Neurocomputing 000 (2017) 1–8

Contents lists available at ScienceDirect

Neurocomputing

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

Coupled local–global adaptation for multi-source transfer learning

Jieyan Liu[a], Jingjing Li[a,b,*], Ke Lu[a,**]

[a] University of Electronic Science and Technology of China, Chengdu, China
[b] University of Queensland, Brisbane, Australia

ARTICLE INFO

Article history:
Received 23 November 2016
Revised 9 May 2017
Accepted 24 June 2017
Available online xxx

Communicated by Deng Cai

Keywords:
Domain adaptation
Subspace learning
Transfer learning

ABSTRACT

This paper presents a novel unsupervised multi-source domain adaptation approach, named as coupled local–global adaptation (CLGA). At the global level, in order to maximize the adaptation ability, CLGA regards multiple domains as a unity, and jointly mitigates the gaps of both marginal and conditional distributions between source and target dataset. At the local level, with the intention of maximizing the discriminative ability, CLGA investigates the relationship among distinctive domains, and exploits both class and domain manifold structures embedded in data samples. We formulate both local and global adaptation in a concise optimization problem, and further derive an analytic solution for the objective function. Extensive evaluations verify that CLGA performs better than several existing methods not only in multi-source adaptation tasks but also in single source scenarios.

1. Introduction

One of the common anxieties among researchers and engineers at the age of big data is that quality labeled samples are quite scarce and, consequently, expensive. Therefore, both brilliant and indolent scientists are eagerly exploring ways to automatically learn labels for novel domains from some related ones where data have been labeled for previous jobs. However, the data distributions are often mismatched between different domains, which cause significant performance degradation when the classifiers trained on source domain are applied to the target domain [1,2].

Recently, the community has witnessed an increasing interest in developing transfer learning [3] algorithms for handling cross-domain problems. As a practical branch of transfer learning, domain adaptation [4–9] has been regarded as a promising technology to learn cross-domain transforms. Yet, existing domain adaptation approaches either assume that source data are from single domain [10–12], or ignore the multiple manifold structures embedded in different domains [13–17]. However, in real-world applications, people often collect and organize data according to semantic meanings, e.g., person name for face images and action name for motion videos. Although this ad hoc practice is reasonable and easy to be implemented, there is one pitfall that a dataset is often a mixture of multiple distinctive domains. For example, if we collect images of Albert Einstein on the web, we can find photographs with different poses/illuminations/ages, as well as caricatures and even photo sketches. As a result, roughly treating those datasets as a single domain would eliminate the distinctions and, consequently, undermine the distinctive ability of the learned model. Furthermore, although some previous work [14,15] have exploited how to discover latent domains, how to effectively utilize both the class information and domain information in a unified framework to boost the adaptation remains to be a vital concern for multi-source transfer learning.

The discussed two concerns in multi-source domain adaptation, discovering latent domains and adapting domain knowledges, are related within a complete system, but they can be addressed independently [2]. Literatures specialized in the first problem can be found in [14,15]. Limited by space, this paper focuses on the second issue. Specifically, we propose a novel approach, named as coupled local–global adaptation (CLGA), by jointly maximizing the global adaptation ability and the local discriminative ability.

This work is motivated by two key insights of multi-source domain adaptation, as illustrated in Fig. 1. The first insight is *to maximize global adaptation ability between source and target dataset*. No matter how many latent domains are mixed in a dataset, they all have one very single mission: serving as source (or target) dataset at a time. Therefore, from a global perspective, source domains are still in a unity, so are target domains. To avoid compromising this kind of data integrity, CLGA globally mitigates the gaps of both marginal and conditional distributions between source dataset and target dataset. Another insight is *to maximize local discriminative*

* Corresponding author at: University of Electronic Science and Technology of China, Chengdu, China.
** Corresponding author.
E-mail addresses: lijin117@yeah.net (J. Li), kel@uestc.edu.cn (K. Lu).

http://dx.doi.org/10.1016/j.neucom.2017.06.051
0925-2312/

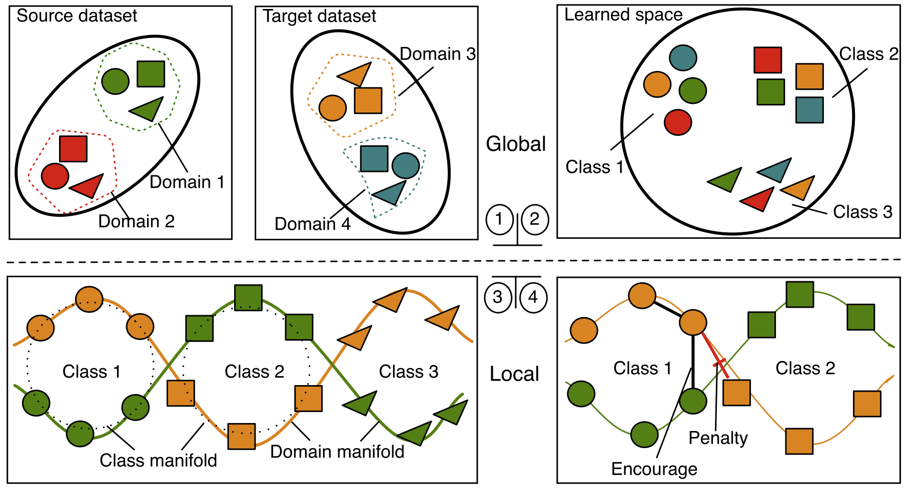

Fig. 1. Illustration of our approach. ① and ② show our idea from the view of global adaptation. Although there are multiple domains in each dataset, different domains have identical mission: serving as source dataset or target dataset. Therefore, we treat them as a unity to globally mitigate gaps of both marginal and conditional distributions. ③ and ④ show our idea from the view of local adaptation. Since multi-domain data are embedded in both class and domain manifolds, we encourage samples in the same class but from different domains to stay close, and also penalize samples in different classes but from the same domain distant from each other.

ability among distinctive domains. Since multi-domain data are embedded in both class and domain manifolds, the distance between two samples in different classes but from the same domain can be shorter than the distance between two samples in the same class but from different domains, e.g., the Euclidean distance between two sketches of different faces can be much shorter than the distance between one sketch and its matching photo. To address this, CLGA pays close attention to the local relationship among multiple distinctive domains and classes. Specifically, during the cross-domain transformation, CLGA encourages samples in the same class but from different domains to stay close, while penalizes samples in different classes but from the same domain distant from each other.

Technically, we extend the nonparametric Maximum Mean Discrepancy (MMD) [6,18] to globally adapt both the marginal and conditional distributions. We also propose a novel graph structure under the framework of graph embedding [19] to locally adapt the multi-manifold structure. We formulate both local and global adaptation into a concise optimization problem, and optimize local–global adaptation via a principled dimensionality reduction procedure. The main contributions of this paper are summarized as follows:

(1) A unified objective of coupled local–global adaptation is proposed for unsupervised multi-source transfer learning. It achieves good performance on several standard benchmarks with significant improvements compared with baselines. To the best of our knowledge, this work is the first one that considers both the adaptation ability from a global transfer view and the discriminative ability from a local multi-manifold perspective.
(2) CLGA is formulated with good generalization ability. It is compliant to semi-supervised settings and single domain adaptation tasks, and it outperforms several existing methods in single source scenarios as well.

In the rest of this paper, we first review some related work reported in recent literatures and highlight the differences of our work from them. Then, we detail our work from formulation to optimization, followed by the experiments in both multi-source and single source domain adaptation tasks are presented to verify the effectiveness of our method. At last, we conclude our work and exploit possible improvements of our model, which can be done in the future work.

2. Related work and discussions

According to the mechanism of different methods, existing domain adaptation algorithms can be roughly categorized into two groups: feature selection and instance re-weighting. Specifically, feature selection methods [4,7,11,12] aim to learn a latent space where common features shared by source domain and target domain can be uncovered. While instance re-weighting approaches [5,10,20] usually try to train a weight sensitive classifier on the source domain, e.g., multiple-kernel SVM, which can be used in the target domain, or use landmarks [21] to re-weight samples. Some previous work [11] also try to exploit benefits from both sides by jointly optimizing feature selection and instance re-weighting. However, most of previous algorithms, in both groups, assume that there is only one domain in either source dataset or target dataset. Hoffman et al. [14] and Gong et al. [15] notice that latent domain information is beneficial to adaptation performance in multi-source tasks. Xiong et al. [22] verify that exploiting manifold structure of latent domains can further modeling the dataset. Li et al. [23–25] investigate the manifold structures under subspace learning. For multi-source domain adaptation, another possible solution is to learn multiple cross-domain transforms, one for each source-target pair [26]. However, it requires prior target class information for training and domain information to choose which transform should be used in the testing stage.

In general, almost all of the previous approaches only focus on one property of the data: global adaptation ability. It is ignored that multi-domain data are embedding on both class and domain manifolds. To the best of our knowledge, this work is the first one that considers both the adaptation ability from a global transfer view and the discriminative ability from a local multi-manifold perspective. Furthermore, we formulate the coupled local–global adaptation problem into a concise objective, and present an analytical solution. It is worth noting that although previous work [27] propose a graph-regularized domain adaptation model, it does not take the multi-domain discrimination into consideration.

3. Coupled local–global adaptation

3.1. Notations

In this paper, we use bold lowercase letters to represent vectors, bold uppercase letters to represent matrices, specifically,