Neurocomputing 167 (2015) 614–624

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Image tag-ranking via pairwise supervision based semi-supervised model

Yonghao He *, Cuicui Kang, Jian Wang, Shiming Xiang, Chunhong Pan

Institute of Automation, Chinese Academy of Sciences, No. 95, Zhongguancun East Road, Haidian District, Beijing, China

ARTICLE INFO

Article history:
Received 11 November 2014
Received in revised form
12 April 2015
Accepted 12 April 2015
Communicated by X. Gao
Available online 25 April 2015

Keywords:
Image tag-ranking
Pairwise supervision
Learning to rank

ABSTRACT

Image tag-ranking, the task to sort tags based on their relevance to the related images, has become a hot topic in the field of multimedia. Most existing methods do not incorporate the tag-ranking order information into the models, which is actually very important to solve the issue of image tag-ranking. In this paper, by taking advantage of such important information, we propose a novel model which uses images with ranked tag lists as its supervision information. In the proposed method, each ranked tag list is decomposed into a number of image–tag pairs, all of which are pooled together for training a scoring function. With this pairwise supervision, the model is able to capture the intrinsic ranking structures. In addition, unsupervised data, namely images with unranked tag lists, is also integrated for digging the binary order: relevant or irrelevant. By leveraging both the pairwise supervision and unsupervised structural information, our model sufficiently exploits the tag relevance to images as well as the ranking structures of tag lists. Extensive experiments are conducted on both image tag-ranking and tag-based image search with three benchmark datasets: SUNAttribute, Labelme and MSRC, demonstrating the effectiveness of the proposed model.

1. Introduction

In recent years, Internet users are willing to share their personal information (such as blogs, videos and pictures) and enjoy the information from others at the same time, which brings the prosperity of social networks. In order to take advantage of the uploaded information from users, the providers of social networks encourage users to attach meaningful tags while uploading the associated information. For example, when users share their pictures, system may ask users to type some keywords (tags) or to choose some recommended keywords that best describe the contents of their pictures. By doing this, system can facilitate applications such as image search and interests group recommendation. Actually, these applications can be further improved by utilizing the carefully sorted keywords. For instance, a user who loves cats may upload a picture containing a cute cat walking in the wild, and he may attach the keywords in a random order: “grass”, “tree”, “path”, “cat” and “sky”. One interests group recommendation system that considers the order of keywords is very likely to mis-categorize the user to the groups such as “Nature Photography” by simply analyzing that the first three keywords (“grass”, “tree”, “path”) are closely related to the “nature”, which deviates from the real intention of the user. However, if the uploaded keywords can be properly sorted before being fed into the recommendation system, say “cat”, “path”, “grass”, “tree” and “sky”, it is more likely to categorize the user to the “Cat Fans” group.

The goal of image tag-ranking is to sort tags according to their relevance to the contents of the images. The issue of tag-ranking has been investigated in [1], where the statistics of position distribution of the most important tags is presented: for 1200 images with at least 10 tags randomly selected from Flickr,[1] there are only less than 10% of the images having their most relevant tags at the first place. Furthermore, we also make an analogous analysis on the benchmark datasets, the SUNAttribute [2], Labelme [3] and MSRC [4], which are used in the experiments. Since each tag is assigned to a relevance level from 0 to 3, corresponding to irrelevant to the most relevant, we calculate the average relevance levels of the top tags. In Fig. 1, it is obviously observed that the original average relevance levels of the top tags are lower than the optimal ones, which indicates that there are many highly relevant tags that are not placed at the top positions.

By now, there have been some methods proposed for image tag-ranking. Most of them, such as [1,5–7], are unsupervised methods.[2] In

* Corresponding author.
E-mail address: yhhe@nlpr.ia.ac.cn (Y. He).

http://dx.doi.org/10.1016/j.neucom.2015.04.027
0925-2312/

[1] http://www.flickr.com

[2] In the field of image tag-ranking, “unsupervised” means that the tags of the images are in random orders, not properly ranked. And “supervised” indicates that the tags are ranked according to their relevance to the associated images. We emphasize that the “tags” themselves, existing for both “unsupervised” and “supervised” scenarios, are not related to unsupervised or supervised information in the case that the tags in many classification problems are supervised information.

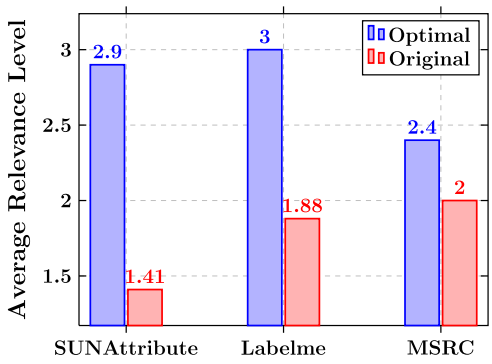

Fig. 1. Average relevance levels at the first position on SUNAttribute, Labelme and MSRC.

practise, unsupervised data is much more than supervised data, which makes researchers put lots of endeavor to utilize unsupervised information. Since supervised data (the images with ranked tag lists) is so limited that there are very few algorithms proposed based on it. For example, [8,9] present a semi-supervised method and a purely supervised method, respectively.

In this paper, a pairwise supervision based semi-supervised model is proposed to address the issue of image tag-ranking (we name it **PSTR**). In the literature, pairwise supervision is well studied in learning to rank (L2R) techniques [10], which motivates the proposed model. It addresses the ranking problem by viewing a ranked list as a number of item pairs to preserve the relative ranking structure, which is mostly ignored by the existing tag-ranking methods. By utilizing the pairwise supervision, the proposed model will gain the global view of ranking structure by decomposing the ranked tag lists into image–tag pairs. Moreover, we also integrate the unsupervised information into our model. The tags in unranked tag lists are deemed as weak ranking information, since we can tell whether the tags are relevant or irrelevant. So the proposed model is semi-supervised by using both the pairwise supervision and unsupervised information. The main contributions of our work are listed below:

- The idea of pairwise supervision is introduced into the proposed model for image tag-ranking. The pairwise supervision treats the tag lists as image–tag pairs—the items in the lists are no longer viewed independently and have mutual interactions, which makes the model capable of predicting the tag relevance by considering all pairs in a list.
- We leverage the unsupervised data which is viewed as the weak ranking information to facilitate the proposed model. The final objective function in the semi-supervised model consists of two components: pairwise supervision item and unsupervised item, and it can be elegantly optimized with a closed-form solution.
- Two experiments (image tag-ranking and tag-based image search) have been carried out to compare the proposed model with state-of-the-art algorithms on three benchmark datasets. The experimental results show that the proposed method can produce better ranked tag lists.

The remainder of this paper is structured as follows. Section 2 briefly summarizes the related work. In Section 3, the motivation deriving from L2R algorithms is first introduced. Then the proposed model is described in detail. Extensive experiments are shown in Section 4. Finally, the conclusions are drawn in Section 5.

2. Related work

The issue of image tag-ranking has attracted remarkable attentions, and various methods have been proposed for it. *Automatic image annotation* [11–15], which automatically assigns meaningful and content-related tags to the corresponding images, provides preliminary insights into this issue. These methods only generate coarse tag lists for the untagged images. They do not consider the orders of the tags. Whereas, image tag-ranking sorts tags in existing tag lists. Following image annotation, *tag refinement* methods, such as [16–18], are required for more precise image–tag association. These methods are built upon image annotation, namely taking the results from automatic image annotation as the initializations, and subsequently explore which tags are more appropriate to be annotated. One problem is that the initial tags may not be correctly provided by image annotation methods, whereas in image tag-ranking tags are all supposed to appropriately describe the images.

To address the issue of image tag-ranking, a number of specifically designed methods have been proposed. As mentioned in Section 1, they are divided into three categories, i.e., unsupervised, semi-supervised and supervised methods.

Unsupervised methods: For tag-ranking, most methods, such as [1,5–7], are designed in the unsupervised fashion. Ref. [1] is the first attempt to address image tag-ranking problem. In [1], Liu et al. assign initial relevance scores to the tags by using kernel density estimation (KDE), and then perform a random walk based refinement on a tag–tag similarity graph. Li et al. [5] propose to learn the tag relevance by neighbor voting. The idea of [5] includes two steps: (1) calculate the image nearest neighbors; (2) accumulate the tag votes within the top K neighbors. The method in [6] assumes that in the view of an image, the image can be represented as a weighted combination of the relevant tags. And for a tag, the tag can be expressed as a weighted combination of the representative images. Then a image–tag correlation matrix is learned under the criterion that two images with high similarity are close in the tag view, and vice versa. The final tag relevance scores are the elements in the image–tag correlation matrix. Differently, Sun et al. [7] use commercial search engines as auxiliaries to collect images for each tag, and then a Bayesian based model is proposed to estimate the initial relevance scores for the tags by using the collected images. Finally, a random walk is performed on the tag graph to refine the tag scores. To address the issue of personalized tag recommendation, Zhao et al. [19] propose a graph based ranking method, leveraging the benefit of traditional manifold ranking. This method can achieve good performance to recommend tags for users. Thus, we can see that these unsupervised methods rely on the tag relevance propagation via visual similarity, but ignoring the ranking structure information within the tag lists.

Semi-supervised methods: To our knowledge, [8] is the only method in the semi-supervised fashion. The aim of this method is to obtain a projection matrix that projects visual features to tag relevance space. The supervised component is formulated as the linear regression between the tag relevance scores and the projected image visual features. The unsupervised component is a regularizer that restricts the large relevance scores only appearing for the tags that are annotated to the associated images. Finally, it results in a quadratic programming problem. In our view, a single linear projection cannot capture complicated relationships between visual feature space and tag relevance space, and the tag-biased regularization is also not related to the inner ranking structures among the tag lists.

Supervised methods: Lan et al. [9] propose a Max-Margin Rifled Independence Model for tag-ranking. The main idea is that the max-margin formalism with riffled independence factorization proposed in [20] can perform structure learning. Therefore, this model can predict the tag orders in the tag permutation space.

Besides the above methods, *learning to rank* (L2R) techniques [10] have the potential to accomplish the task of image tag-ranking. However, there is no existing methods utilizing L2R to