Engineering Applications of Artificial Intelligence 58 (2017) 157–170

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Joint predictive model and representation learning for visual domain adaptation

Marzieh Gheisari, Mahdieh Soleymani Baghshah*

Computer Engineering Department, Sharif University of Technology, Tehran

ARTICLE INFO

Keywords:
Unsupervised domain adaptation
Predictive function
Shared subspace
Representation learning
Image classification

ABSTRACT

Traditional learning algorithms cannot perform well in scenarios where training data (source domain data) that are used to learn the model have a different distribution with test data (target domain data). The domain adaptation that intends to compensate this problem is an important capability for an intelligent agent. This paper presents a domain adaptation method which learns to adapt the data distribution of the source domain to that of the target domain where no labeled data of the target domain is available (and just unlabeled data are available for the target domain). Our method jointly learns a low dimensional representation space and an adaptive classifier. In fact, we try to find a representation space and an adaptive classifier on this representation space such that the distribution gap between (the marginal and the conditional distribution of) the two domains is minimized and the risk of the adaptive classifier is also minimized. To evaluate the proposed method, we conduct several experiments on image classification datasets. Experimental results verify the superiority of our method to the existing domain adaptation methods and the proposed method outperforms the other methods with a large margin in some of the domain adaptation problems. These results demonstrate the effectiveness of learning the representation space and the adaptive classifier simultaneously.

1. Introduction

Traditional supervised learning methods learn a classification model with labeled training data and use the obtained model to predict labels of unknown test samples. These algorithms usually assume that training (i.e. source) and test (i.e. target) data are sampled from the same distribution. However, when we train a classifier on a source domain and directly use the resulted classifier for a target domain (having a different distribution from the source domain), the accuracy can significantly degrade (Gong et al., 2013). Therefore, we intend to design a classifier which is robust to the mismatch of training and test data distributions. Domain adaptation methods learn to adapt the sample distribution of the source domain to that of the target domain and find classifiers leveraging labelled data of a related source domain to predict labels for data of the target domain. Therefore, the source and the target domains are assumed to be related, but not identical.

Domain adaptation methods can be studied in both semi-supervised and unsupervised tasks. In the semi-supervised domain adaptation, in addition to the labeled data of the source domain, some labeled data from the target domain are also accessible although the labeled data of the target domain are not independently useful to build a proper classifier. Unsupervised domain adaptation methods attend a more challenging problem where no labeled data are available for the target domain. However, in the unsupervised domain adaptation, it is assumed that unlabeled data of the target domain are available during domain adaptation. In this paper, we focus on the unsupervised domain adaptation which is well suited to many real-world applications since we do not access labeled data of the target domain in many real-world cases. Unsupervised domain adaptation methods usually assume that the distribution mismatch can be reduced by reweighting samples of the source domain (in order to match the target distribution), or by projecting data of both the source and the target domains into a common subspace in which data distributions of these domains are somewhat indistinguishable.

In the recent years, domain adaptation has received a lot of attention in many applications. Until now, domain adaptation methods have been used in Natural Language Processing (NLP) and machine vision applications such as face detection (Shekhar et al., 2013; Cao et al., 2013), document classification (Shi et al., 2010; Dong et al., 2016), sentiment analysis (Glorot et al., 2011; Liu et al., 2015; Xia et al., 2013), handwritten digit recognition (Hosseinzadeh and Razzazi, 2016), and event recognition (Duan et al., 2010; Yang et al., 2007; Zhu and Shao, 2014).

Domain adaptation in many image classification applications

* Corresponding author.
E-mail addresses: gheisari@ce.sharif.edu (M. Gheisari), soleymani@sharif.edu (M. Soleymani Baghshah).

http://dx.doi.org/10.1016/j.engappai.2016.12.004
Received 18 February 2016; Received in revised form 19 August 2016; Accepted 5 December 2016

(called visual domain adaptation) is useful since domain shift (the difference between the distribution of training and test data) occurs frequently in these tasks and collecting many labeled data from all domains is not possible (Saenko et al., 2010). Some real world conditions that may alter the image statistics include changes in sensor type, scene lighting, object pose, background, and in the extreme case all of them. For example, training data might be images captured from objects at frontal pose under very restricted illumination while test data contains images captured from various viewpoints under poor lighting conditions. As another example, human faces on color images may be available for training a face detector while we need to detect human faces on infrared images (Patel et al., 2015). The availability of various devices and large Internet databases has led to create a big demand for visual domain adaptation recently.

In this paper, we propose a general framework that we called Joint Adaptive Classifier and Representation Learning (JACRL). Our method learns both an adaptive classifier and a low dimensional representation space jointly by minimizing the structural risk functional and the embedded distribution gap between (marginal and conditional distribution of) the two domains and also maximizing the manifold consistency of the adaptive classifier. In this way, we use unlabeled data of the target domain both to learn a new representation space in which the distribution of the target domain is accommodated to that of the source domain and also to learn an adaptive classifier (in the new representation space) via a semi-supervised method that incorporates unlabeled data of the target domain. Labeled data of the source domain are also utilized to train the classifier as well as to find the conditional distributions of the source domain which are required when we intend to reduce the difference between the two domains in the shared subspace. In the proposed JACRL framework, we try to learn a new representation space and an adaptive classifier simultaneously. Indeed, we learn a new representation space in which a more proper adaptive classifier can be found. Many supervised methods, such as Regularized Least Squares (RLS) and Support Vector Machines (SVM), can be entered to the proposed framework for domain adaptation tasks. We also present the kernelized version of our method to make it more general, which usually lead to better performance.

The rest of the paper is organized as follows: we first review some related works in Section 2. In Section 3, the general requirements as some definitions are presented. Our unsupervised domain adaptation method is introduced in Section 4. Experimental settings and results on image classification tasks are presented in Section 5. Finally, we conclude our work in Section 6.

2. Related work

The existing domain adaptation methods usually can be categorized into instance-based, feature-based, and model-based methods:

Instance-based methods (Jiang, 2008; Sun et al., 2011; Jiang and Zhai, 2007; Aljundi et al., 2015) try to make the source distribution close to the target distribution by reweighting or selecting training samples in the original feature space. Kernel Mean Matching (KMM) (Huang et al., 2006) learns the resampling weights by matching distributions of training and test sets in an RKHS. Some instance weighting methods such as (Jiang and Zhai, 2007) minimize the expected loss over the target distribution by incorporating instance-dependent weights into the loss function in order to find a proper model for the target domain. More recently, a method was proposed in (Gong et al., 2013) based on the notion of landmarks where the landmarks are a subset of labeled instances in the source domain that are most similarly distributed to data of the target domain.

Feature-based methods are based on transforming the original feature space into another space in which the source and the target domain have common characteristics and they are approximately indistinguishable. In the new feature space, standard classifiers can be trained on the source domain and the learnt classifier can also be applied to the target domain. Blitzer et al., (2006) proposed a Structural Correspondence Learning (SCL) method which is a domain adaptation method for NLP tasks such as sentiment classification. SCL selects some domain independent pivot features that behave similarly in both the domains and leverages these features to learn an embedded space by identifying correspondences between the features. Pan et al., (2008) proposed the Maximum Mean Discrepancy Embedding (MMDE) method to learn a low-dimensional linear projection which tends to move the two distributions closer to each other. These authors also proposed the Transfer Component Analysis (TCA) method (Pan et al., 2011) that is a more efficient algorithm overcoming the high computational cost of MMDE. Recently, (Long et al., 2013) learns a new representation by matching both the marginal and conditional distributions. Saenko et al., (2010) introduced the domain adaptation problem for visual recognition tasks in a semi-supervised setting and proposed a method based on metric learning with cross-domain constraints. This method learns a symmetric transformation to project data of the source and the target domains to a domain invariant space. Gopalan et al., (2011) extended this method to the unsupervised setting inspired by an incremental learning approach. In fact, a finite number of subspaces is sampled along the geodesic path connecting the source and the target subspace on the Grassmann manifold and some intermediate meaningful subspaces between the source and the target domain are found. Then, the source and the target data are projected onto these subspaces. Geodesic Kernel Flow (GFK) (Gong et al., 2012) extended this idea to consider all the subspaces along the geodesic path, instead of a finite number of sampled subspaces, and the source and the target data are projected onto an infinite number of subspaces (new representation is obtained using all of these subspaces).

Model-based methods try to directly design an adaptive classifier with a good performance on the target domain. They adapt parameters of the source classifier to new parameters that are proper for the target domain. Most of the existing model-based methods have been proposed for semi-supervised setting (where some labeled data from the target domain are available). Some studies (Yang et al., 2007; Zisserman and Aytar, 2011; Jiang et al., 2008) try to find an optimal model for the target domain by changing the SVM formulation. Yang et al., (2007) proposed the Adaptive Support Vector Machine (ASVM) that learns a new decision boundary for the target data by leveraging the classifier trained on the source data. (Marconcini and Bruzzone, 2010) proposed the Domain Adaptation Support Vector Machine (DASVM) that iteratively learns a classifier for the target data (where the training set is updated iteratively). In each iteration, DASVM labels the unlabeled training samples of the target domain and simultaneously removes those labeled data of the source domain that seem to be useless (for training a classifier on the target domain). Duan et al., (2012) proposed the Domain Adaptation Machine (DAM) that uses pre-learnt classifiers with the labeled samples of source domains and also considers a regularization term (based on the smoothness assumption) which enforces the target classifier to have similar decisions to the pre-learnt classifiers. A small number of works also aims to directly learn an adaptive classifier by adding a regularization term (showing a distance measure) to the optimization problem of the supervised learning methods (such as SVMs) (Yang et al., 2007; Tao et al., 2012; Long et al., 2014; Quanz and Huan, 2009).

Both the instance-based and the feature-based approaches intend to reduce the distribution gap between the training and test data. However, generally, they divide the domain adaptation task into two separate steps. All these approaches attempt to reduce the distribution gap between the domains firstly and then train a classifier based on reweighted or transformed instances. In these methods, reweighting or transformation makes the two domains look more similar, but it may lose some discriminative information. On the other hand, model-based methods adapt model parameters of the source domain to the target domain without any changes to the representation of data.

None of the above approaches consider representation learning and