Image Labeling on a Network: Using Social-Network Metadata for Image Classification

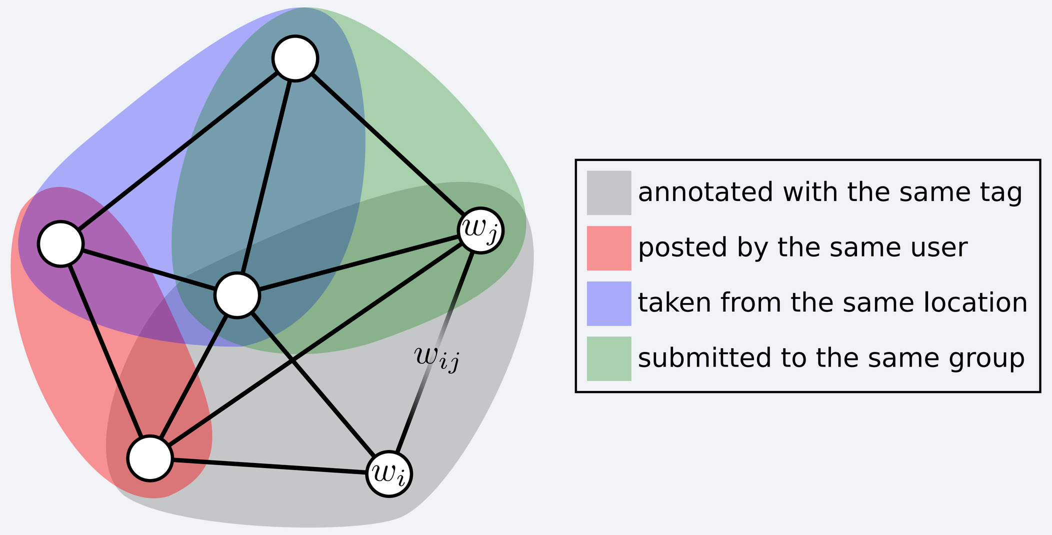
Julian McAuley, Jure Leskovec

Stanford

Abstract

We study the use of social network metadata for image classification. Existing multimodal classification frameworks use metadata such GPS, EXIF, tags, and user profiles. However, online photo sharing networks like Flickr include several additional sources of metadata that can be harnessed for image classification. We build relational models for such types of metadata.

Building a graph of related images



We form edges between images with common metadata. Edge features include the number of common tags, groups, collections, and galleries, as well as location and user profile information.

Model

We model image labels in terms of image features $\phi(x_i)$, and image relationships $\phi(x_i, x_j)$. We then label an *entire dataset* according to

$$\underset{Y \in \{-1,1\}^N}{\operatorname{argmax}} \sum_{i=1}^{N} y_i \underbrace{\langle \phi(x_i), \theta^{\text{node}} \rangle}_{w_i} + \sum_{i=1}^{N} \sum_{j=1}^{N} \delta(y_i = y_j) \underbrace{\langle \phi(x_i, x_j), \theta^{\text{edge}} \rangle}_{w_{ij}},$$

which can be done efficiently using *graph cuts* so long as related images prefer to have similar labels. We learn the optimal model $(\theta^{\text{node}}; \theta^{\text{edge}})$ using Structured Learning approaches.

Data

	CLEF	PASCAL	MIR	NUS	ALL
Number of photos	**4546**	**10189**	**14460**	**244762**	**268587**
Number of users	2663	8698	5661	48870	58522
Number of tags	21192	27250	51040	422364	450003
Number of groups	10575	6951	21894	95358	98659
Number of comments	77837	16669	248803	9837732	10071439
Number of sets	6066	8070	15854	165039	182734
Number of galleries	1026	155	3728	100189	102116
Number of locations	1007	1222	2755	22106	23745
Number of labels	**99**	**20**	**14**	**81**	**214**

We augment four popular datasets using metadata from Flickr. Our data is available at `i.stanford.edu/~julian/`

Evaluation

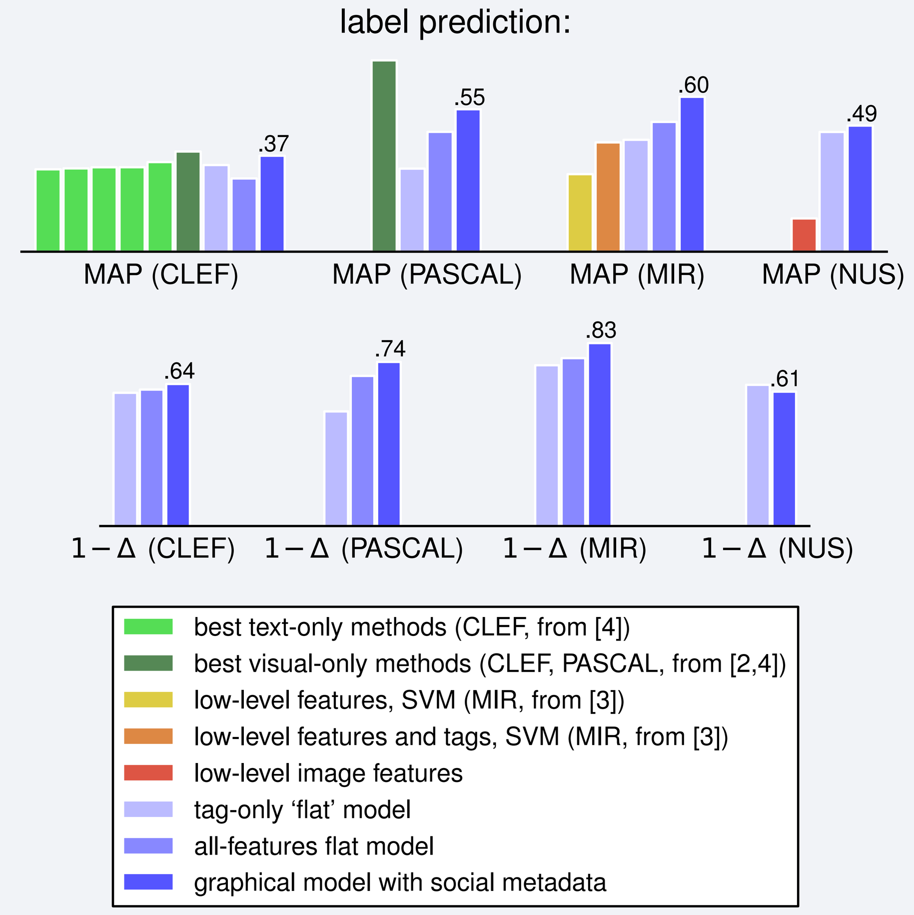
We evaluate our method using published classification results from the four benchmark datasets we consider. For computational reasons our method is optimized to minimize the Balanced Error Rate

$$\Delta(Y, Y_c) = \frac{1}{2}\left[\underbrace{\frac{|Y^{\text{pos}} \setminus Y_c^{\text{pos}}|}{|Y_c^{\text{pos}}|}}_{\text{false positive rate}} + \underbrace{\frac{|Y^{\text{neg}} \setminus Y_c^{\text{neg}}|}{|Y_c^{\text{neg}}|}}_{\text{false negative rate}}\right],$$

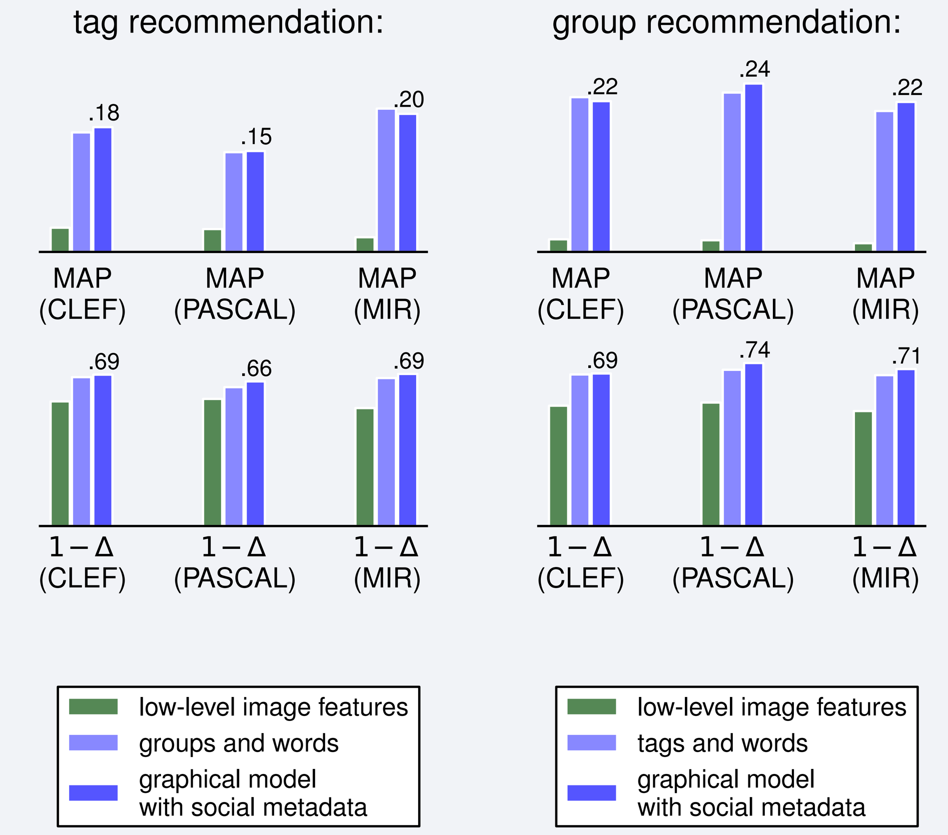
though we also report the Mean Average Precision for the sake of comparison.

In addition, we use our model to recommend tags and groups for each image.

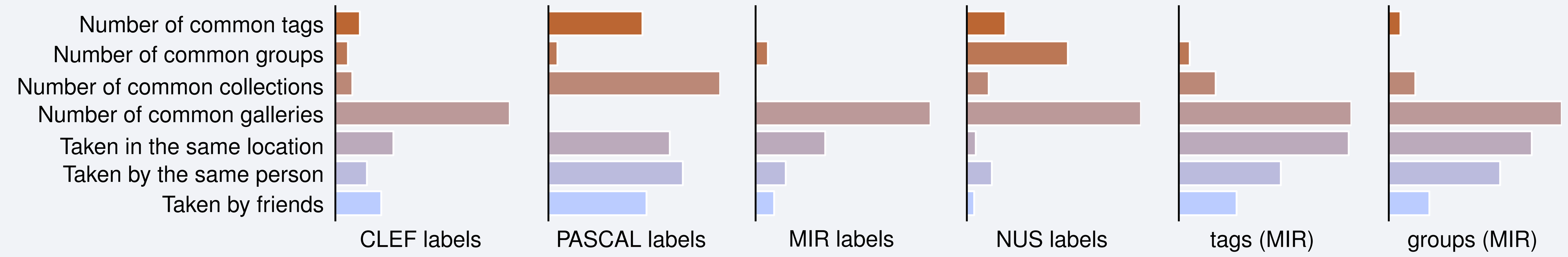
Image labeling results



Tag and group prediction

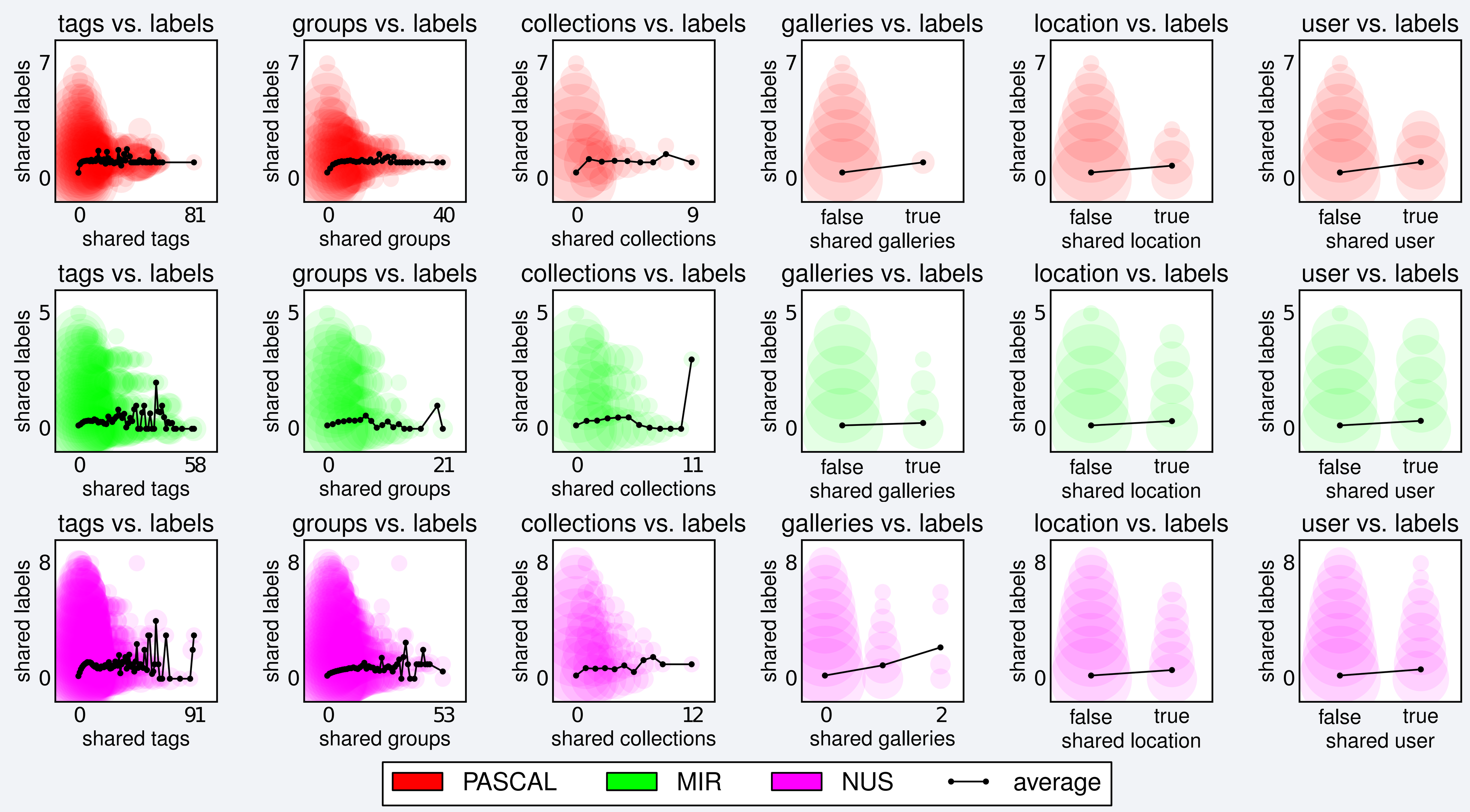


Which social network features are useful?



We confirm existing findings that tag and GPS data are useful for classification, while also finding that other sources of metadata are informative.

Do images with similar metadata have similar labels?



For all forms of metadata, we find that images with similar metadata tend to have similar labels.

References

[1] J. J. McAuley and J. Leskovec. Image labeling on a network: Using social-network metadata for image classification. In *ECCV*, 2012.

[2] M. Everingham, L. Van Gool, C Williams, J. Winn, and A. Zisserman. The PASCAL visual object classes (VOC) challenge. *IJCV*, 2010.

[3] Mark Huiskes and Michael Lew. The MIR Flickr retrieval evaluation. In *CIVR*, 2008.

[4] Stefanie Nowak and Mark Huiskes. New strategies for image annotation: Overview of the photo annotation task at ImageCLEF 2010. In *CLEF*, 2010.

[5] Tat-Seng Chua, Jinhui Tang, Richang Hong, Haojie Li, Zhiping Luo, and Yan-Tao Zheng. NUS-WIDE: A real-world web image database from the National University of Singapore. In *CIVR*, 2009.

jmcauley@cs.stanford.edu, jure@cs.stanford.edu