

Exploring Tag Relevance for Image Tag Re-ranking

Jie Xiao
Department of Computer Science
The University of Texas at
San Antonio
jxiao@cs.utsa.edu

Wengang Zhou
Department of Computer Science
The University of Texas at
San Antonio
zhwgeis@gmail.com

Qi Tian
Department of Computer Science
The University of Texas at
San Antonio
qitian@cs.utsa.edu

ABSTRACT

In this paper, we propose to explore the relevance between tags for image tag re-ranking. The key component is to define a global tag-tag similarity matrix, which is achieved by analysis in both semantic and visual aspects. The text semantic relevance is explored by the Latent Semantic Indexing (LSI) model [1]. For the visual information, the tag-relevance can be propagated by reconstructing exemplar images with visually and semantically consistent images. Based on our tag relevance matrix, a random-walk approach is leveraged to discover the significance of each tag. Finally, all tags in an image are re-ranked by their significance values. Extensive experiments show its effectiveness on an image dataset with a large tags vocabulary.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Search process

General Terms

Algorithms, Measurement, Performance.

Keywords

Tag re-ranking, random walk, latent semantic indexing.

1. INTRODUCTION

Recently, millions of tagged images are available online in social community. However, tags are labeled in a random order and cannot accurately describe the image content. Therefore, it hinders the applications in the real world. Image tag re-ranking becomes an interesting topic in research community [2] and industry. In this paper, to tackle this problem, we explore the latent semantic relevance among tags from text and visual perspectives.

In sum, this paper has three main contributions: (1) we propose a new scheme to measure tag relevance by integrating a subset of images sharing the same target tag into a global tag-tag similarity matrix; (2) we adopt Latent Semantic Indexing to measure tag-tag text similarity and integrate them with different schemes; and (3) we design a three-step approach to propagate the tag relevance from trustable images that are visually similar to the representative exemplar image.

2. ALGORITHM FRAMEWORK

Given an image with several tags, our goal is to re-rank those tags, so that the more relevant the tag is to the image, the higher it is ranked. We exploit the tag-tag relevance from the perspective of text similarity and visual similarity, respectively.

Denote \mathbf{I} as an image dataset with n images, and \mathbf{T} as tag vocabulary with m tags. We explore tag-tag semantic relevance in a tag-specific manner. Each image I_i is labeled with a set of tags, denoted as S_i . For each tag t_i , we find all images from \mathbf{I} that contain t_i . We denote such image set as $U(t_i)$, and a tag-image matrix can be constructed as follows,

$$D_{t_i}(j, k) = \begin{cases} 1 & \text{if image } I_k \text{ contains tag } t_j \\ 0 & \text{otherwise} \end{cases}$$

For each tag t_i , we study $\mathbf{D}_{t_i} \in \mathbf{R}^{m \times n_{t_i}}$, a tag-image matrix with n_{t_i} images in $U(t_i)$ sharing the same tag t_i and all the tags occurred in those n_{t_i} images. $D_{t_i}(j, k)$ denotes the occurrence of tag t_j on image I_k . In Section 2.1, we study the tag-tag text similarity matrix $\mathbf{P}_t^{t_i}$ by Latent Semantic Indexing [1] on tag occurrence. And a tag-tag visual similarity matrix $\mathbf{P}_v^{t_i}$ is formulated by the propagated tag relevance from trustable images in Section 2.2. We further apply integration functions below to obtain global tag-tag visual similarity \mathbf{P}_v and text similarity \mathbf{P}_t

$$\mathbf{P}_t = f_t(\mathbf{P}_t^{t_i}) \quad (1)$$

$$\mathbf{P}_v = f_v(\mathbf{P}_v^{t_i}) \quad (2)$$

$$\mathbf{P} = \lambda \mathbf{P}_v + (1 - \lambda) \mathbf{P}_t \quad (3)$$

Then, we embed the global tag-tag similarity \mathbf{P} into a random walk based framework for image tag re-ranking.

$$\mathbf{R} = \alpha \mathbf{P} \mathbf{R} + (1 - \alpha) \mathbf{V} \quad (4)$$

where \mathbf{R} is the relevance score vector, \mathbf{P} is a transition matrix, P_{ij} indicates the probability of the transition from tag t_i to tag t_j , \mathbf{V} describes the relevance between the image and a labeled tag. α is a damping factor ranging from 0 to 1. The process can be illustrated as each tag t_i gives a score with the probability of $(1 - \alpha)$ or gives a score to other tags based on transition probability P_{ij} . Eventually, it converges and the tags with high relevance score will be re-ranked to the top.

2.1 Latent Semantic Tag-Tag Similarity

Given images sharing the same tag t_i in the form of tag-image matrix \mathbf{D}_{t_i} . We apply Latent Semantic Indexing [1] to estimate the similarity among these tags. More specifically, we do Singular Value Decomposition on \mathbf{D}_{t_i} ,

$$\mathbf{D}_{t_i} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}' \quad (5)$$

where \mathbf{U} and \mathbf{V} are the left and right singular vector, $\mathbf{\Sigma}$ is the diagonal matrix of singular values. By keeping the top k largest singular value in $\mathbf{\Sigma}$, and set the remaining as zeros, we can obtain $\hat{\mathbf{D}}_{t_i}$. The similarity matrix between two terms is $\mathbf{P}_t^{t_i} = \hat{\mathbf{D}}_{t_i} \hat{\mathbf{D}}_{t_i}'$.

We design different integration functions f_t in Eq. (1). $f_t^{(1)}$ is a row-based selection. For tag t_i and its tag-tag similarity matrix $P_t^{t_i}$, we copy the row where t_i lies to the global tag-tag similarity matrix $P_t^{(1)}$. $P_t^{(2)}$ is the summation of $P_t^{t_i}$ with weighting function $f_t^{(2)}$.

$$f_t^{(2)} = \text{sqrt}\left(\log\left(\frac{n}{n_t}\right) \times \frac{n_t}{n}\right) \quad (6)$$

2.2 Tag-Tag Visual Similarity

It is a common way to visually represent a tag by images containing it. It unavoidably introduces noise while comparing visual similarity since images contain multiple objects. Therefore, we divide the images containing the same tag into groups so that images within the same group are more visually similar. For a given tag t_i and D_{t_i} , we propose a three-step approach to collect propagated tags from trustable images:

(1) Clustering images sharing the same tag into groups by Affinity Propagation algorithm [3]. Each group consists of an exemplar image and group member images.

(2) Applying Locality-constrained Linear Coding (LLC) [4] to reconstruct exemplar image by group members. The members with high contribution should be semantically similar to the exemplar, because they lie in the same group sharing the same latent sub-concept and they all strongly agree with the exemplar. We consider those images as trustable images.

(3) Propagating weighted tag information from the trustable images to form tag relevance vector for a certain group. The aggregated tag relevance vectors from groups are used to describe the similarity between t_i and other tags.

As illustrated in Figure 1, all of beach images are first clustered into groups. A red triangle represents the exemplar of each group. The green dots are member images with high contribution for the reconstruction and their tag information will be propagated with certain weight.

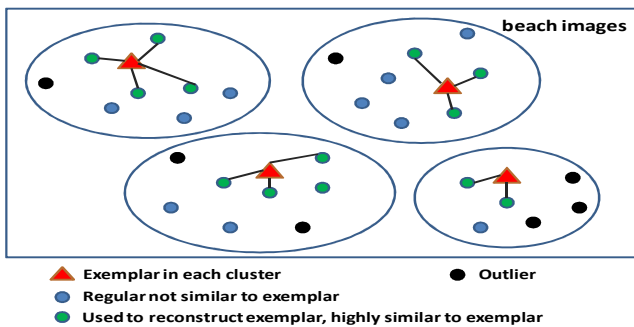


Figure 1. A toy case of reconstructing exemplar by similar images

3. EXPERIMENTS

We evaluate the proposed approaches on NUS-WIDE dataset [5] with 269,648 images. A challenging tag set with 4,926 unique tags is obtained by filtering the noisy tags and tags with less than 30 images in the training set. The ranking accuracy is measured by NDCG@5, which measures the consistency between the top-N ranked tags and the ground truth. To build the ground-truth dataset, 150 images are manually labeled with the relevance for each tag on a scale of 5. For each image, we extract a 712-dimension feature, including 225-dimension block-wise color moment feature and 512-dimension GIST feature. In experiment, we estimate image and tag similarity by kernel density estimation.

We use all images in each group to reconstruct the exemplar image and integrate the tag relevance from all groups with L_2 norm.

Baseline We implement Liu’s method [2] for comparison and use raw data as baseline. In terms of accuracy, Liu’s result is 73.36% and the baseline is 65.98%. Another baseline is to use Google distance in [2] as P_t in PageRank, and its accuracy is 68.37%.

We conduct extensive experiments with different integration schemes and combinations for visual similarity and text similarity in Table 1. P_v , $P_t^{(1)}$ and $P_t^{(2)}$ indicate the performance purely based on itself. ‘G’ refers to the Google distance used in Liu [2]. ‘ P_v+G ’ refers to Google distance combined with our visual tag-tag similarity. $P_t^{(1)}$ and $P_t^{(2)}$ are tag-tag text similarity based on integration functions $f_t^{(1)}$ and $f_t^{(2)}$ mentioned in section 2.1

Table 1. Comparison on accuracy (%)

raw	G[2]	Liu[2]	$P_v + G$	$P_v + P_t^{(1)}$	$P_v + P_t^{(2)}$	P_v	$P_t^{(1)}$	$P_t^{(2)}$
65.98	68.37	73.36	78.47	78.56	78.58	78.37	77.87	77.96

Accuracy As shown in Table 1, our best performance is 78.58%. It improves 5.22% over Liu [2], improves 12.6% over the raw data baseline, and 10.21% over the Google distance. In our approach, we apply all images to reconstruct exemplar image with L_2 norm among groups of image for each tag, $\alpha = 8$, $\lambda = 4$. P_v , $P_t^{(1)}$ and $P_t^{(2)}$ can work well individually, and their combinations also have some improvement.

Efficiency Our approach is 8% faster than Liu’s approach [2], because in Liu’s work the visual similarity is calculated online by K-NN search for every tag. Ours is estimated by the aggregated vote for representative images as well as the weighted image cluster groups. It considers the general cases and is efficient in run-time.

4. CONCLUSION

In this paper, we propose a new approach for image tag re-ranking by exploiting tag-tag relevance using Latent Semantic Indexing on tag occurrence. We also study the tag-tag visual relevance by propagating tags based on their contribution to exemplar reconstruction. We evaluate our approach on a large scale vocabulary dataset with 4,926 tags. Extensive experiments demonstrate the superiority of our approach over others.

5. ACKNOWLEDGEMENT

This work was supported in part to Dr. Qi Tian by ARO grant W911BF-12-1-0057, NSF IIS 1052851, Faculty Research Awards by Google, NEC Laboratories of America and FXPAL, respectively.

6. REFERENCES

- [1] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., Harshman, R., Indexing by latent semantic analysis, *Journal of the American Society for Information Science*, 1990
- [2] Liu, D., Hua, X.S., Yang, L., Wang, M., Zhang, H.J., Tag ranking, In *Proc. of WWW*, 2009
- [3] Frey, B. J. and Dueck, D., Clustering by Passing Messages Between Data Points, *Science*, vol. 315, 2007
- [4] Wang, J., Yang, J., Yu, K., Lv, F., Huang, T. and Gong, Y., Locality-constrained Linear Coding for image classification, In *Proc. of CVPR*, 2010
- [5] Chua, T., Tang, J., Hong, R., Li, H., Luo, Z., and Zheng, Y. NUS-WIDE: a real-world web image database from National University of Singapore, In *Proc. of CIVR*, 2009