Social Group Suggestion from User Image Collections

Jie Yu
Kodak Research Labs
Rochester, NY 14616
jie.yu@kodak.com

Xin Jin
University of Illinois at Urbana-Champaign
xinjin3@illinois.edu

Jiawei Han
University of Illinois at Urbana-Champaign
hanj@cs.uiuc.edu

Jiebo Luo
Kodak Research Labs
Rochester, NY 14616
jiebo.luo@kodak.com

ABSTRACT
Photo-sharing services have attracted millions of people and helped construct massive social networks on the Web. A popular trend is that users share their image collections within social groups, which greatly promotes the interactions between users and expands their social networks. Existing systems have difficulties in generating satisfactory social group suggestions because the images are classified independently and their relationship in a collection is ignored. In this work, we intend to produce suggestions of suitable photo-sharing groups from a user's personal photo collection by mining images on the Web and leveraging the collection context. Both visual content and textual annotations are integrated to generate initial prediction of the events or topics depicted in the images. A user collection-based label propagation method is proposed to improve the group suggestion by modeling the relationship of images in the same collection as a sparse weighted graph. Experiments on real user images and comparisons with the state-of-the-art techniques demonstrate the effectiveness of the proposed approaches.

Categories and Subject Descriptors
H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing – *Algorithms, Indexing methods;*

General Terms
Algorithms, Theory, Experimentation, Performance.

Keywords
social image, group recommendation, label propagation

1. INTRODUCTION
Media-sharing websites such as *Flickr* and *YouTube* contain massive social networks, where people are connected by their common interests on certain types of images or videos and generate millions of sub-communities [1]. For example, Figure 1 shows the webpage of the "Architecture" Flickr group with more than 29000 members and 400000 images. Most of the activities within a group start from sharing images: users would select and contribute their own images to the related group, comment on other member's photos, and discuss related photographic techniques. Such interactions open a new way to greatly expand one's social network. Consequently, the formation of groups has gained great popularity and attracted enormous number of users [2].

Currently, a user has to manually assign each photo in his/her collection to the appropriate group, which requires matching the subject of each image with the topics of various interest groups. This work is tedious and often prohibits users from exploiting the relevant group resources.

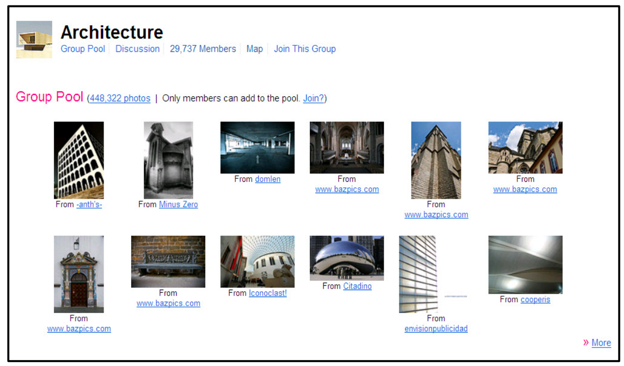

Figure 1. Social groups attract users based upon their shared interest, e.g., a group featuring architecture photos on Flickr.

In this paper, we present a new method to make accurate group suggestions for each image using both the conventional content analysis on each image and a novel collective analysis of all the images in a user collection. Our system uses Support Vector Machine to produce initial group prediction for each image based on visual content and topic analysis. We further leverage the relationship between the images in the same user collection to improve the prediction via a novel technique called collection-based sparse label propagation.

2. METHODOLOGY
2.1 Social Group Suggestion
We collect 16145 images from 767 Flickr users, which are contributed to 24 popular Flickr groups. Using SimRank [3] and spectral clustering [4], we further automatically categorize the groups into 11 classes, i.e. *Flower, Animal, Architecture, Portrait, Seashore, Night, Sky, Sunset/Sunrise, Red, Green,* and *B&W*. Statistical analysis suggests that users often contribute their images to all the groups in the same group category. Similarly, it would be desirable to suggest all the groups in the category that matches the visual content of user image. Therefore, the social group suggestion task is converted to as a classification problem.

Theoretically, any classification method can be plugged into our framework to learn the subjects of different group categories. For a image X, its initial social group prediction Y^0 can be obtained using a classifier F as:

$$Y^0 = F(X) \quad (1)$$

2.2 Collection graph representation
Graph representation has been an active topic in the data mining and machine learning fields. As data points are represented as graph nodes, the intrinsic relationship between data can be

captured by graph edge W. Intuitively, the similarity between samples can be preserved by finding the optimal W that minimizes the reconstruction loss:

$$\min \sum_i \| x_i - \sum_{j \neq i} w_{ij} x_j \|^2 \qquad (2)$$

Instead of using the optimal W found in (2), we constrain the W to be sparse for three reasons: 1) Recent findings in neural science suggest that human vision system interprets images based on the sparse representation of the visual features [5]; 2) Theoretically, the sparse W does not make the local distribution assumption and provides an interpretive explanation of the correlation weights. 3) Practically, the shrinkage of coefficients in combining predictors often improves prediction accuracy. Additionally, we further constrain that the samples should only be reconstructed by using similar ones in the same user's collection. Therefore, the sparse graph representation for user collection can be obtained from the following optimization problem:

$$W = \min_w \sum_i \| x_i - \sum_{\mathrm{i,j \in same\ collection}, j \neq \mathrm{i}} w_{ij} x_j \|^2 \quad \text{s.t.} \sum_{i,j} | w_{ij} | < s \qquad (3)$$

This optimization problem in (3) could be solved by several algorithms such as LASSO and modified Least Angle Regression.

2.3 Sparse label propagation

It is reasonable to assume that similar images from the same user's collection should have similar prediction. Therefore, the prediction among images from the same user collection should be propagated in the following iterative process.

$$Y^{t+1} = (1 - \Lambda) W \cdot Y^t + \Lambda Y^0 \qquad (4)$$

Note that Λ is a diagonal matrix with

$$\lambda_{i,i} = \max \frac{y_{i,j}^0}{\sum_j y_{i,j}^0} \qquad (5)$$

where $y_{i,j}^0$ is the initial prediction of sample x_i for class j. The definition in (5) indicates that Λ describes the confidence of the initial prediction. Therefore, the refined prediction can be learned from visually similar samples with confident prediction. The final prediction for images is obtained by updating equation (4) until convergence or for certain number of iterations.

3. EXPERIMENTS AND ANALYSIS

We implement the proposed Sparse Label Propagation (SLP) method as discussed in Section 2. The number of iteration is set to 100. Four popular visual descriptors, tiny image, color histogram, GIST [6], and CEDD [7], and topic representation of user annotations [8] are extracted to represent the images in compact feature space. Support Vector Machine is trained to produce initial group suggestion as the baseline. Due to its resemblance to our method, Linear Neighborhood Propagation (LNP) [9] is also implemented for comparison.

The unique property of our research is that we try to leverage user collection information. Therefore, instead of randomly selecting images from the entire data set, we construct our training set by randomly selecting 200 users and all of the images in their collections. The rest of the images are used for testing.

The average accuracy for all test samples in each category is shown in Figure 2. Several conclusions can be drawn from the results: 1) Compared with SVM, SLP improves the average accuracy from 55% to 62% and the relative improvement is about 12.7%. It shows that the collection-level context is successfully leveraged in the collection-based spare propagation method. 2) In 10 out of 11 categories, our new method helps make better prediction. It performs slightly worse than the baseline on the architecture images. Upon reviewing the misclassified images, we find the potential reason could be the visual diversity for buildings and the noisy background in some user collections. 3) Both SLP and the baseline classifier outperform LNP because LNP can not use the collection information and the predictions from the trained classifier.

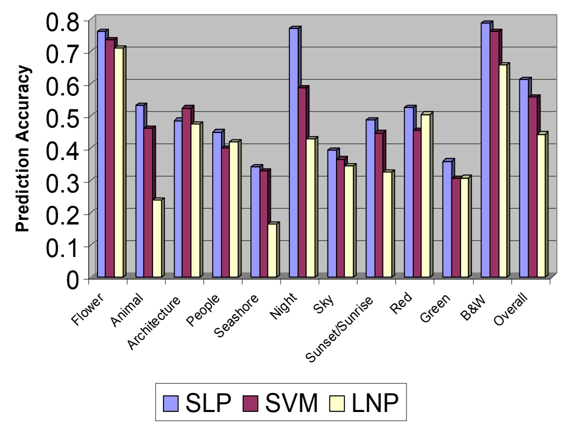

Figure 2. Average prediction accuracy of SLP, SVM and LNP

4. CONCLUSIONS AND FUTURE WORK

In this paper, we propose to use the entire user image collection to collectively suggest social groups based on the image content and user annotations. State-of-the-art machine learning techniques are used to build classifiers and generate the initial group suggestion for each image based on its text annotation and visual descriptors. User collection information is further leveraged to enhance the prediction accuracy through collection-based sparse label propagation. The proposed methods have been tested on group suggestion tasks for real user collections and demonstrated superior performance over other techniques.

5. REFERENCES

[1] Pedro, J., and Siersdorfer, S. Ranking and Classifying Attractiveness of Photos in Folksonomies, *World Wide Web Conference*, 2009.

[2] Hays, J. and Efros, A. IM2GPS: estimating geographic information from a single image, *In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2008.

[3] Jeh, G., and Widom, J. Simrank: a measure of structural context similarity, In *Proc. of the 8th Int'l Conference on Knowledge discovery and data mining*, 2002.

[4] Kannan, R., Vempala, S., and Vetta, A. On clusterings: Good, bad and spectral, *Journal of the ACM*, 51(3):497–515, 2004.

[5] Griffiths, T., and Steyvers, M. Finding scientific topics, *Proc. Natl. Acad. Sci. USA*, April 2004.

[6] Oliva, A., and Torralba, A. Modeling the shape of the scene: a holistic representation of the spatial envelope, *Intl. Journal of Computer Vision*, Vol. 42(3), 2001.

[7] Chatzichristofis, S., and Boutalis, Y. CEDD: Color and Edge Directivity Descriptor: A Compact Descriptor for Image Indexing and Retrieval, *Computer Vision Systems*, pp. 312-322., 2008.

[8] Blei, D., Ng, A., and Jordan, M. Latent Dirichlet Allocation, *Journal of Machine Learning Research*, 2003.

[9] Wang, F., and Zhang, C. Label Propagation through Linear Neighborhoods, *IEEE Trans. on Knowledge and Data Engineering*, 2008.