Information Sciences 262 (2014) 159–171

Contents lists available at ScienceDirect

Information Sciences

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

Effective automatic image annotation via integrated discriminative and generative models

Mei Wang [a,*], Xiaoling Xia [a], Jiajin Le [a], Xiangdong Zhou [b]

[a] School of Computer Science and Technology, Donghua University, China
[b] School of Computer Science, Fudan University, China

ARTICLE INFO

Article history:
Received 28 December 2012
Received in revised form 17 September 2013
Accepted 7 November 2013
Available online 5 December 2013

Keywords:
Automatic image annotation
Generative model
Discriminative model
Discriminative hyperplane tree
Hierarchical classification

ABSTRACT

In this paper, we present a novel image annotation method that leverages on the advantages of both generative and discriminative models. To label an image, we first identify a visual neighborhood in the training image set based on generative approach. Then, the neighborhood is refined by an optimal discriminative hyperplane tree classifier based on concept feature. The tree classifier is built according to a local topic hierarchy, which is adaptively constructed by exploiting the semantic contextual correlations of the corresponding visual neighborhood. Experiments conducted on the ECCV2002 and TRECVID 2005 benchmarks demonstrate the effectiveness and efficiency of the proposed method.

1. Introduction

Automatic Image Annotation (AIA) refers to the association of a predefined list of keywords or captions to images. It is a key step towards semantic keyword based image retrieval, which is considered to a convenient and easy way for retrieving images on the web. The early annotation approaches rely on human annotators, which suffers from the problems of labor intensity and subjectivity. With the rapid growth of image archives, the need of AIA becomes increasingly important to facilitate semantic keyword based image retrieval.

Many previous works have been conducted on AIA. One of the most extensively researched approaches is the generative model based image annotation, such as the relevance models [16,8,19]. By predicting the joint probability of semantic keywords and visual features from unlabeled images, the generative model based methods have shown significant scalability in the number of keywords of interest, and provided a natural ranking of keywords for each image to be annotated. However, due to the "semantic gap", such an "unsupervised labeling process" is easily disturbed by images with high visual similarities but different semantics [41]. On the other way, by taking each semantic keyword as a class, the discriminative techniques such as SVM have also been applied to address the problem of image annotation [9,41] due to their strong discrimination/classification power. However, their scalability to a large number of classes is not satisfactory. It is well known that there are some complementary properties between the generative models and the discriminative models. Therefore, recent research efforts focus on exploring new learning methods that leverage on the advantages of both the generative and discriminative models [18,11]. Lasserre et al. [18] propose a principled approach to blend generative and discriminative approaches.

* Corresponding author. Tel.: +86 21 67792164.
E-mail addresses: wangmei@dhu.edu.cn (M. Wang), sherlysha@dhu.edu.cn (X. Xia), lejiajin@dhu.edu.cn (J. Le), xdzhou@fudan.edu.cn (X. Zhou).

0020-0255/$ - see front matter © 2014 Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.ins.2013.11.005

Grabner et al. [11] propose eigenboosting method to combine discriminative and generative information. However, the semantic overlaps and data imbalance among different semantic classes induced by multi-label characteristics of image annotation heavily impair the classification power of the discriminative method, and present a challenge in combining it with generative model for AIA.

Semantic hierarchies have been shown to be useful for AIA [30,29,7]. Srikanth et al. [30] exploited the hierarchical dependencies between annotation words derived from WordNet to generate improved visual lexicons for the translation-based approaches. Shi et al. [29] built a multi-level concept hierarchy from lexical resource WordNet and proposed a Bayesian learning framework of hierarchical multinomial mixture model (BHMMM) for automatic image annotation. In most existing work, the semantic hierarchy (ontology) is useful to establish relationships between annotation words at different abstraction level. Fig. 1(left) shows an example in the predefined WordNet-based hierarchy, in which "bus" and "train" have close relationship because they both belong to the lexical abstract concept C_1 "vehicle". The common visual features or similar contextual information shared in the sibling concepts are usually exploited to improve image annotation [29,7]. However, the concepts and categories in a predefined ontology are sometimes not sufficient and adaptive enough to *fully* describe the diversity of the *semantic topics or scenes* of the real world images. As Fig. 1(right) shows, the complete semantics of the images are better represented by a set of semantically correlated keywords, named latent topic, learned from the real data set adaptively. For example, the semantics of image 1 is more completely represented by the latent topic T_1 jointly defined by the frequently co-occurring words "bus", "building" and "street" rather than using the abstract concept "vehicle". Similarly, the semantic of image 2 is better modeled by topic T_2. As we expect the natural separation between different latent topics to be greater than that between the WordNet-based abstract concepts, we expect images under different latent topics to be more effectively classified by discriminative learning approaches. Obviously, such co-occurrence information is hard to obtain from the WordNet-based hierarchy directly. Hence we demonstrate that the latent topic hierarchy can be learned from the training data adaptively.

In this paper, we propose a novel image annotation method that integrates a discriminative approach into a traditional generative model based annotation framework. To achieve this, we first build a generative model based on visual features to identify visual neighborhood of images with high visual generative probabilities in the training space. We then exploit the contextual correlation between the semantic labels of training images in the visual neighborhood to build a local topic hierarchy. We apply SVM at each level of the topic hierarchy to learn an optimal discriminative hyperplane to build our hyperplane tree classifier, from which we derive the discriminate factor through the top-down classification process. The discriminate factor is used to progressively refine the obtained visual generative neighborhood and improve the final annotation estimation. We believe that our approach effectively exploits both visual and semantic knowledge of training corpus and readily integrates the advantages of generative model with strong discriminative capability.

The main contributions of this paper are as follows. (1) We propose a new image annotation method by integrating generative and discriminative models. A visual-based generative model is used to build a visual neighborhood, while the semantic label correlation is exploited to build the local topic hierarchy for use in a discriminative process. (2) Instead of training classifier for each semantic keyword, our hyperplane tree classifier is learned based on the topic hierarchy to achieve stronger discriminative power. The experimental results on the ECCV2002 benchmark [6] and the TRECVID 2005 benchmark [27] show that our method outperforms the state-of-the-art generative model based and discriminative model based annotation methods.

The rest of the paper is organized as follows. Section 2 introduces the related work in image annotation. Section 3 shows our motivation. Section 4 presents the proposed annotation method. We discuss the results of our experiments in Section 5, and conclude the paper in Section 6.

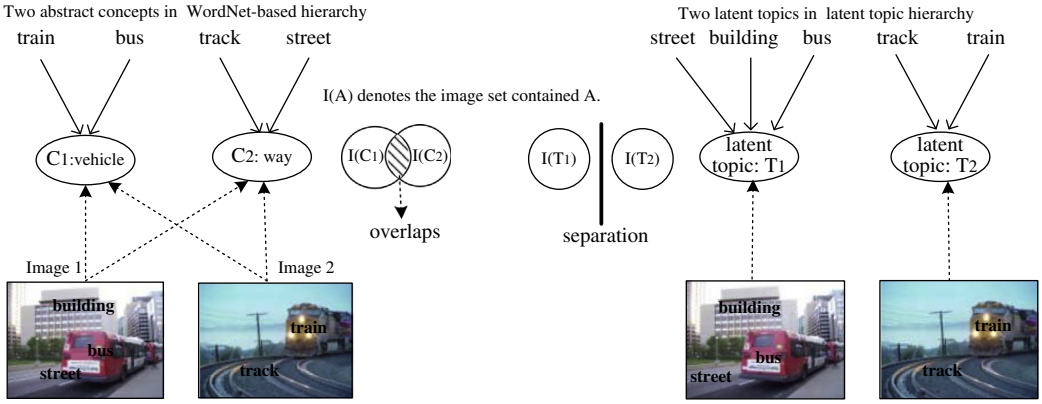

Fig. 1. The examples of the WordNet-based hierarchy and the latent topic hierarchy.