Signal Processing: *Image Communication* 34 (2015) 61–71

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Personalized image annotation via class-specific cross-domain learning

Zhiming Qian*, Ping Zhong, Runsheng Wang

College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, Hunan, China

ARTICLE INFO

Article history:
Received 29 September 2014
Received in revised form
31 January 2015
Accepted 23 March 2015
Available online 31 March 2015

Keywords:
Personalized image annotation
Cross-domain learning
Nearest neighbor model
Multiple kernel learning

ABSTRACT

In this paper, we propose to learn users' own profiles for image annotation with the purpose of facilitating image searching towards users' intentions. Considering that the size of a user's annotation vocabulary is usually small and different users have different visual understanding towards a specific tag, we perform personalization in a class-specific form to summarize a user's annotation profiles for each tag in his vocabulary. In particular, we first exploit a generic annotation dataset with a class-specific weighted nearest neighbor (cs-WNN) model by combining the techniques of multiple kernel learning and nearest neighbor modeling. Next, a new personalization method, namely class-specific cross-domain learning (cs-CDL), is proposed to achieve users' own annotation profiles (i.e. the user-specific parameters of cs-WNN models) by exploiting users' annotation datasets. Experimental results have been reported over several challenging image databases to validate the effectiveness of the proposed method for both generic and personalized image annotation.

1. Introduction

Rapid technological advances in image capturing and sharing have resulted in the explosive growth of personal and web image collections, and consequently, how to search these images according to a user's intention is a key problem for improving many relevant applications, such as searching sharing photos from circles of friends, advertising users of product photos, and so on. In practice, most users prefer to search images using textual queries other than image examples due to the fact that textual descriptions are much easier to express users' intentions than visual descriptions, encouraging the research on image annotation which aims to learn dependencies between images and tags. However, image annotation is also a challenging task related to understanding what we see in a visual scene due to the semantic gap between low-level visual features and high-level semantics [1].

Much work [2–5] has been proposed to bridge the semantic gap by exploiting large quantities of well-labeled images. Unfortunately, these efforts have completely ignored the user interact. Clearly, users have their own labeling preferences due to the diversity of their interests. For example, a traveler might label the images in Fig. 1 with 'Amsterdam' or 'Cologne', while an architect prefer to label them with 'Gothic' or 'Bridge'. This suggests that an off-the-shelf image annotation system is unlikely to be universally acceptable by different users. So how we can utilize users' image datasets and their labeling histories to learn their annotation profiles is the key point we investigate. Previous work [6–8] on this line mostly focuses on learning users' annotation models on their own datasets. However, the problems of cold start for users (i.e. new users have little historical behavior) and the sparsity of users' datasets (i.e. the low proportion of a user's images in

* Corresponding author. Tel.: +86 18573107036; fax: +86 84575492.
E-mail addresses: qianzhiming@nudt.edu.cn (Z. Qian), zhongping@nudt.edu.cn (P. Zhong), rswang@nudt.edu.cn (R. Wang).

http://dx.doi.org/10.1016/j.image.2015.03.008
0923-5965/

Fig. 1. Example images from Flickr and their associated tags.

the generic image dataset) have been increasingly intractable for these methods [9]. To address these problems, personalized methods [1,10] learn users' own annotation profiles by personalizing annotation models using both generic and user-specific data statistics. However, these methods haven't quantified the difference between these two data statistics, resulting in the nonoptimal solution of personalization. Recently, cross-domain learning techniques [11,12] are proposed to solve the above cold start and sparsity problems by leveraging the difference between a generic domain and users' specific domains, allowing users to build their own modeling profiles using relatively more abundant data. For personalized image annotation, there also exist domain variations between generic and users' image datasets, and thus cross-domain learning techniques can also be beneficial for personalization.

In this paper, we present a method of cross-domain learning, namely class-specific cross-domain learning (cs-CDL), to personalize a class-specific weighted nearest neighbor (cs-WNN) model for learning users' own annotation profiles. The nearest neighbor tag transfer mechanism introduced in [3] are widely used in inferring tags from labeled data to unlabeled data due to its efficiency of dealing with the increasing amount of training data which can be easily obtained from some social websites (e.g. Flickr). Considering that different users have different visual understanding towards a specific tag, we extend the nearest neighbor model to the cs-WNN model by introducing class-specific modeling and multiple kernel learning. By personalizing the above cs-WNN models for tags in users' own vocabularies with cs-CDL, we can achieve the optimal annotation models for different users to reveal the diversity of users' visual understanding. Besides, we also consider the correspondences among different tags, and present a smoothing method based on nonnegative matrix factorization (NMF) for achieving more coherent labeling information. Two core contributions of this paper are summarized as follows:

(1) We first present the cs-WNN model for generic image annotation by directly maximizing the log-likelihood of class-specific tag predictions. This model integrates class-specific modeling and multiple kernel learning, and thus it can effectively learn the nonlinearity of features in the visual space of each tag.
(2) We then present cs-CDL to build users' own annotation profiles that correspond to the user-specific parameters of cs-WNN models. Because only the specific modeling parameters with respect to the tags in users' own vocabularies are involved in personalization, cs-CDL is more effective than the methods that perform personalization for all tags in the generic vocabulary.

The remainder is organized as follows. We first discuss related work in Section 2. Next, Section 3 illustrates the framework of the proposed method for both generic and personalized image annotation. Experimental results and conclusions are provided in Sections 4 and 5, respectively.

2. Related work

In this section, we outline research contributions that are most relevant to our work, including recent progresses on both generic and personalized image annotation.

2.1. Generic image annotation

For generic image annotation, there are two well-accepted groups: generative methods and discriminative methods. By predicting the joint probability of visual features and tags from unlabeled images, generative methods [2,13] have shown significant scalability in the number of tags of interest and provide a natural ranking of tags for each image to be annotated. However, these methods are usually not rich enough to accurately capture the intricate dependencies between images and tags, which are easily disturbed by images with high visual similarities but different semantics. On the other hand, by taking each tag as a class, the discriminative methods [5,14] have also been used to address the problem of image annotation due to their strong classification power. The scalability of discriminative methods is not satisfactory, however, when coping with a large number