ISSN: 2277-9655
Impact Factor: 4.116

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

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

S.Keerthivarshini*, Prof. . Janarthanan

[1]*M.E. student,*[2]*Asst Professor , Thirumalai Engineering College,kilambi, kanchipuram, Tamil Nadu, India.*

DOI: 10.5281/zenodo.56889

ABSTRACT

Increasingly developed social sharing websites, like Flickr and YouTube, allow users to create, share, annotate and comment Medias. The large-scale user-generated meta-data not only facilitate users in sharing and organizing multimedia content, but provide useful information to improve media retrieval and management. Personalized search serves as one of such examples where the web search experience is improved by generating the returned list according to the modified user search intents. In this paper, we exploit the social annotations and propose a novel framework simultaneously considering the user and query relevance to learn to personalized image search. The basic premise is to embed the user preference and query-related search intent into user-specific topic spaces. Since the users' original annotation is too sparse for topic modelling, we need to enrich users' annotation pool before user specific topic spaces construction.

KEYWORDS: Uploaded images,social media, Online information services, web-based services.

INTRODUCTION

All Keyword-based search has been the most popular search paradigm in today's search market. Despite simplicity and efficiency, the performance of keyword-based search is far from satisfying. Investigation has indicated its poor user experience - on Google search, for 52% of 20,000 queries, searchers did not find any relevant results. This is due to two reasons. Queries are in general short and nonspecific, e.g., the query of "IR" has the interpretation of both information retrieval and infra-red. Users may have different intentions for the same query, e.g., searching for "jaguar" by a car fan has a completely different meaning from searching by an animal specialist. One solution to address these problems is personalized search, where user-specific information is considered to distinguish the exact intentions of the user queries and re-rank the list results. Given the large and growing importance of search engines, personalized search has the potential to significantly improve searching experience.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.

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.

LITERATURE SURVEY

Some previous systems shows different studies on automatically assign the privacy settings. Jonathan Anderson proposed a paradigm called Privacy Suites [2] which allows users to easily choose "suites" of privacy settings. A privacy suite can be created by an expert using privacy programming. Privacy Suites could also be created directly through existing configuration UIs or exporting them to the abstract format. The privacy suite is distributed through existing distribution channels to the members of the social sites. The disadvantage of a rich programming language is less understandability for end users. Given a sufficiently high-level language and good coding practice, motivated

users should be able to verify a Privacy Suite. The main goal is transparency, which is essential for convincing influential users that it is safe to use.

Fabeah Adu-Oppong developed privacy settings based on the concept of social circles [3]. It provides a web based solution to protect personal information. The technique named Social Circles Finder, automatically generates the friend's list. It is a technique that analyses the social circle of a person and identifies the intensity of relationship and therefore social circles provide a meaningful categorization of friends for setting privacy policies. The application will identify the social circles of the subject but not show them to the subject. The subject will then be asked questions about their willingness to share a piece of their personal information. Based on the answers the application finds the visual graph of users [15].

Kambiz Ghazinour designed a recommender system known as Your Privacy Protector [4] that understands the social net behaviour of their privacy settings and recommending reasonable privacy options. It uses user's personal profile, User's interests and User's privacy settings on photo albums as parameters and with the help of these parameters the system constructs the personal profile of the user. It automatically learned for a given profile of users and assign the privacy options. It allows users to see their current privacy settings on their social network profile, namely Facebook, and monitors and detects the possible privacy risks. Based on the risks it adopts the necessary privacy settings.

Alessandra Mazzia introduced PViz Comprehension Tool [5], an interface and system that corresponds more directly with how users model groups and privacy policies applied to their networks. PViz 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, we also address the important sub-problem of producing effective group labels. PViz is better than other current policy comprehension tools Facebook's Audience View and Custom Settings page.

Peter F. Klemperer developed a tag based access control of data [6] shared in the social media sites. A system that creates access-control policies from photo management tags. Every photo is incorporated with an access grid for mapping the photo with the participant's friends. The participants can select a suitable preference and access the information. Photo tags can be categorized as organizational or communicative based on the user needs. There are several important limitations to our study design. First, our results are limited by the participants we recruited and the photos they provided. A second set of limitations concerns our use of machine generated access-control rules. The algorithm has no access to the context and meaning of tags and no insight into the policy the participant intended when tagging for access control. As a result, some rules appeared strange or arbitrary to the participants, potentially driving them toward explicit policy-based tags like "private" and "public.

Our contribution

Our work is related to the concept on privacy setting configuration in social sites, recommendation systems, and privacy analysis of online images. We propose a novel personalized image search framework by simultaneously considering user and query information. The user's preferences over images under certain query are estimated by how probable he/she assigns the query-related tags to the images.
A ranking based tensor factorization model named RMTF is proposed to predict users' annotations to the images. To better represent the query-tag relationship, we build user-specific topics and map the queries as well as the users' preferences onto the learned topic spaces.

A. User-Specific Topic Modelling

Users may have different intentions for the same query, e.g., searching for "jaguar" by a car fan has a completely different meaning from searching by an animal specialist. One solution to address these problems is personalized search, where user-specific information is considered to distinguish the exact intentions of the user queries and re-rank the list results. Given the large and growing importance of search engines, personalized search has the potential to significantly improve searching experience.

B. Personalized Image Search

In the research community of personalized search, evaluation is not an easy task since relevance judgment can only be evaluated by the searchers themselves. The most widely accepted approach is user study, where participants are

asked to judge the search results. Obviously this approach is very costly. In addition, a common problem for user study is that the results are likely to be biased as the participants know that they are being tested. Another extensively used approach is by user query logs or click through history. However, this needs a large-scale real search logs, which is not available for most of the researchers.

C. ***Ranking – Multi Correlation based***

Photo sharing websites differentiate from other social tagging systems by its characteristic of self-tagging: most images are only tagged by their owners. The tagger statistics for Flickr and the webpage tagging system delicious. We can see that in Flickr, 90% images have no more than 4 taggers and the average number of tagger for each image is about 1.9. However, the average tagger for each webpage in delicious is value 6.1. The severe sparsity problem calls for external resources to enable information propagation. In addition to the ternary interrelations, we also collect multiple intra-relations among users, images and tags. We assume that two items with high affinities should be mapped close to each other in the learnt factor subspaces. In the following, we first introduce how to construct the tag affinity graph, and then incorporate them into the tensor factorization framework. To serve the ranking based optimization scheme, we build the tag affinity graph based on the tag semantic relevance and context relevance. The context relevance of tag is simply encoded by their weighted co-occurrence in the image collection

System Overview

The A3P system consists of two main components: A3Pcore 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 A3P-social: (i) The user does not have enough data for the type of the uploaded image to conduct policy prediction; (ii) The A3Pcore detects the recent major changes among the user's community about their privacy practices along with user's increase of social networking activities like addition of new friends, new posts on one's profile etc. In above cases, it would be beneficial to report to the user the latest privacy practice of social communities that have similar background as the user. The A3P-social groups users into social communities with similar social context and privacy preferences, and continuously monitors the social groups. When the A3P-social is invoked, it automatically identifies the social group for the user and sends back the information about the group to the A3P-core for policy prediction. At the end, the predicted policy will be displayed to the user. If the user is fully satisfied by the predicted policy, he or she can just accept it. Otherwise, the user can choose to revise the policy.

A3P framework

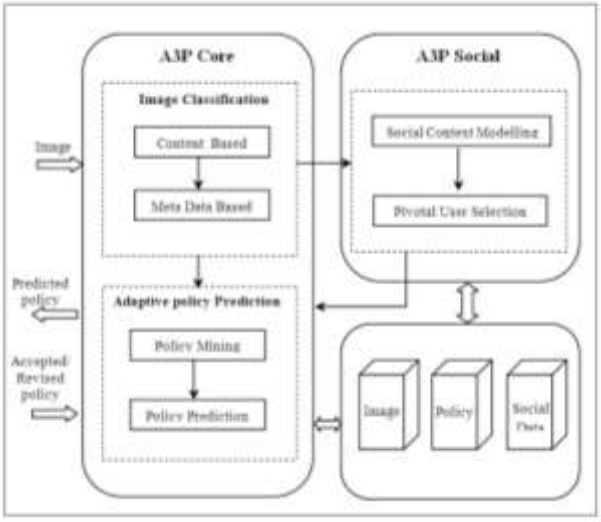
Users can express their privacy preferences about their content disclosure preferences with their socially connected users via privacy policies. Our policies are inspired by popular content sharing sites i.e. Facebook, Picasa, Flickr, although the actual implementation depends on the specific content management site structure and implementation.

In the definition, users in S can be represented by their identities, roles e.g., family, friend, co-workers, or organizations e.g., non-profit organization, profit organization. ID will be the set of images in the user's profile. Each image has a unique ID along with some associated metadata like tags "vacation", "birthday". Images can be further grouped into albums. As for A, we consider four common types of actions: {view, comment, and tag, download}. Last, the condition component C specifies when the granted action is effective. C is a Boolean expression on the grantees' attributes like time, location, and age. For better understanding, an example policy is given by an example. Alice would like to allow her friends and co-workers to comment and tag images in the album named "vacation album" and the image named "summer.jpg" before year 2012. Her privacy preferences can be expressed by the following policy: P: [{friend, co-worker}, {vacation album, summer.jpg}, {comment, tag}, (date< 2012)].The policy prediction algorithm provides a predicted policyof 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: policy normalization; policy mining; and Policy prediction.

1 Policy Normalization

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. An example of policy normalization is shown below. Example 2: Consider policy P in Example 1. Suppose that the album "vacation album" contains k images, namely img1 .jpg, img2 .jpg, imgk.jpg. P is normalized into the following set of atomic rules.

ISSN: 2277-9655
Impact Factor: 4.116



2 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 look for 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 Policy Prediction

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.

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. 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. 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 α as the percentage of people in the specified subject category who satisfy the condition in the policy. For example, a user has 5 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=0.6. The larger the value of α, the more people are allowed to access the image and hence the policy is less restricted. Therefore, we subtract (1-α) from l to obtain the final strictness level. 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.

RESULTS AND DISCUSSION

We keep monitoring the average strictness level of existing policies in each category of images. The average strictness level is defined as follows: where Lpi denote the strictness level of policy Pi, and Np is the total number of policies Notice that the average strictness level is computed by excluding 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 stricter or much looser. 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 4 different strictness levels. Also, the change on the policy preferences being more than 4 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.

Formulae:

$$L = l - (1 - \alpha) \quad , \qquad l_{avg} = \frac{\sum_{i=1}^{N_p} l_{p_i}}{N_p}$$

Tables:

Table 1. 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%

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.

ACKNOWLEDGEMENTS

We take this opportunity to thank our professors and college management for support and guidance in bringing this paper.

REFERENCES

[1] Anna Cinzia Squicciarini, Member, IEEE, Dan Lin, Smitha Sundareswaran, and Joshua Wede, "Privacy Policy Inference of User-Uploaded Images on Content Sharing Sites", *IEEE Transactions on Knowledge and Data Engineering, Vol. 27, NO. 1, January 2015*.

[2] J. Bonneau, J. Anderson, and L. Church, "Privacy suites: Shared privacy for social networks," in *Proc. Symp. Usable Privacy Security, 2009.*

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

[4] Kambiz Ghazinour, Stan Matwin and Marina Sokolova, "Your privacy protector: A Recommender System For Privacy Settings In Social Networks", *International Journal of Security, Privacy and Trust Management (IJSPTM) Vol 2, No 4, August 2013.*

[5] Alessandra Mazzia Kristen LeFevre and Eytan Adar, The PViz Comprehension Tool for Social Network Privacy Settings, *Tech. rep., University of Michigan, 2011.*

[6] Peter F. Klemperer, Yuan Liang, Michelle L. Mazurek, "Tag, You Can See It! Using Tags for Access Control in Photo Sharing", *Conference on Human Factors in Computing Systems, May 2012.*

[7] C. A. Yeung, L. Kagal, N. Gibbins, and N. Shadbolt, "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.*

[8] Sergej Zerr, Stefan Siersdorfer, Jonathon Hare, Elena Demidova , I Know What You Did Last Summer!:Privacy-Aware Image Classification and Search , *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, 2012.*

[9] Anna Cinzia Squicciarini, "Privacy Policy Inference of User-Uploaded Images on Content Sharing Sites", *IEEE Transactions On Knowledge And Data Engineering, vol. 27, no. 1, January 2015.*

[10] 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.*

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

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

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

[15] S. Ahern, D. Eckles, N. S. Good, S. King, M. Naaman, and R. Nair, "Over-exposed: Privacy patterns and considerations in onlineand mobile photo sharing," in Proc. Conf. Human Factors Comput. Syst., 2007, pp. 357–366.

[16] Mehmet Erkan Yüksel and Asım Sinan Yüksel, "An Application for Protecting Personal Information on Social Networking Websites", The Fourth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, 2010.