Knowledge-Based Systems 72 (2014) 72–80

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Semi-supervised learning for refining image annotation based on random walk model

Dongping Tian *

Institute of Computer Software, Baoji University of Arts and Sciences, Baoji, Shaanxi 721007, PR China
Institute of Computational Information Science, Baoji University of Arts and Sciences, Baoji, Shaanxi 721007, PR China

A R T I C L E I N F O

Article history:
Received 12 October 2013
Received in revised form 8 August 2014
Accepted 29 August 2014
Available online 9 September 2014

Keywords:
Automatic image annotation
Semi-supervised learning
Gaussian mixture model
Expectation-maximization
Random walk
Image retrieval

A B S T R A C T

Automatic image annotation has been an active research topic in recent years due to its potential impact on both image understanding and semantic based image retrieval. In this paper, we present a novel two-stage refining image annotation scheme based on Gaussian mixture model (GMM) and random walk method. To begin with, GMM is applied to estimate the posterior probabilities of each annotation keyword for the image, during which a semi-supervised learning, i.e. transductive support vector machine (TSVM), is employed to enhance the quality of training data. Next, a label similarity graph is constructed by a weighted linear combination of label similarity and visual similarity of images associated with the corresponding labels. In this way, it can seamlessly integrate the information from image low-level visual features and high-level semantic concepts. Followed by a random walk process over the constructed label graph is implemented to further mine the correlation of the candidate annotations so as to capture the refining results, which plays a crucial role in semantic based image retrieval. Finally, extensive experiments carried out on two publicly available image datasets bear out that this approach can achieve marked improvement in annotation performance over several state-of-the-art methods.

1. Introduction

Automatic image annotation (AIA) is a promising solution to enable the semantic based image retrieval via keywords. It is generally believed that AIA refers to a process to automatically generate textual words to describe the content of a given image. In recent years, the state-of-the-art research on AIA has broadly proceeded along two categories. The first one views image annotation as a supervised classification problem [1] that treats each semantic concept as an independent class and constructs different classifiers for different concepts. This approach predicts the annotations of a new image by computing similarity at the visual level and propagating the corresponding keywords. The second category treats the words and visual tokens in each image as equivalent features in different modalities. Image annotation is then formalized by modeling the joint distribution of visual and textual features on the training data and predicting the missing textual features for a new image. As the representative work of this perspective, Duygulu et al. [2] propose a translation model (TM) to treat AIA as a process of translation from a set of blob tokens, obtained by clustering image regions, to a set of keywords. Jeon et al. [3] put forward cross-media relevance model (CMRM) to annotate images, assuming that the blobs and words are mutually independent given a specific image. Subsequently, CMRM is improved through continuous-space relevance model (CRM) [4], multiple-Bernoulli relevance model (MBRM) [5] and dual cross-media relevance model (DCMRM) [6], etc. In addition, several nearest-neighbor-based methods have also been proposed in the most recent years [7,8]. Despite most of these methods have achieved encouraging results, there are still two problems remain to be solved. First, in most cases, labeled images are often hard to obtain or create in large quantities while the unlabeled ones are easier to collect. So how to efficiently use the unlabeled images to improve the annotation performance is a key issue to formulate effective semantic models. Second, little effort focuses on the semantic context and semantic correlation between image annotations. Even if considered, the image visual information related to the annotation often tends to be ignored, which is liable to cause the phenomenon that different images with the same candidate annotations[1] would obtain the same refinement results.

To address the above issues, we present a novel two-stage refining image annotation scheme based on Gaussian mixture

* Address: Institute of Computer Software, Baoji University of Arts and Sciences, No. 44, Baoguang Road, Baoji, Shaanxi 721007, PR China.

[1] Candidate annotations denote the initial annotations generated by some image annotation models.

http://dx.doi.org/10.1016/j.knosys.2014.08.023
0950-7051/

model and random walk. On the one hand, GMM is employed to estimate the posterior probabilities of each annotation keyword for the image, which can be seen as the initial annotation stage. It is worth noting that a semi-supervised learning, viz. transductive support vector machine, is introduced into GMM to enable the unlabeled images to be fully exploited to improve the GMM performance to some degree. On the other hand, a random walk process over the constructed label graph is implemented to further mine the correlation of the candidate annotations for precise annotation, which can be regarded as the refining annotation stage. Experimental results on two publicly available image datasets validate the effectiveness of the proposed approach. To the best of our knowledge, this study is the first attempt to integrate GMM with random walk as well as semi-supervised learning in the task of refining image annotation.

The rest of this paper is organized as follows. Section 2 discusses some related work about image annotation. In Section 3, Gaussian mixture model is first introduced, and then the TSVM as well as random walk is elaborated respectively, especially for the proposed SGMM-RW refining annotation framework. Experimental results are reported and analyzed in Section 4. Finally, the paper is ended with some important conclusions and future work in Section 5.

2. Related work

During recent years, many methods have been developed for refining image annotation. As a pioneer work, Jin et al. [9] employ WordNet to estimate the semantic correlation between annotation keywords. This method, however, can only achieve a fairly limited effect as it totally ignores the visual content of images. In the approach [10], Wang et al. apply random walk with restarts to refine candidate annotations by integrating word correlations with the original candidate annotation confidences together. Followed by they propose a content based approach by formulating annotation refinement as a Markov process [11]. Later on Jin et al. [12] exploit randomized weighted-max cut algorithm to complete image annotation refinement. Subsequently it is extended by a new methodology for knowledge based refining image annotation in a deterministic polynomial time [13], in which they further investigate various semantic similarity measures between keywords and fuse the outcomes of all these measures together to make a final decision by using Dempster–Shafer evidence combination. Liu et al. [14] rank the image tags according to their relevance with respect to the associated images by tag similarity and image similarity in a random walk model. Xu et al. [15] come up with a new graphical model termed as regularized latent Dirichlet allocation (rLDA) for tag refinement. In recent work [16], an efficient iterative approach is put forward for image tag refinement by pursuing low-rank, content consistency, tag correlation and error sparsity through solving a constrained yet convex optimization problem. Especially in our previous work [17,18], a unified refining image annotation technique is proposed by combining probabilistic latent semantic analysis (pLSA) with random walk/max-bisection model respectively, experiments on several image datasets have validated its computational efficiency and annotation accuracy. In more recent work [19], a survey on refining image annotation techniques is reviewed. Particularly the key aspects of various methods, including their original intentions and annotation models, are comprehensively summarized and analyzed. For more details please refer to the corresponding literature.

Alternatively, semi-supervised learning has been an active topic of research in computer vision and machine learning for decades [20,21]. As the representative work, Li and Sun [22] formulate image annotation problem as a joint classification task based on two-dimensional conditional random fields together with semi-supervised learning, in which the semi-supervised learning is utilized to exploit the unlabeled data to improve the joint classification performance. In [23], a semi-supervised ensemble of classifiers is constructed based on the AdaBoost and naive Bayes is leveraged as its base classifiers. One of the main advantages is that the weights of unlabeled instances are dynamic and proportional to the probability given by the previous stage. Besides, TSVM-HMM [24] is proposed for automatic image annotation by integrating the discriminative classification with the generative model to mutually complete their advantages for better annotation performance. Zhu and Liu [25] conduct image annotation based on a semi-supervised learning model and random walk with restart algorithm so as to well integrate the information of both candidate annotations and the corpus. Recently, Shao et al. [26] put forward a semi-supervised topic modeling for image annotation by introducing a harmonic regularizer based on the graph Laplacian of the data into the probabilistic semantic model for learning latent topics of the images. Ismail and Bchir [27] present a semi-supervised possibilistic clustering and feature weighting algorithm for AIA. More recently, to make up for the drawback of ignoring manifold structures revealed by the unlabeled data, Yuan et al. [28] propose a semi-supervised cross-domain learning method with group sparsity for image annotation by applying both labeled and unlabeled training data with their manifold structural information. Meanwhile, a multi-view semi-supervised bipartite ranking model, which allows using the information contained in unlabeled sets of images, is proposed for refining image annotation [29]. Following this work, Chen et al. [30] treat image annotation as a multi-instance semi-supervised learning problem by constructing a graph based semi-supervised learning classifier to produce several keywords for each unseen image. Table 1 summarizes several automatic image annotation methods reported in the literatures.

3. Framework of the SGMM-RW model

In this paper, we propose an approach for refining image annotation by integrating Gaussian mixture model (GMM) with random walk (abbreviated as SGMM-RW). This method mainly involves two stages. Specifically, Gaussian mixture model is leveraged to estimate the posterior probabilities of each annotation keyword for the image in the first stage, during which TSVM is applied to enhance the quality of training data. In the second stage, random walk is employed over the constructed label graph to further mine the correlation of the candidate annotations for precise results. To summarize, the novelty of SGMM-RW lies in two aspects: exploiting GMM to accomplish the initial semantic annotation task and implementing random walk process over the constructed label similarity graph to refine the candidate annotations generated by the GMM. Fig. 1 illustrates the framework of SGMM-RW proposed in this paper.

3.1. Gaussian mixture model

A Gaussian mixture model (GMM) is a parametric statistical model which assumes that the data originates from a weighted sum of several Gaussian sources. More formally, a GMM is a weighted sum of M component Gaussian densities as given by the following equation.

$$p(x|\lambda) = \sum_{i=1}^{M} \omega_i g(x|\mu_i, \Sigma_i) \tag{1}$$

where x is a D-dimensional continuous-valued data vector, w_i, $i = 1, 2, \ldots, M$, denote the mixture weights and satisfy the constraint $\sum_{i=1}^{M} \omega_i = 1$, $g(X|\mu_i, \Sigma_i)$, $i = 1, 2, \ldots, M$, are the compo-