Engineering Applications of Artificial Intelligence 73 (2018) 80–91

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Sample-to-sample correspondence for unsupervised domain adaptation☆

Debasmit Das *, C.S. George Lee

School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA

ARTICLE INFO

Keywords:
Unsupervised domain adaptation
Correspondence
Convex optimization
Image classification
Sentiment classification

ABSTRACT

The assumption that training and testing samples are generated from the same distribution does not always hold for real-world machine-learning applications. The procedure of tackling this discrepancy between the training (source) and testing (target) domains is known as domain adaptation. We propose an unsupervised version of domain adaptation that considers the presence of only unlabelled data in the target domain. Our approach centres on finding correspondences between samples of each domain. The correspondences are obtained by treating the source and target samples as graphs and using a convex criterion to match them. The criteria used are first-order and second-order similarities between the graphs as well as a class-based regularization. We have also developed a computationally efficient routine for the convex optimization, thus allowing the proposed method to be used widely. To verify the effectiveness of the proposed method, computer simulations were conducted on synthetic, image classification and sentiment classification datasets. Results validated that the proposed local sample-to-sample matching method out-performs traditional moment-matching methods and is competitive with respect to current local domain-adaptation methods.

1. Introduction

In traditional machine-learning problems, we assume that the test data is drawn from the same distribution as the training data. However, such an assumption is rarely encountered in real-world situations. For example, consider a recognition system that distinguishes between a cat and a dog, given labelled training samples of the type shown in Fig. 1(a). These training samples are frontal faces of cats and dogs. When the same recognition system is used to test in a different domain such as on the side images of cats and dogs as shown in Fig. 1(b), it would fail miserably. This is because the recognition system has developed a bias in being able to only distinguish between the face of a dog and a cat and not side images of dogs and cats. Domain adaptation (DA) aims to mitigate this dataset bias (Torralba and Efros, 2011), where different datasets have their own unique properties. Dataset bias appears because of the distribution shift of data from one dataset (i.e., source domain) to another dataset (i.e., target domain). The distribution shift manifests itself in different forms. In computer vision, it can occur when there is changing lighting conditions, changing poses, etc. In speech processing, it can be due to changing accent, tone and gender of the person speaking. In remote sensing, it can be due to changing atmospheric conditions, change in acquisition devices, etc. To encounter this discrepancy in distributions, domain adaptation methods have been proposed. Once domain adaptation is carried out, a model trained using the adapted source domain data should perform well in the target domain. The underlying assumption in domain adaptation is that the task is the same in both domains. For classification problems, it implies that we have the same set of categories in both source and target domains.

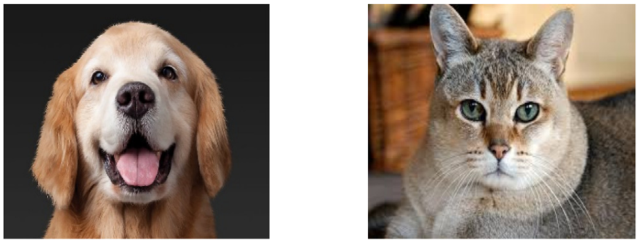
Domain adaptation can also assist in annotating datasets efficiently and further accelerating machine-learning research. Current machine-learning models are data hungry and require lots of labelled samples. Though huge amount of unlabelled data is obtained, labelling them requires lot of human involvement and effort. Domain adaptation seeks to automatically annotate unlabelled data in the target domain by adapting the labelled data in the source domain to be close to the unlabelled target-domain data.

In our work, we consider *unsupervised domain adaptation* (UDA), which assumes absence of labels in the target domain. This is more realistic than semi-supervised domain adaptation, where there are also a few-labelled data in the target domain. This is because labelling data might be time-consuming and expensive for real-world situations. Hence we need to effectively exploit fully labelled source-domain data and fully unlabelled target-domain data to carry out domain adaptation. In our case, we seek to find correspondences between each source-domain

☆ This work was supported in part by the National Science Foundation under Grant IIS-0916807. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

* Corresponding author.
E-mail addresses: das35@purdue.edu (D. Das), csglee@purdue.edu (C.S.G. Lee).

https://doi.org/10.1016/j.engappai.2018.05.001
Received 6 December 2017; Received in revised form 27 March 2018; Accepted 1 May 2018
0952-1976/



(a) Source domain.

(b) Target domain.

Fig. 1. Discrepancy between the source domain and the target domain. In the source domain, the images have frontal faces while the target domain has images of the whole body from the side view-point.

sample and each target-domain sample. Once the correspondences are found, we can transform the source-domain samples to be close to the target-domain samples. The transformed source-domain samples will then lie close to the data space of the target domain. This will allow a model trained on the transformed source-domain data to perform well with the target-domain data. This not only achieves the goal of training robust models but also allows the model to annotate unlabelled target-domain data accurately.

The remainder of the paper is organized as follows: Section 2 discusses related work of domain adaptation. Section 3 discusses the background required for our proposed approach. Section 4 discusses our proposed approach and formulates our unsupervised domain adaptation problem into a constrained convex optimization problem. Section 5 discusses the experimental results and some comparison with existing work. Section 6 discusses some limitations. Section 7 concludes with a summary of our work and future research directions. Finally, the Appendix shows more details about the proof of convexity of the optimization objective function and derivation of the gradients.

2. Related work

There is a large body of prior work on domain adaptation. For our case, we only consider homogeneous domain adaptation, where both the source and target domains have the same feature space. Most of previous DA methods are classified into two categories, depending on whether a deep representation is learned or not. In that regard, our proposed approach is not deep-learning-based since we directly work at the feature level without learning a representation. We feel that our method can easily be extended to deep architectures and provide much better results. For a comprehensive overview on domain adaptation, please refer to Csurka's survey paper (Csurka, 2017).

2.1. Non-deep-learning domain-adaptation methods

These non-deep-learning domain-adaptation methods can be broadly classified into three categories — instance re-weighting methods, parameter adaptation methods, and feature transfer methods. Parameter adaptation methods (Jiang et al., 2008; Bruzzone and Marconcini, 2010; Duan et al., 2009; Yang et al., 2007) generally adapt a trained classifier in the source domain (e.g., an SVM) in order to perform better in the target domain. Since these methods require at least a small set of labelled target examples, they cannot be applied to UDA.

Instance Re-weighting was one of the early methods, where it was assumed that conditional distributions were shared between the two domains. The instance re-weighting involved estimating the ratio between the likelihoods of being a source example or a target example to compute the weight of an instance. This was done by estimating the likelihoods independently (Zadrozny, 2004) or by approximating the ratio between the densities (Kanamori et al., 2009; Sugiyama et al., 2008). One of the most popular measures used to weigh data instances, used in Gretton et al. (2009) and Huang et al. (2007), was the Maximum Mean Discrepancy (MMD) (Borgwardt et al., 2006) computed between the data distributions in the two domains. *Feature Transfer* methods, on the other hand, do not assume the same conditional distributions between the source and target domains. One of the simplest methods for DA was proposed in Daumé III (2009), where the original representation is augmented with itself and a vector of the same size is filled with zeros — the source features become $(\mathbf{x}_s, \mathbf{x}_s, 0)$ and the target features become $(\mathbf{x}_t, 0, \mathbf{x}_t)$. Then an SVM is trained on these augmented features to figure out which parts of the representation is shared between the domains and which are the domain-specific ones. The idea of feature augmentation inspires the Geodesic Flow Sampling (GFS) (Gopalan et al., 2014, 2011) and the Geodesic Flow Kernel (GFK) (Gong et al., 2012, 2013), where the domains are embedded in d-dimensional linear subspaces that can be seen as points on the Grassmann manifold, corresponding to the collection of all d-dimensional subspaces. The Subspace Alignment (SA) (Fernando et al., 2013) learns an alignment between the source subspace obtained by Principal Component Analysis (PCA) and the target PCA subspace, where the PCA dimensions are selected by minimizing the Bregman divergence between the subspaces. Similarly, the linear Correlation Alignment (CORAL) (Sun et al., 2016) algorithm minimizes the domain shift using the covariance of the source and target distributions. Transfer Component Analysis (TCA) (Pan et al., 2011) discovers common latent features having the same marginal distribution across the source and target domains. Feature transformation proposed by Chen et al. (2012) exploits the correlation between the source and target sets to learn a robust representation by reconstructing the original features from their noisy counterparts. All these previous methods learned a global transformation between the source and target domains. In contrast, the Adaptive Transductive Transfer Machines (ATTM) (Farajidavar et al., 2014) learned both a global and a local transformation from the source domain to the target domain that is locally linear. Similarly, the optimal transport for domain adaptation (Courty et al., 2017) considers a local transport plan for each source example.

2.2. Deep domain-adaptation methods

Most deep-learning methods for DA follow a twin architecture with two streams, representing the source and target models. They are then trained with a combination of a classification loss and a discrepancy loss (Long et al., 2016, 2015; Tzeng et al., 2014; Ghifary et al., 2015; Sun and Saenko, 2016) or an adversarial loss. The classification loss depends on the labelled source data, and the discrepancy loss diminishes the shift between the two domains. On the other hand, adversarial-based methods encourage domain confusion through an adversarial objective with respect to a domain discriminator. The adversarial loss tries to encourage a common feature space through an adversarial objective with respect to a domain discriminator. Tzeng et al. (2017) proposes a unified view of existing adversarial DA methods by comparing them according to the loss type, the weight-sharing strategy between the two streams, and on whether they are discriminative or generative. The Domain-Adversarial Neural Networks (DANN) (Ganin et al., 2016) integrates a gradient reversal layer into the standard architecture to promote the emergence of features that are discriminative for the main learning task in the source domain and indiscriminate with respect to