Exploring Abstract Concepts for Image Privacy Prediction in Social Networks (Student Abstract)

Gabriele Galfré, Cornelia Caragea
Department of Computer Science, University of Illinois at Chicago
Chicago, Illinois
{ggalfr2, cornelia}@uic.edu

Abstract

Automatically detecting the private nature of images posted in social networks such as Facebook, Flickr, and Instagram, is a long-standing goal considering the pervasiveness of these networks. Several prior works to image privacy prediction showed that object tags from images are highly informative about images' privacy. However, we conjecture that other aspects of images captured by abstract concepts (e.g., religion, sikhism, spirituality) can improve the performance of models that use only the concrete objects from an image (e.g., temple and person). Experimental results on a Flickr dataset show that the abstract concepts and concrete object tags complement each other and yield the best performance when used in combination as features for image privacy prediction.

Introduction

Social network users share quotidian details and post pictures of their significant milestones and private events. The smartphones and other mobile devices facilitate image sharing virtually at any time. New privacy concerns are rising and mostly emerge due to users' lack of understanding that semantically rich images may reveal sensitive information.

Recently, Squicciarini et al. (2017) and Zerr et al. (2012) explored learning models for image privacy prediction using user tags and visual features and found that user tags are informative for classifying images as *private* or *public*. However, the user tags are at the sole discretion of users, and hence, they typically tend to be noisy and incomplete. To address this limitation, Tonge and Caragea (2016) proposed to automatically extract object tags from the images' content using Convolutional Neural Networks and showed that using these deep object tags are helpful for image privacy prediction. Still, these deep object tags correspond to the $1,000$ object categories available in the ImageNet dataset and capture only concrete object types (not abstract).

We conjecture that in addition to the concrete object tags, other aspects of images captured by abstract concepts can be used to better discriminate images as *private* or *public*.

Figure 1 shows an example of a private images (by a consensus of multiple annotators) and its abstract concepts and concrete object tags. Using the combination of abstract and concrete tags can help describe the image's content better, and hence, more accurately infer the prediction of the image's privacy.

religion, *sikhism*, *spirituality*, **person**, **temple**.

Figure 1: Example of a private image and its *abstract* (red) concepts and **concrete** (blue) object tags

The idea of abstractness has been investigated in term of the cognition process in the human mind (Paivio 2013) and how it affects decision making and learning tasks when they involve the representation of abstract and concrete concepts. Fascinating theories have been formulated about what these two aspects of knowledge represent essentially in terms of the brain activity (Schwartz and Mitchell 2019). Several authors agreed on defining as concrete what can be experienced directly through senses and physical actions, while abstract are those concepts that need a certain level of rational processing to be represented.

To this end, we propose to explore abstract concepts for predicting images' privacy in order to capture additional information that is not encoded in the concrete tags. Our results show that adding abstract concepts to the concrete object tags yields better performance compared with concrete tags alone. To our knowledge, this is the first work to explore the impact of "abstractness" for privacy prediction.

Proposed Features

The main purpose of this work is to evaluate the impact of abstract concepts on image privacy prediction as compared with concrete object tags alone. In order to achieve this goal and obtain the desired insight, we encode images as follows: given an image I and a set of tags T corresponding to I, we split T into two subsets corresponding to concrete object tags T^C and abstract concepts T^A. Subsequently, T^C

Copyright © 2020, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Experiment Setup	Tags Type	SVM			NB			RF			CNN		
		Acc.	Pri. F1	Pub. F1	Acc.	Pri. F1	Pub. F1	Acc.	Pri. F1	Pub. F1	Acc.	Pri. F1	Pub. F1
1 Tag Equal	C	51.21	49.24	49.37	52.76	51.37	53.82	62.19	71.92	71.98	60.48	51.65	66.56
	A+C	60.51	66.94	67.10	61.51	56.47	65.50	67.86	74.91	74.92	67.93	57.75	74.12
2 Tags Equal	C	57.71	64.22	64.36	56.31	51.91	59.97	62.77	72.06	72.08	63.25	52.24	70.11
	A+C	64.68	71.83	71.84	68.79	63.14	72.94	71.18	76.92	76.93	72.89	63.18	78.53
3 Tags Equal	C	59.08	66.39	66.53	61.71	54.3	67.05	63.62	71.47	71.48	65.81	54.82	72.45
	A+C	65.54	72.76	72.77	70.32	64.84	74.34	72.92	77.83	77.86	74.02	65.1	79.29
4 Tags Equal	C	60.96	68.3	68.35	62.8	56.77	67.36	65.13	72.69	72.71	67.35	57.16	73.59
	A+C	67.25	74.13	74.15	71.49	66.18	75.37	73.91	78.71	78.75	73.91	64.51	79.32

Table 1: Results (percentage values (%) of Accuracy and F1-scores) of image privacy prediction using various classifiers.

and T^A are encoded as bag of words for various machine learning and deep learning models.

Rabinovich et al. (2018) recently used a weakly supervised neural network model to label a set of almost $100,000$ words with abstractness numeric scores between 0 and 1. Low values indicate words with more concrete connotations, while high values are related to concepts representing linguistic constructs elevated from the perceptible/abstract sphere. Thus, because of its size, for the purpose of splitting the set of tags T into T^C and T^A adopting the threshold score value of 0.5, we directly use this set of *abstractness* scored words to encode images with concrete tag features and concrete + abstract tag features.

Experiments

Dataset and Evaluation Setting: We evaluate the quality of abstract concepts on a subset of Flickr images sampled from the PicAlert dataset (Zerr et al. 2012). PicAlert contains images on various subjects, which are manually labeled as *public* or *private* by external viewers. We regard the set of tags T as corresponding to the user tags. The pre-processing pipeline adopted for handling user tags consists in tokenization, punctuation removal, (possible) english translation, non-alphabetic symbols and stopwords removal, and lemmatization. Our final goal is to tailor a dataset to be able to study the impact of abstract and concrete tags, therefore two constraints have been selected for choosing samples: a minimum of 2 abstract tags and 2 concrete tags per image, and a minimum of 10 tags per image. The final dataset consists of $2,925$ images ($1,072$ private and $1,853$ public) annotated with $4,235$ different scored words (913 abstract and $3,322$ concrete). The **Train**, **Validation**, and **Test** sets represent 60%:20%:20% from the entire dataset, respectively. We use five random seeds to obtain the splits and averaged results across the five runs on the **Test** set.

Results and Observations: Table 1 shows the results (overall accuracy and the F1-score) on the image privacy prediction task using various classifiers, Support Vector Machine (SVM), Naïve Bayes (NB), Random Forest (RF), and Tag-based Convolutional Neural Network (CNN). In experiments, we used k concrete tags for the C experiment, and $2k$ tags for the A + C experiment (k concrete tags - the same tags as in the C experiment, augmented with k abstract tags).

As can be seen from the table, augmenting the set of concrete object tags with abstract concepts always improves the performance of all classifiers in terms of all performance measures, regardless of the number of tags used in an experiment, with k ranging from 1 to 4. We can also observe that the difference in performance is higher for smaller values of k and becomes smaller and smaller as k increases. Among all four classifiers compared, Random Forest provides the highest F1-score for the private class. Moreover, when we compare the C experiment with 4 tags (4 concrete tags - last block) with the A + C experiment with 2 equal tags (2 concrete and 2 abstract tags - 4 tags in total - second block of results), we can see that the A + C results with 4 tags outperform the C results with 4 tags in total, for all performance measures and all classifiers.

Conclusion and Future work

We explored the use of "abstractness" of concepts to augment concrete object tags for image privacy prediction and showed that they can improve the prediction performance of this task. Specifically, the best performance is achieved when we consider the combination of abstract concepts with concrete object tags. We conclude that abstract concepts and concrete object tags complement each other and help boost the performance. In the future, more sophisticated methods to combine these types of tags can be explored.

Acknowledgments. This research is supported by NSF.

References

Paivio, A. 2013. Dual coding theory, word abstractness, and emotion: a critical review of kousta et al. (2011). *Journal of experimental psychology. General* 142 1:282–7.

Rabinovich, E.; Sznajder, B.; Spector, A.; Shnayderman, I.; Aharonov, R.; Konopnicki, D.; and Slonim, N. 2018. Learning concept abstractness using weak supervision. *CoRR* abs/1809.01285.

Schwartz, D., and Mitchell, T. M. 2019. Understanding language-elicited EEG data by predicting it from a fine-tuned language model. *CoRR* abs/1904.01548.

Squicciarini, A.; Caragea, C.; and Balakavi, R. 2017. Toward automated online photo privacy. *ACM Transactions on the Web* 11(1).

Tonge, A., and Caragea, C. 2016. Image privacy prediction using deep features. In *AAAI*.

Zerr, S.; Siersdorfer, S.; Hare, J.; and Demidova, E. 2012. Privacy-aware image classification and search. In *ACM SIGIR*. NY, USA: ACM.