Towards Uncovering the Intrinsic Data Structures for Unsupervised Domain Adaptation Using Structurally Regularized Deep Clustering

Hui Tang, Xiatian Zhu, Ke Chen, Kui Jia, and C. L. Philip Chen, *Fellow, IEEE*

Abstract—Unsupervised domain adaptation (UDA) is to learn classification models that make predictions for unlabeled data on a target domain, given labeled data on a source domain whose distribution diverges from the target one. Mainstream UDA methods strive to learn domain-aligned features such that classifiers trained on the source features can be readily applied to the target ones. Although impressive results have been achieved, these methods have a potential risk of damaging the *intrinsic* data structures of target discrimination, raising an issue of generalization particularly for UDA tasks in an inductive setting. To address this issue, we are motivated by a UDA assumption of *structural similarity* across domains, and propose to directly uncover the intrinsic target discrimination via constrained clustering, where we constrain the clustering solutions using structural source regularization that hinges on the very same assumption. Technically, we propose a hybrid model of *Structurally Regularized Deep Clustering*, which integrates the regularized discriminative clustering of target data with a generative one, and we thus term our method as H-SRDC. Our hybrid model is based on a deep clustering framework that minimizes the Kullback-Leibler divergence between the distribution of network prediction and an auxiliary one, where we impose structural regularization by learning domain-shared classifier and cluster centroids. By enriching the structural similarity assumption, we are able to extend H-SRDC for a pixel-level UDA task of semantic segmentation. We conduct extensive experiments on seven UDA benchmarks of image classification and semantic segmentation. With no explicit feature alignment, our proposed H-SRDC outperforms all the existing methods under both the inductive and transductive settings. We make our implementation codes publicly available at https://github.com/huitangtang/H-SRDC.

Index Terms—Domain adaptation, deep clustering, inductive learning, image classification, semantic segmentation

1 Introduction

In many practical applications of machine learning, the problem of interest is concerned with learning from data on a domain where, due to practical constraints and/or expenses, data annotations are difficult to acquire, and a standard supervised training cannot be readily applied; instead, labeled data on a *different but related* domain can be obtained relatively easily. This creates a learning scenario in which one is tempted to leverage the labeled data on the *source* domain to help train machine learning models for a transferrable use on the unlabeled *target* domain, i.e., the problem of *unsupervised domain adaptation (UDA)* [1], [2]. UDA typically assumes a shared label space between the source and target domains, and its technical challenge arises from the assumed existence of distribution divergence between the two domains.

A rich literature of UDA research has been developed in the past decades [3]. Among them, mainstream methods [4], [5], [6], [7], [8], [9], [10] are motivated by the seminal theories [2], [11], [12] that bound the expected errors of classification models on the target domain by quantities involving classifier-induced divergence between feature distributions of the two domains, e.g., those recent ones based on adversarial training of deep networks [5], [6], [7]. Consequently, these methods strive to minimize domain divergence by learning aligned features between the two domains, such that classifiers trained on the features of source domain can be readily applied to the target ones. Despite the impressive results achieved, these methods have a potential risk of damaging the *intrinsic* data structures of target discrimination, as analyzed recently in [13]. We note that more importantly, this shortcoming would become severer in the practically more useful setting of *inductive* UDA (analogous to the setting of inductive transfer learning in [1]), where the objective is to learn classification models as off-the-shelf ones such that they can be used for held-out data sampled from the same target domain; adapting classifiers to the damaged discrimination of target data by feature alignment of existing methods would be less effective for inductive UDA, since the held-out target data still follow the undamaged, intrinsic data discrimination.

To overcome such limitations in existing methods, we first revisit the general UDA assumptions made in existing

- *Hui Tang, Ke Chen, and Kui Jia are with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, China. E-mail: eehuitang@mail.scut.edu.cn, {chenk, kuijia}@scut.edu.cn.*
- *Xiatian Zhu is with the Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, GU2 7XH Guildford, U.K. E-mail: eddy.zhuxt@gmail.com.*
- *C. L. Philip Chen is with the School of Computer Science and Engineering, South China University of Technology, Guangzhou 510641, China. E-mail: philip.chen@ieee.org.*

Manuscript received 24 Nov. 2020; revised 6 Apr. 2021; accepted 5 June 2021. Date of publication 9 June 2021; date of current version 9 Sept. 2022.
(Corresponding author: Kui Jia.)
Recommended for acceptance by D. Damen.
Digital Object Identifier no. 10.1109/TPAMI.2021.3087830