Information Sciences 281 (2014) 648–660

Contents lists available at ScienceDirect

Information Sciences

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

Automatic image annotation by semi-supervised manifold kernel density estimation

Ping Ji [a,*], Na Zhao [b], Shijie Hao [b], Jianguo Jiang [b]

[a] Department of Electronic Information and Electrical Engineering, Hefei University, Hefei 230601, PR China
[b] School of Computer and Information, Hefei University of Technology, Hefei 230009, PR China

ARTICLE INFO

Article history:
Available online 11 September 2013

Keywords:
Image annotation
Semi-supervised learning
Kernel density estimation
Manifold

ABSTRACT

The insufficiency of labeled training data is a major obstacle in automatic image annotation. To tackle this problem, we propose a semi-supervised manifold kernel density estimation (SSMKDE) approach based on a recently proposed manifold KDE method. Our contributions are twofold. First, SSMKDE leverages both labeled and unlabeled samples and formulates all data in a manifold structure, which enables a more accurate label prediction. Second, the relationship between KDE-based methods and graph-based semi-supervised learning (SSL) methods is analyzed, which helps to better understand graph-based SSL methods. Extensive experiments demonstrate the superiority of SSMKDE over existing KDE-based and graph-based SSL methods.

1. Introduction

The rapidly increasing large-scale image data makes their effective management [19,13,48] and accessing [27] highly desired. Metadata have shown their superiority in image representation at syntactic and semantic levels. These metadata can be used for image retrieval, summarization and indexing. To generate the metadata, automatic annotation is an elementary step, which can be formulated as a classification task and accomplished by a learning-based method. More specifically, statistical models are usually built based on pre-labeled data to accomplish the task. However, manually labeling images is usually a time-consuming and labor-intensive process. This brings about the problem of training data insufficiency in practice, which thus leads to inaccurate annotation results.

Extensive research efforts have been dedicated to automatic image annotation, and semi-supervised learning (SSL) methods have recently shown great potential in solving the problem of training data insufficiency. Essentially, automatic image annotation can be formulated as a SSL [11,37,57] task, in which only a small proportion of images are labeled while the rest are left for label prediction. By leveraging a large amount of unlabeled data based on certain assumptions, SSL methods are expected to build more accurate models than those built based on purely supervised methods.

Recently, graph-based SSL methods that benefit from label smoothness assumption have been introduced [6,56,59]. These methods define a graph where the vertices are labeled/unlabeled samples and the edges reflect the similarities between vertex pairs. A labeling function is then estimated on the graph. The label smoothness over the graph is characterized in a regularization framework, which is composed of a regularization term and a loss function term. Graph-based SSL algorithms have shown encouraging performance in many machine learning and multimedia applications [60,61,64,63], in particular when labeled data are extremely limited. However, we have to notice several issues that are still not clear enough for graph-based SSL methods. First, a supervised version of graph-based SSL sometimes outperforms its

* Corresponding author. Tel.: +86 13855147539.
E-mail addresses: jiping.hfuu@gmail.com (P. Ji), hfut.hsj@gmail.com (S. Hao), jgjiang@hfut.edu.cn (J. Jiang).

0020-0255/$ - see front matter
http://dx.doi.org/10.1016/j.ins.2013.09.016

semi-supervised version, i.e., unlabeled samples may degrade the performance of generative SSL methods [14]. Second, graph-based SSL models are generally considered to be transductive and they cannot be easily extended to out-of-sample data. Several different approaches have been proposed to tackle this problem [15,51]. But it is not yet clear that which way is optimal. Third, as a graph edge measures the similarity between two vertices, several novel graph construction strategies, in addition to using Euclidean distance, can be used to improve classification performance [39,45]. From these above issues, we can observe that further improvements are still needed for the existing methods.

Kernel Density Estimation (KDE) is a non-parametric density estimation approach, which avoids the model assumption problem [62]. Recently, semi-supervised KDE (SSKDE) has been proposed to investigate both labeled and unlabeled multimedia data [46]. Furthermore, an improved SSKDE, which adaptively estimates kernel density, is introduced in [47]. These methods are limited by the isotropic nature of the chosen Gaussian window. Despite the globally non-linear structure of feature space in the multimedia scenario, it is plausible to assume that feature vectors can be embedded in a locally linear subspace. By considering this structure, a Manifold KDE is proposed in [41] to generate much smoother classification boundaries, whereas the previous KDE methods usually introduce boundaries with "hole" or "zig-zag" artifacts.

In this paper, we extend the Manifold KDE to a novel method called Semi-Supervised Manifold KDE (SSMKDE). The method combines the strengths inherited from SSL and KDE, and thus addresses both problems of training data insufficiency and model assumption mentioned above. We also show that several different graph-based SSL algorithms can be derived from SSKDE. Therefore, KDE can be viewed as the supervised version of graph-based SSL methods. We employ the proposed SSMKDE for image annotation and experiments demonstrate the effectiveness of our algorithm.

The contributions of our work are as follows. First, SSMKDE leverages both labeled and unlabeled samples and formulates all data in a manifold structure, which improves annotation accuracy. Second, we analyze the relationships between KDE-based methods and graph-based SSL methods, which are helpful in better understanding graph-based SSL methods.

For clarity, we organize the rest of this paper as follows. Section 2 briefly introduces related work. In Section 3, we introduce the proposed semi-supervised manifold KDE for automatic image annotation. In Section 4, we provide a discussion on the relationships between KDE-based methods and graph-based SSL methods. Experiments are provided in Section 5, followed by concluding remarks in Section 6.

2. Related work

2.1. Image annotation

Generally, image annotation approaches can be divided into three paradigms, i.e. generative models, discriminative models and nearest neighbor based methods [22]. Generative models can be further categorized into topic models [4,31,49] and mixture models [8]. Discriminative models learn a separate classifier for each class [21]. As for nearest neighbor based methods, various kinds of features are combined (called Joint Equal Contribution, JEC) to describe the similarities between images, and a simple kNN-based keyword transferring method is used [30]. Instead of the JEC strategy, TagProp [22] learns the combination weights of the feature groups and a word-specific model for each keyword is learned. Recently, the family of sparsity methods is also widely employed in image annotation. An input and output structural grouping sparsity is introduced into a regularized regression model for image annotation in [23]. Zhang et al. [52] introduce a regularization-based feature selection algorithm to leverage both the sparsity and clustering properties of features. All the above-mentioned research is assumed to have sufficient labeled training data. However, when training data are insufficient, the performance of these methods may severely degrade.

2.2. Semi-supervised learning

As we stated above, large collection of unlabeled images and the high cost of manual labeling trigger the research on semi-supervised learning methods [42–44]. Although there are large bodies of SSL research such as self-training [35], co-training [7], transductive SVM [53], and graph-based methods [59] (for in-depth reading, literatures such as [11,37,57] are recommended), many of them are computationally expensive [53] or ineffective when the assumed models are inaccurate [14]. So we argue that new models that reveal the implicit structures of multimedia data should be developed. Recently, Wang et al. propose semi-supervised kernel density estimation (SSKDE), in which both labeled and unlabeled data are leveraged to estimate class conditional probability densities. Shao et al. [38] propose a semi-supervised topic model for image annotation, in which a harmonic regularization based on the graph Laplacian is introduced into the probabilistic semantic model. Zhao et al. [55] build a cooperative kernel sparse representation (SR) method for image annotation with co-training two SRs in the kernel space. In [50], the authors propose a semi-supervised long-term Relevance Feedback (RF) algorithm to refine the multimedia data representation. The proposed long-term RF algorithm utilizes both the multimedia data distribution in multimedia feature space and the history RF information provided by users. Zhang et al. [54] propose a generic framework for video annotation via semi-supervised learning. A Fast Graph-based Semi-Supervised Multiple Instance Learning (FGSSMIL) algorithm, which aims to build a generic framework for various video domains, is proposed. These works inspire us that SSL methods can be elegantly incorporated into a multimedia annotation framework at various aspects.