Deep Subdomain Adaptation Network for Image Classification

Seri Lee
Computer Science and Engineering
Seoul National University
Seoul, Republic of Korea
sally20921@snu.ac.kr
ß

***Abstract*—For a target task where the labeled data are unavailable, domain adaptation can transfer a learner from a different source domain. Previous deep domain adaptation methods mainly learn a global domain shift, i.e., align the global source and target distributions without considering the relationships between two subdomains within the same category of different domains, leading to unsatisfying transfer learning performance without capturing the fine-grained information. Recently, more and more researchers pay attention to subdomain adaptation that focuses on accurately aligning the distributions of relevant subdomains. However, most of them are adversarial methods that contain several loss functions and converge slowly. Based on this, we present a deep subdomain adaptation network (DSAN) that learns a transfer network by aligning the relevant subdomain distributions of domain-specific layer activations across different domains based on a local maximum mean discrepancy (LMMD). Our DSAN is very simple but effective, which does not need adversarial training and converges fast. The adaptation can be achieved easily with most feedforward network models by extending them with LMMD loss, which can be trained efficiently via backpropagation.**

***Index Terms*—Domain adaptation, fine-grained, subdomain**

I. Introduction

In recent years, deep learning methods have achieved impressive success in computer vision [1], which, however usually needs large amounts of labeled data to train a good deep network. In the real world, it is often expensive and laborsome to collect enough labeled data. For a target task with the shortage of labeled data, there is a strong motivation to build effective learners that can leverage rich labeled data from related source domain. However, this learning paradigm suffers from the shift of data distributions across different domains, which will undermine the generalization ability of machine learning models.

Learning a discriminative model in the presence of the shift between the training and test data distributions is known as domain adaptation or transfer learning. Previous shallow domain adaptation methods bridge the source and target domains by learning invariant feature representations or estimate instance importance without using target labels. Recent studies have shown that deep networks can learn more transferable features for domain adaptation, by disentangling explanatory factors of variations behind domains. The latest advantages have been achieved by embedding domain adaptation modules in the pipeline of deep feature learning to extract domain-invariant representations.

The previous deep domain adaptation methods, mainly learn a global domain shift, i.e., aligning the global source and target distributions without considering the relationships between two subdomains in both domains (a subdomain contains the samples within the same class). As a result, not only all the data from the source and target domains will be confused, but also the discriminative structures can be mixed up. This might lose the fine-grained information for each category.

After global domain adaptation, the distributions of the two domains are approximately the same, but the data in different subdomains are too close to be classified accurately. This is a common problem in previous global domain adaptation methods. Hence, matching the global source and target domains may not work well for diverse scenarios.

With regard to the challenge of global domain shift, recently, more and more researchers pay attention to subdomain adaptation (also called semantic alignment or matching conditional distribution) which is centered on learning a local domain shift, i.e., accurately aligning the distribution of the relevant subdomains within the same category in the source and target domains. After subdomain adaptation, with the local distribution that is approximately the same, the global distribution is also approximately the same. However, all of them are adversarial methods that contain several loss functions and converge slowly.

Based on the subdomain adaptation, we propose a deep subdomain adaptation network (DSAN) to align the relevant subdomain distributions of activations in multiple domain-specific layers across domains for unsupervised domain adaptation. DSAN extends the feature representation ability of deep adaptation networks (DANs) by aligning relevant subdomain distributions as mentioned earlier. A key improvement over previous domain adaptation methods is the capability of subdomain adaptation to capture the fine-grained information for each category, which can be trained in an end-to-end framework. To enable proper alignment, we design a local maximum mean discrepancy (LMMD), which measures teh Hilbert-Schmidt norm between kernel mean embedding of empirical distributions of the relevant subdomains in source and target domains with considering the weight of different

samples.

The LMMD method can be achieved with most feedforward network models and can be trained efficiently using standard backpropagation. In addition, our DSAN is very simple and easy to implement. Note that the most remarkable results are achieved by adversarial methods recently. Experiments show that DSAN, which is a nonadversarial method, can obtain remarkable results for standard domain adaptation on both object recognition tasks and digit classification tasks.

II. Related Work

A. Domain Adaptation

Recent years have witnessed many approaches to solve the visual domain adaptation problem, which is also commonly framed as the visual data set bias problem. Previous shallow methods for domain adaptation include reweighting the training data so that they can more closely reflect those in the test distribution, and finding a transformation in a lower dimensional manifold that draws the source and target subspaces closer.

Recent studies have shown that deep networks can learn more transferable features for domain adaptation, by disentangling explanatory factors of variations behind domains. The latest advances have been achieved by embedding domain adaptation modules in the pipeline of deep feature learning to extract domain-invariant representations. Two main approaches are identified among the literature. The first is statistic moment matching-based approach, i.e., MMD, central moment discrepancy (CMD), and second-order statistics matching. The seoncd commonly used approach is based on an adversarial loss, which encourage samples from different domains to be nondiscriminative with respect to domain labels, i.e., domain adversarial net-based adaptation methods borrowing the idea of GAN. Generally, the adversarial approaches can achieve better performance than the statistic moment matching-based approaches. In addition, most state-of-the-art approaches are domain adversarial net-based adaptation methods. Our DSAN is an MMD-based method. We show that DSAN without adversarial loss can achieve remarkable results.

B. Maximum Mean Discrepancy

MMD has been adopted in many approaches, for domain adaptation. In addition, there are some extensions of MMD. Conditional MMD and joint MMD measure the Hilbert-Schmidt norm between kernel mean embedding of empirical conditional and joint distributions of the source and target data, respectively. Weighted MMD alleviates the class weight bias by assigning class-specific weights to source data. However, our local MMD measures the discrepancy between kernel mean embedding relevant subdomains in source and target domains with considering the weight of different samples.

C. Subdomain Adaptation

Recently, we have witnessed considerable interest and research for subdomain adaptation that focuses on accurately aligning the distributions of the relevant subdomains. Multiadversarial domain adaptation (MADA) captures the multimode structures to enable fine-grained alignment of different data distributions based on multiple-domain discriminator. Moving semantic transfer network (MSTN) learns the semantic representations for unlabeled target samples by aligning labeled source centroid and pseudolabeled target centroid. CDAN conditions the adversarial adaptation models on discriminative information conveyed in the classifier predictions. Co-DA constructs multiple diverse feature spaces and aligns source and target distributions in each of them individually while encouraging that alignments agree with each other with regard to the class predictions on the unlabeled target examples. The adversarial loss is adopted by all of them.

III. Deep Subdomain Adaptation Network

To divide the source and target domains into multiple subdomains that contain the samples within the same class, the relationships between the samples should be exploited. It is well known that the samples within the same category are more relevant. However, data in the target domain is unlabeled. Hence, we would use the output of the networks as the pseudolabels of target domain data, which will be detailed later.

References

[1] Sicheng Zhao, Xiangyu Yue, Shanghang Zhang, Bo Li, Han Zhao, Bichen Wu, Ravi Krishna, Joseph E Gonzalez, Alberto L Sangiovanni-Vincentelli, and Sanjit A Seshia. A review of single-source deep unsupervised visual domain adaptation. *IEEE Transactions on Neural Networks and Learning Systems 2162-237X*, 2020.