only on social groups that have similar privacy strictness level for same type of images. Using the base-line approach, we note that regardless of the individual privacy inclination of the users, the best accuracy is achieved in case of explicit images and images dominated by the appearance of children. In both cases, users maintain more consistent policies, and our algorithm is able to learn them effectively. The largest variability, and therefore worse results occur for images denoting scenery, where the error rate is 15.2 percent. Overall, the accuracy achieved by grouping users by strictness level is 86.4 percent.

Direct User Evaluation

Table 2. Result of Direct User Evaluation



Item Type	Count	Ratio
Total Policies	1025	
Exactly Matched Policies	944	92.1%
Policies with 1 error	67	6.4%
Policies with 2 errors	10	1.1%
Policies with 3 errors	4	0.4%

Table 2 reports the results for the direct user evaluation. Among a total of 1,025 predicted policies, we achieve over 92 percent accuracy (SD ¼ 0:047), in that each participant rejected about two policies on average (1.98). The overall accuracy of the predicted policies in the direct user evaluation is significantly better than the performance in the offline evaluation (tð127Þ ¼ 3:346; p < :01). This demonstrates that users may not have a strong preference regarding privacy settings for individual pictures, and that a system like A3P that can accurately predict preferences will lead to an acceptable level of privacy for users. For mismatched policies, we further examined the type of error. We found that there were total 97 mismatched items (i.e., mismatched subjects, actions and conditions) in those policies. About 60 percent of the errors were due to false positive, which means the predicted policy contains more items than the actual policy. We also noticed that 82.7 percent of the mismatched policies have two components, the subject and action component, fully matched. The most common errors occur within the condition component as this component is the most flexible and can vary significantly if users want to add special constraints. Interestingly, the errors were reported mainly in the first three or four policies displayed to the user. This demonstrates the adaptive nature of our A3P system. Upon correcting mismatched policies, our system’s accuracy increases. We also expect that with more user data and a longer execution of the A3P system, the prediction accuracy will be further increased, as the system adapts to users’ privacy preferences.

Conclusion

In this, 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. This effectively tackled the issue of cold-start, leveraging social context information. This experimental study proves that A3P is a practical tool that offers significant improvements over current approaches to privacy.

References

1. Acquisti A.and Gross R., “Imagined communities: Awareness, information sharing,

and privacy on the facebook," in Proc. 6th Int. Conf. Privacy Enhancing Technol. Workshop, 2006, pp. 36–58.

2. Bonneau J., Anderson J. and Danezis G., "Prying data out of a social network," in Proc. Int. Conf. Adv. Soc. Netw. Anal.Mining., 2009, pp.249–254.

3. Chen H.-M., Chang M.-H., Chang P.-C., Tien M.-C., Hsu W. H. and Wu J.-L., "Sheepdog: Group and tag recommendation for flickr photos by automatic search-based learning," in Proc. 16th ACM Int. Conf. Multimedia, 2008, pp. 737–740.

4. Choudhury M. D., Sundaram H., Lin Y.-R., John A. 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.

5. Da Silva Torres R. and Falcao A., "Content-based image retrieval: Theory and applications," Revista de Informatica Teorica e Aplicada, vol. 2, no. 13, pp. 161–185, 2006.

6. Datta R., Joshi D., Li J. and Wang J., "Image retrieval: Ideas, influences, and trends of the new age," ACM Comput. Surv., vol. 40, no.2, p. 5, 2008.

7. Kapadia A., Adu-Oppong F., Gardiner C. K. and Tsang P. P., "Social circles: Tackling privacy in social networks," in Proc. Symp. Usable Privacy Security, 2008.

8. Klemperer P., Liang Y., Mazurek M., Sleeper M., Ur B., Bauer L., Cranor L. F., Gupta N. 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.

9. Lerman K., Plangprasopchok A. and Wong C., "Personalizing image search results on flickr," CoRR, vol. abs/0704.1676, 2007.

10. Lipford H., Besmer A., and J. Watson, "Understanding privacy settings in facebook with an audience view," in Proc. Conf. Usability, Psychol., Security, 2008.

11. Liu Y., Gummadi K. P., Krishnamurthy B. and Mislove A.,"Analyzing facebook privacy settings: User expectations vs. reality," in Proc. ACMSIGCOMMConf. Internet Meas. Conf., 2011, pp. 61–70.

12. Mazzia A. and LeFevre K. "The PViz comprehension tool for social network privacy settings," in Proc. Symp. Usable Privacy Security, 2012.

13. Rabbath M., Sandhaus P. and Boll S., "Analysing facebook features to support event detection for photo-based facebook applications," in Proc. 2nd ACM Int. Conf. Multimedia Retrieval, 2012, pp. 11:1–11:8.

14. Ravichandran R., Benisch M., Kelley P. and Sadeh N., "Capturing social networking privacy preferences," in Proc. Symp. Usable Privacy Security, 2009.

15. Squicciarini A. C., Sundareswaran S., Lin D. and Wede J., "A3p: Adaptive policy prediction for shared images over popular content sharing sites," in Proc. 22nd ACM Conf. Hypertext Hypermedia, 2011, pp.261–270.

16. Strater K. and Lipford H., "Strategies and struggles with privacy in an online social networking community," in Proc. Brit. Comput. Soc. Conf. Human-Comput. Interact., 2008, pp.111–119.

17. Vailaya A., Jain A. and H. J. Zhang, (1998). On image classification: City images vs. landscapes. Pattern Recog. [Online]. 31(12), pp. 1921–1935.

18. Yeung C. A., Kagal L., Gibbins N. and Shadbolt N., "Providing access control to online photo albums based on tags and linked data," in Proc. Soc. Semantic Web: Where Web 2.0 Meets Web 3.0 at the AAAI Symp., 2009, pp. 9–14.

19. Yu J., Joshi D. and Luo J., "Connecting people in photo-sharing sites by photo content and user annotations," in Proc. IEEE Int. Conf. Multimedia Expo, 2009, pp.1464–1467.

20. Zerr S., Siersdorfer S., Hare J. and Demidova E., "Privacy-aware image classification and search," in Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval, 2012, pp. 35–44.