Neurocomputing 139 (2014) 202–219

Contents lists available at ScienceDirect

Neurocomputing

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

Sparsity regularization label propagation for domain adaptation learning

JianWen Tao [a,*], Wenjun Hu [b], Shitong Wang [c]

[a] School of Information Science and Engineering, Ningbo Institute of Technology, Zhejiang University, Ningbo 315100, China
[b] School of Information and Engineering, Huzhou Teachers College, Huzhou 313000, China
[c] School of Digital Media, Jiangnan University, Wuxi, China

ARTICLE INFO

Article history:
Received 25 August 2013
Received in revised form
29 December 2013
Accepted 11 February 2014
Communicated by Deng Cai
Available online 12 April 2014

Keywords:
Domain adaptation learning
Sparse representation
Label propagation
Maximum mean discrepancy
Multiple kernel learning

ABSTRACT

Recently, domain adaptation learning (DAL) has shown surprising performance by utilizing labeled samples from the source (or auxiliary) domain to learn a robust classifier for the target domain of the interest which has a few or even no labeled samples. In this paper, by incorporating classical graph-based transductive SSL diagram, a novel DAL method is proposed based on a sparse graph constructed via kernel sparse representation of data in an optimal reproduced kernel Hilbert space (RKHS) recovered by minimizing inter-domain distribution discrepancy. Our method, named as Sparsity regularization Label Propagation for Domain Adaptation Learning (SLPDAL), can propagate the labels of the labeled data from both domains to the unlabeled one in the target domain using their sparsely reconstructed objects with sufficient smoothness by using three steps: (1) an optimal RKHS is first recovered so as to minimize the data distributions of two domains; (2) it then computes the best kernel sparse reconstructed coefficients for each data point in both domains by using l_1-norm minimization in the RKHS, thus constructing a sparse graph; and (3) the labels of the labeled data from both domains is finally propagated to the unlabeled points in the target domain over the sparse graph based on our proposed sparsity regularization framework, in which it is assumed that the label of each data point can be sparsely reconstructed by those of other data points from both domains. Furthermore, based on the proposed sparsity regularization framework, an easy way is derived to extend SLPDAL to out-of-sample data. Promising experimental results have been obtained on both a serial of toy datasets and several real-world datasets such as face, visual video and text.

1. Introduction

Constructing learning algorithms for data that may not be independent and identically distributed (i.i.d.) is one of the newly emerging research topics in data mining and machine learning [1,2]. To address the non-i.i.d problem, domain adaptation learning (DAL) [3] was developed to label the target domain data by using some other labeled source domain data. It arises when the data distribution in the testing domain is different from that in the training domain. The need for domain adaptation is prevalent in many real-world applications. For example, spam filters can be trained on some public collections of spam and ham emails. However, when applied to an individual's email box, one should personalize the spam filters to fit the individual's own distribution of email data in order to achieve better performance. Although such kind of domain adaptation problem is fundamental to machine learning, it only started gaining much attention very recently in the context of transfer learning [4,5]. Recently, a lot of methods have been proposed to solve the domain adaptation problems in video concept detection [6], text classification [7], face recognition, image notation and ranking learning [8].

By sufficiently exploiting existing machine learning algorithms, one of major problems in DAL is how to reduce the divergence of distributions between source and target domains [44]. Thus, the first important target of DAL is to efficiently measure the distribution discrepancy between source and target domains by recovering a function from a given space of functions in a reproducing kernel Hilbert space (RKHS). There has witnessed several research works aiming to measure the distance between distributions of both domains [5,6]. Since some specific feature representation would be able to reduce the distribution discrepancy between two domains as much as possible, discovering a good feature representation across domains is crucial in DAL [2]. For instance, Bliter et al. [3] showed that the performance of the hyper-plane classifier that could best separate the data could provide a good method for measuring distribution distance for different data representations.

* Corresponding author.
E-mail address: jianwen_tao@aliyun.com (J. Tao).

http://dx.doi.org/10.1016/j.neucom.2014.02.044
0925-2312/

Along this same line, Gretton et al. [26] showed that for a given class of functions, the measure could be simplified by computing the discrepancy between two means of the distributions in a reproducing kernel Hilbert space, thus resulting in the maximum mean discrepancy (MMD) measure criterion. The particular form of this criterion makes it easier to be incorporated into the corresponding optimization problems. Based on MMD measure criterion, several DAL methods have been recently proposed in the literature [5].

In the past several years, the graph-based semi-supervised learning (SSL) [11,12] approach has been becoming a hot topic due to its elegant mathematical formulation and its distinctive effectiveness in combining labeled and unlabeled data through label propagation [13–16]. Generally, there are two types of SSL tasks on graphs: (1) transductive SSL [10], which aims at predicting the labels of the unlabeled vertices only, and (2) inductive SSL [11], which tries to induce a decision function that has a low error rate on the whole sample space. In general, the existing SSL methods work well only under a common hypothesis that the training and testing data are drawn from the same distribution and/or feature space. When the distribution has changed, the models learned from the prior information need to be reconstructed using the newly available training data. In DAL, if we ignore the domain difference and treat the labeled source domain instances as labeled data and the unlabeled target domain instances as unlabeled data, then a SSL problem is resulted. Intuitively, the SSL algorithms can be adopted in the domain adaptation problems. The subtle difference between SSL and DAL is that (1) the amount of labeled data in SSL is assumed small but relatively large in DAL, and (2) both the training data and testing data in SSL come from an identical though unknown distribution while those in DAL come from different but related data distributions. Several existing works extended state-of-the-art SSL methods into DAL, such as [7,17].

The past few years also witness the fast advancement of sparse representation (SR) [22,23], which has been successfully applied to many practical problems in signal processing, statistics, and pattern recognition [22,33–35]. Recent researches like [22] have showed that classifiers based on SR are exceptionally effective and achieve by far the best recognition rate on some face databases. Very recently, Fan et al. [35] proposed a robust sparse regularization semi-supervised classification algorithm based on sparse representation (SR) [22,23], and Cheng et al. [21] also proposed a novel label propagation algorithm based on sparsity induced similarity measure strategy.

Enlightened by recent advances in sparse representation by l_1-optimization [22] and great successes in graph-based SSL in many specific applications [12], in the paper, we propose a novel DAL method based on a sparse graph constructed via kernel sparse representation of data in an optimal reproduced kernel Hilbert space (RKHS) recovered by minimizing inter-domain distribution divergence. Our method, named as Sparsity regularized Label Propagation for Domain Adaptation Learning (SLPDAL), can propagate the labels of the labeled data from both domains to the unlabeled one in the target domain using their sparsely reconstructed objects with sufficient smoothness. Besides, we also propose a sparsity preserving regularization framework, in which it is assumed that the label of each data point can be sparsely reconstructed by those of other data points from both domains, and theoretically verify that the SLPDAL algorithm can be derived from this regularization framework. Compared with state-of-the-art methods, the distinct and favorable properties of SLPDAL and the main contributions of this study can be highlighted as follows:

(1) By initially constructing a sparse graph in some RKHS recovered by minimizing the maximum distribution discrepancy between source and target domains, we propose a novel sparse label propagation DAL algorithm based on classical graph-based SSL diagram. The proposed algorithm inhibits and extends the potential advantages of traditional graph-based SSL into DAL, thus smoothly propagating the labels of labeled data in both domains to the unlabeled data in target domain.
(2) SLPDAL can remarkably improve the performance of typical DAL methods by only resorting to very few labeled data and/or unlabeled data, because it is not based on the manifold assumption, but mainly focuses on the discriminative power which can be naturally achieved by minimizing an l_1-regularization objective function. Besides, by using the sparsity concentration index (SCI) [22], SLPDAL can automatically remove the noise points and recover the incomplete data objects from datasets.
(3) Graph construction in SLPDAL relies on a sparse representation criterion [22,23] which is generally superior to the nearest neighbor criterion usually adopted in traditional SSL methods, especially for high-dimensional data. The neighborhood size and edge weight of each sample in the corresponding graph construction are automatically determined in one single step by an l_1-optimization programming procedure. As a result, different samples will get different neighborhood sizes, which is more adaptive to complex data distributions.
(4) The idea behind SLPDAL is quite general, and can be easily applied to graph-based SSL algorithm only through ignoring the domain difference and treating the labeled source domain samples as labeled training data and the unlabeled target domain samples as unlabeled testing data.

The remainder of this paper is organized as follows. In Section 2, we briefly review the related works. In Section 3 we motivate and detail the SLPDAL algorithm for classifying data points in target domain in a RKHS recovered by minimizing the distribution discrepancy between domains. In Section 4, we discuss about the robustness of the proposed method and its extension to out-of-sample learning. Experimental results on several toy and real-world data sets from different domains are reported in Section 5 involving parameter selection and results analysis. Finally, Section 6 concludes the paper.

2. Notations and related works

We refer to the training domain as the source domain where labeled data is abundant, and the testing domain as the target domain where labeled data is not available. For a pattern classification problem, a domain D is given by a distribution $P(\mathbf{x}, y)$, $\mathbf{x} \in X$, $y \in Y$, which is the true underlying distribution of the task, where X and Y denote all possible instances and the corresponding class labels for the task, respectively. In the paper, the transpose of vector or matrix is denoted by the superscript', the trace of a matrix A is represented as $tr(A)$, I denotes identity matrix, and $\mathbf{e}_d$ is a d-dimensional vector of ones. Consider the problem in a reproducing kernel Hilbert space (RKHS) H induced by a nonlinear mapping $\phi : \mathbb{R}^d \to H$ [9]. For a properly chosen ϕ, the inner product $< \cdot, \cdot >$ in H is defined as $\langle \phi(\mathbf{x}_1), \phi(\mathbf{x}_2) \rangle_H = K(\mathbf{x}_1, \mathbf{x}_2)$, where $\mathbf{x}_1, \mathbf{x}_2 \in X$, and $K(,) : X \times X \to \mathbb{R}$ is a positive semi-definite kernel function.

2.1. Distribution discrepancy measure for DAL

Let us denote the data set from the target domain as $X^t = \{\mathbf{x}_j^t\}_{j=1}^m$, where $\mathbf{x}_j^t \in X$. For DAL, testing data set X^t are drawn from a target domain D^t which is different from the source domain D^s of training samples $X^s = \{(\mathbf{x}_i^s, y_i^s)\}_{i=1}^n$, where $\mathbf{x}_i^s \in X$ and $y_i^s \in Y$ is