

Hyrarchical Classification of Images to Predict Privacy Policy to Uploaded Images on Content Sharing Site

Pranay Diwakar Kute¹, H. A. Hingoliwala²

¹Department Computer Engineering, Jayawantrao Sawant College of Engineering, Hadapsar Pune-28, Savitribai Phule Pune University, Pune, India

²Professor, Computer Engineering, Jayawantrao Sawant College of Engineering, Hadapsar Pune-28, Savitribai Phule Pune University, Pune, India

Abstract: *Social Network is an e-service which is booming for sharing contents. Reliable communication is done with social networking sites. Though these sites are a new attacking area for hackers for hacking; they are able to misuse the data through these sites. Some users over CSS affect user's privacy on their personal contents, where some users keep on sending unwanted comments and messages by taking advantage of the user's inherent trust in their relationship network. Towards addressing this need, we propose a Content-Predicated Relegation system to avail users to get automatically set privacy of their images. We examine the role of gregarious context, image content, and metadata as possible bespeakers of user's privacy predilections. Our solution relies on an image relegation framework for image categories which may be associated with homogeneous policies and on a policy presage algorithm to automatically engender a policy for each incipiently uploaded image.*

Keywords: Social media, Content sharing sites, Privacy, Meta data, Content-Based Classification system, policy prediction algorithm

1. Introduction

Hundreds of millions of people on Social Networking (SN) can swap their content through text, media like image, audio, video, etc. It provides a content sharing mechanism and connects people across the world. Users of social media can define a personal profile and modify it as they wish. Through this SM, users may engage with each other for various purposes like business, leisure, and knowledge sharing. People use social networks to get in touch with further people, and create and contribute content that includes personal information, images, and videos. The service providers have admission to the content presented by their users and have the right to collect data and share them to unauthorized users. A very familiar service provided in SN is to produce proposition for finding new friends, groups, and events using mutual filtering techniques. The prosperity of the SN predicated on the number of users it magnetizes, and cheering users to integrate more users to their circle and to apportion data with other users in the SN. So the information will go across the world. End users are nevertheless often not cognizant of the size or nature of the spectators accessing their data and the sense of understanding engendered by organism among digital friends. 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. e may upload some other photos of landscapes which he took as a hobby and for these photos, he may set privacy predilection sanctioning 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.

2. Related Work

P. Klemperer, Y. Liang, M. Mazurek, M. Sleeper, B. Ur, L. Bauer, L. F. Cranor, N. Gupta, and M. Reiter, 2012 find that (a) tags engendered for organizational purposes can be repurposed to engender efficient and plausibly precise access-control rules; (b) users tagging with access control in mind develop coherent strategies that lead to significantly more precise rules than those associated with organizational tags alone; and (c) participants can understand and actively engage with the concept of tag-based access control. [1].

A. Mazzia, K. LeFevre, and A. E. 2012 In the paper we introduced PViz which sanctions the utilizer to understand the overtness of her profile according to automatically constructed, natural sub-groupings of friends, and at different calibers of granularity. Because the utilizer must be able to identify and distinguish automatically-constructed groups, we witalh address the consequential sub-quandary of engendering efficacious group labels. We conducted an extensive utilizer study comparing PViz to current policy comprehension implements (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. Utilizing feedback from the user study, we further iterated on our design, constructing PViz 2.0, and conducted a follow-up study to evaluate our refinements 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.[2].

H. Sundaram, L. Xie, M. De Choudhury, Y. Lin, and A. Natsev 2012 introduced a model which is a part of a method to raise the reusability potential of content which in turn is expected to lower production costs of new content. [3].

S. Zerr, S. Siersdorfer, J. Hare, and E. Demidova 2012. Introduced the classification models for clearly searching for private photos. Also it introduced for query results to serve users with a better coverage of private and public content. Large-scale classification also experiments reveal insights into the predictive performance of various visual as well as textual features, and a user evaluation of query result rankings demonstrates the viability of our approach. [4].

S. Jones and E. O'Neill 2011. Found how decisions are made at the time of granular access control mechanism for sharing photographs are influenced by contextual factors as well as properties relating to detect the contacts. We build analytical models using logistic regression to understand relationships between variables which affect sharing decisions. We also detect how predefined, static groups for privacy control cope with the challenge of sharing large quantity of content related with numerous different contexts, also test whether they need to be adjusted to suit specific contexts. [5].

Y. Liu, K. P. Gummadi, B. Krishnamurthy, and A. Mislove 2011. In this paper, our objective on measuring the disparity between the desired as well as actual privacy settings. We are also quantifying the magnitude of the problem of managing privacy of content. We deploy a survey, implemented as a Facebook application, to 200 Facebook users recruited via Amazon Mechanical Turk. We find that there are 36% of content remains shared with the default privacy settings. We withal find that the overall privacy settings match users' prospects only 37% of the time, and when erroneous, virtually always expose content to more users than expected. Conclusively, we explore how our results have potential to avail users in culling felicitous privacy settings by examining the utilizer-engendered friend lists. [6].

A. C. Squicciarini, S. Sundareswaran, D. Lin, and J. Wede 2011. We introduced the Framework determines the best privacy policy for the uploaded images on the social networking site. It includes an Image classification framework. It is for association of images with same policies as well as a policy prediction technique to automatically generate a privacy policy for images uploaded by uses. [7].

J. Deng, A. C. Berg, K. Li, and L. Fei-Fei 2010. In this paper we have studied large scale categorization which includes a series of challenging experiments on classification with more than 10, 000 image classes. We find that a) computational problem is crucial in algorithm design; b) conventional wisdom from a couple of hundred image categories on relative performance of various classifiers does not hold when the number of categories increases; c)

there is a surprisingly strong relationship between the structure of WordNet as well as the toughness of visual categorization; d) classification can be improved by exploiting the semantic hierarchy. [8].

D. Liu, X.-S. Hua, M. Wang, and H.-J. Zhang 2010 This paper introduces a social image "retagging" scheme. The aims of this scheme are assigning images with better content descriptors. The refining process is arranged as an optimization framework. This framework is based on the consistency between "visual similarity" as well as "semantic similarity" in social images. An effective iterative bound optimization algorithm is applied to learn the optimal tag assignment. In addition, as many tags are intrinsically not closely-related to the visual content of the images. We employ a knowledge-based method to differentiate visual content related from unrelated tags as well as then constrain the tagging vocabulary of our automatic algorithm within the content related tags. [9].

3. Proposed Work

For getting groups of images that may be associated with similar privacy preferences, we propose a content-based classification. This classification classifies the images first based on their contents as well as then refines each and every category into subcategories based on their metadata. Images which do not have metadata will be grouped by content only. Such a content-based classification provides a higher priority to image content and then minimizes the influence of missing tags. Note that it is possible that some of the images are included in various categories meanwhile they contain the typical content features or metadata of respective categories. These two categories are divided again into subcategories based on tags associated with the images. As a result, we get 2 subcategories under each and every theme respectively.

4. Architectural View

Some users over CSS influence user's privacy on their private contents, where the users keep on distributing superfluous comments and messages by attractive advantage of the user's intrinsic trust in their connection network. The overall architecture of the proposed work has given in figure 1.0. This paper switches the most widespread issues

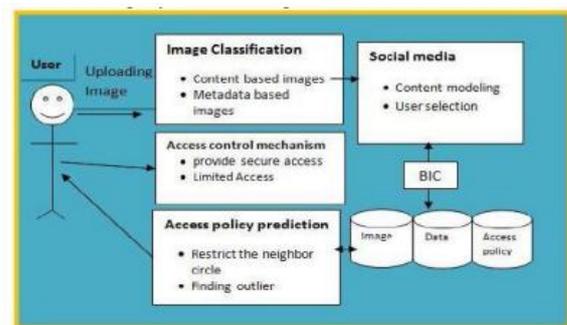


Figure 1: System Architecture

and threats objective different CSS freshly. In CSS, privacy is frequently a key apprehension by the users. As millions of people are willing to interrelate with others, it is also a new

harass ground for image misuses. They are dispersing the images and contents. This paper will demonstrate and argue the most widespread issues and threats targeting different CSS today. Finally, it finds just the right privacy policy scheme for that privacy. This proposes a privacy policy forecast and access boundaries along with overcrowding

scheme for social sites using data mining techniques. It helps to detect and defend distrustful activates, which violates user's privacy in CSS by making an allowance for the following parameters: i) Text annotation, which emerge in the uploaded contents; ii) Image and policy descriptions; iii) Detection of superfluous commends.

| <i>Sr No.</i> | <i>Paper</i> | <i>Technique</i> | <i>Advantages</i> | <i>Disadvantage</i> |
|---------------|-----------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| 1 | Fast algorithms for mining association rules in large databases | AprioriHybrid algorithm, KID3 algorithm | enable marketers to develop and implement customized marketing programs and strategies | Ck is in the wrong order and needs to be sorted on itemsets. |
| 2 | Connecting content to community in social media via image content, user tags and user communication | EM-algorithm | connect user content to the correct community | interactions might always not be directly observable from the image content |
| 3 | Image retrieval: Ideas, influences, and trends of the new age | metric indexing algorithm, automatic annotation algorithms, maximization (EM) algorithm | organize digital picture archives by their visual content. | Work only on problematic due to the semantic gap |
| 4 | What does classifying more than 10,000 image categories tell us? | GIST+NN BOW+NN BOW+SVM SPM+SVM approach | reducing the gap between machine performance and human abilities. | Related to classification is the problem of detection |
| 5 | Interestingness measures for data mining: A survey | Apriori algorithm | Measuring the interestingness of discovered patterns | the key issue is the representation of the user's knowledge |
| 6 | Tag, you can see it!: Using tags for access control in photo sharing | decision-tree algorithm | properly configuring access control can be easy and time-saving | limited way concerning usability |
| 7 | Retagging social images based on visual and semantic consistency | iterative bound optimization algorithm | removes gap between the tags and the actual content of the images | needs to know the full set of relevant tags |
| 8 | Analyzing facebook privacy settings: User expectations vs. reality | community detection algorithms | Overall privacy settings match users' expectations | privacy problem remains unknown |

5. Conclusion

This paper describes privacy policy techniques for user uploaded data images in various content sharing sites. Based on the user social behavior and the user uploaded image, the privacy policy can applied. Content-based classification system is used, which provide users easy and properly, configured privacy setting for their uploaded image. By using this we can easily prevent unwanted discloser and privacy violations. Unwanted disclosure may lead to misuse of one's personal information. Users automate the privacy policy settings for their uploaded images with the help of privacy policy prediction. Based on the information available for a given user the system provides a comprehensive framework to infer privacy preferences. The system is a practical tool.

References

- [1] 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.
- [2] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," Adv. Artif. Intell., vol. 2009, p. 4, 2009.
- [3] X. Sun, H. Yao, R. Ji, and S. Liu, "Photo assessment based on computational visual attention model," in Proc. 17th ACM Int. Conf. Multimedia, 2009, pp. 541–544. [Online]. Available: <http://doi.acm.org/10.1145/1631272.1631351>
- [4] H. Sundaram, L. Xie, M. De Choudhury, Y. Lin, and A. Natsev, "Multimedia semantics: Interactions between content and community," Proc. IEEE, vol. 100, no. 9, pp. 2737–2758, Sep. 2012.
- [5] A. Vailaya, A. Jain, and H. J. Zhang, (1998). On image classification: City images vs. landscapes. Pattern Recog. [Online]. 31(12), pp. 1921–1935. Available: <http://www.sciencedirect.com/science/article/pii/S003132039800079X>
- [6] R. A. Wagner and M. J. Fischer, "The string-to-string correction problem," J. ACM, vol. 21, no. 1, pp. 168–173, 1974.
- [7] Wordnet - A lexical database for the English language. [Online]. Available: <http://wordnet.princeton.edu/>
- [8] C.-H. Yeh, Y.-C. Ho, B. A. Barsky, and M. Ouhyoung, "Personalized photograph ranking and selection system," in Proc. Int. Conf. Multimedia, 2010, pp. 211–220. [Online]. Available: <http://doi.acm.org/10.1145/1873951.1873963>
- [9] 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.
- [10] J. Yu, D. Joshi, and J. Luo, "Connecting people in photo-sharing sites by photo content and user annotations," in Proc. IEEE Int. Conf. Multimedia Expo, 2009, pp.1464–1467.
- [11] S. Zerr, S. Siersdorfer, J. Hare, and E. Demidova, "Privacy-aware image classification and search," in Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval, 2012, pp. 35–44.

- [12] S. Zerr, J. H. Stefan Siersdorfer, and E. Demidova, (2012). Picalert! data set. [Online]. Available: <http://l3s.de/picalert/>
- [13] N. Zheng, Q. Li, S. Liao, and L. Zhang, "Which photo groups should I choose? A comparative study of recommendation algorithms in flickr," J. Inform. Sci., vol. 36, pp. 733–750, Dec. 2010.
- [14] J. Zhuang and S. C. H. Hoi, "Non-parametric kernel ranking approach for social image retrieval," in Proc. ACM Int. Conf. Image Video Retrieval, 2010, pp. 26.

Author Profile



Mr. Pranay D. Kute is currently pursuing M.E (Computer) from Department of Computer Engineering, Jayawantrao Sawant College of Engineering, Pune, India. Savitribai Phule Pune University, Pune, Maharashtra, India -411007. He received her B.E (Computer) Degree from St. Vincent Pallotti College of Engineering, Nagpur, India. Rashtra Sant Tukdoji Maharaj Nagpur University, Nagpur, Maharashtra, India - 441108. area of interest is data mining.



Prof. H. A. Hingoliwala received his M Tech. He is currently working as Head of Department and Assoc. Prof. (Computer) at Department of Computer Engineering, Jayawantrao Sawant College of Engineering, Pune, India. Savitribai Phule Pune University, Pune, Maharashtra, India - 411007. His area of interest is Image Processing and Networking.

