Discrepancy-Aware Collaborative Representation for Unsupervised Domain Adaptation

Chao Han
School of Electronics and Information
Northwestern Polytechnical University
Xi'An, China
hanc@mail.nwpu.edu.cn

Deyun Zhou
School of Electronics and Information
Northwestern Polytechnical University
Xi'An, China
dyzhou@nwpu.edu.cn

Yu Xie
School of Electronic Engineering
Xidian University
Xi'An, China
sxlljcxy@gmail.com

Yu Lei
School of Electronics and Information
Northwestern Polytechnical University
Xi'An, China
leiy@nwpu.edu.cn

Jiao Shi
School of Electronics and Information
Northwestern Polytechnical University
Xi'An, China
jiaoshi@nwpu.edu.cn

Maoguo Gong
School of Electronic Engineering
Xidian University
Xi'An, China
gong@ieee.org

***Abstract*—Domain adaptation aims at learning from the labeled source domain to build an accurate classifier for a related but different target domain. Existing methods attempt to reduce domain discrepancy explicitly by means of statistical properties yet ignore the inherent differences among samples. In this paper, we present a novel solution for domain adaptation based on collaborative representation, named Discrepancy-Aware Collaborative Representation (DACR). Inspired by the success of nearest regularization, DACR develops a novel indicator to measure the discrepancy among every source sample and target domain. Then the indicator is employed in sparse regularization thus ensure that samples with small discrepancy have larger weights in the learned representation. Extensive experiments verify that DACR is able to achieve comparable performance with existing methods while significantly reducing computing complexity.**

***Index Terms*—Transfer Learning, Domain Adaptation, Collaborative Representation**

I. Introduction

Traditional machine learning usually performs poorly when the training and testing data have different distributions, moreover, to label the data from each source is time-consuming. To solve this challenging scenario, domain adaptation (DA) has attracted the interests of many researchers. Specifically, domain adaptation learns from a well-labeled source domain to make predictions in an unlabeled target domain while the two domains are different but related, which has been successfully applied in many fields [1]–[4].

Existing methods try to reduce the domain discrepancy by mapping original data to feature space. As a theoretical basis, Ben *et al.* pointed out the ideal representation should have less source discriminant error and domain discrepancy [5]. A widely used method to measure the discrepancy between domains is moments (see in Fig. 1 (a)), either the first moment (maximum mean discrepancy, MMD [6]) or the second moment (correlation alignment, CORAL [7]). Pan *et al.* derived the close form solution with the goal of minimizing MMD and proposed Transfer Component Analysis (TCA) [8]. Long *et al.* further extended TCA by assuming that the conditional probability of two domains are also different, they proposed Joint Distribution Adaptation (JDA) to jointly align both marginal and conditional probability [9]. Sun *et al.* aligned the covariances of two domains and put forward CORrelation ALignment (CORAL) [7]. Since deep learning achieves remarkable performance in various areas, domain adaptation with deep architectures also makes great progress. Many deep methods simply use moments to regularize features in the middle layers. Tzeng *et al.* trained the ALEXNET8 [10] network while adding an MMD layer before the classification layer to align features from two domains [11]. Long *et al.* adapted more layers and adopted Multi-Kernel MMD to improve the performance [12]. Sun *et al.* also extended their CORAL to deep neural networks [13]. Adversarial training is also popular, which trains a feature extractor and a domain discriminator simultaneously [14]–[16].

Although feature-based methods have been widely used, they still need a classifier to make predictions. Naturally, the idea of classifier adaptation comes up, which does not change the feature dimension but deals with the discrepancy when constructing a classifier. Collaborative representation (CR) based classification is an variant of nearest subspace classifier which shows attractive performance in most cases. Adaptation with CR is also a hot topic. Tang *et al.* exploited the local-neighbor geometrical information to learn adaptive representation of target samples [17]. Zhang *et al.* learned a common dictionary for two domains and presented a kernelized CR method for domain adaptation [18]. Nearest regularized CR is proposed for hyperspectral classification [19], but it performs well even in the domain adaptation setting. The key point is that distance between samples (see

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61603299 and 61602385), the National Natural Science Foundation of Shaanxi Province(Grant Nos. 2018JQ6003 and 2018JQ6030) and the China Postdoctoral Science Foundation (Grantnos.2017M613204 and 2017M623246).

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

in Fig. 1 (b)) can measure the discrepancy between domains. However, employing sample-to-sample distance means that we must test the target samples one by one, which makes it time-consuming.

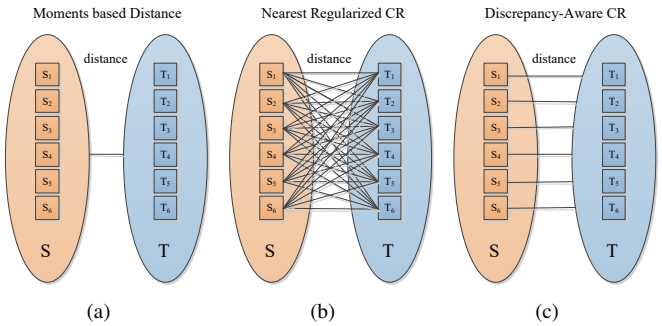

Fig. 1. Graphical illustration of different discrepancy quantization. The oval in orange denotes the source domain, and S_i denotes samples in source domain. Correspondingly, the oval in blue and T_i denote target domain and target samples, respectively. The solid lines denote different distance to measure the domain discrepancy. (a) Moments Based (domain-to-domain), it computes the distance between two domains. (b) Nearest Regularized CR (sample-to-sample), it computes the distance between any two samples in two domains. (c) Discrepancy-Aware CR (sample-to-domain), it computes just one value for each source sample.

From the success of moments based methods and nearest regularized CR, we can learn that the crucial issue for domain adaptation is how to measure the discrepancy and adjust features or classifiers accordingly. In this paper, we propose a CR based method for unsupervised domain adaptation, named Discrepancy-Aware Collaborative Representation (DACR). Our key underlying idea is that samples in source domain should contribute differently due to the distance between each source sample and target domain is different, which is consistent with most instance-based methods [20]. DACR computes domain discrepancy by reconstruction error, then adds them to the sparsity regularization item thus samples closer to the target domain are encouraged to have larger weights. The contributions of the paper can be summarized as follows.

- We proposed a novel domain adaptation solution (DACR) based on collaborative representation. DACR models the discrepancy explicitly and employs them to learn more reasonable representations. Experiments show that the proposed method can achieve reliable performance while greatly improve computation efficiency.
- A discrepancy measurement for characterizing the sample-to-domain relation is presented. Unlike existing methods that measure the distance of domain-to-domain (MMD, CORAL) or sample-to-sample (nearest regularization), our method computes the discrepancy between source samples and target domain. We show that this newly-defined measurement can be derived from domain discrepancy and work like a common regularization.

The rest of the paper is organized as follows. Section 2 details the domain adaptation problem and related works, then introduces collaborative representation based classification. Our method is introduced in Section 3 and experimental evaluation is presented in Section 4. At last, we summarize this paper and discuss future work in Section 5.

II. Related Works

In this section, we begin with a brief introduction to transfer learning and domain adaptation. Then, collaborative representation based classification is given since the proposed method is highly related to it.

A. Transfer Learning and Domain Adaptation

Transfer learning aims to leverage the common knowledge exists in data from different sources so as to fulfill recognition or detection tasks [20]. There are two basic elements in transfer learning, **Domain** ($\mathcal{D}$) and **Task** ($\mathcal{T}$). A domain can be seen as a set of data ($\mathcal{X}$) drawn from distribution $P(\mathcal{X})$. Specifically, there are at least two domains in transfer learning, source domain ($\mathcal{D}_s$) providing a vast amount of knowledge (such as labels in supervised learning) and target domain ($\mathcal{D}_t$) with less or zero knowledge. A task contains the label space ($\mathcal{Y}$) and the mapping function $f(\cdot)$.

Domain adaptation is an attractive branch of transfer learning, which assumes that the distributions of source and target domain are different but related ($P(\mathcal{X}_s) \approx P(\mathcal{X}_t)$). Further, when target domain is completely unlabeled ($\mathcal{Y}_t = \emptyset$), it is called unsupervised domain adaptation. Previous works attempt to reduce the discrepancy caused by different distributions. A general idea is to first evaluate the discrepancy with some measurements and then minimize it. Existing measures can be roughly divided into two categories: 1) Moments based. Both the first order moment (MMD [8]) and the second order moment (CORAL [7]) based measurements are widely used. 2) Adversarial training [21] based. The main idea is to establish a domain discriminator during training.

B. Collaborative Representation based Classification

Sparse representation based face recognition (SRC) [22] shows remarkable performance compared to traditional methods and attracts the attention of many researchers. Zhang *et al.* pointed out the nature that it is the collaborative representation, rather than the l_1-norm sparsity that plays a vital role in classification [23]. Inspired by this, the authors also presented an efficient classification scheme, named collaborative representation based classification (CRC), which has been applied in many areas.

Given a query sample (y), CRC tries to collaboratively represent it with the whole training set ($X = [x_1, x_2 \cdots x_n]$) with coefficients ρ.

$$\rho = argmin_\rho ||y - X \cdot \rho||_2^2 + \lambda ||\rho||_2^2 \tag{1}$$

Intuitively, the first item $||y - X \cdot \rho||_2^2$ indicates that we hope the reconstruction error can be small and $||\rho||_2^2$ is the sparsity item. λ is introduced to balance two objectives. Noticed that the sparsity item uses l_2-norm but not as the l_1-norm used in SRC, the result is a massive saving in time since we can

solve ρ in a closed form. By setting the partial derivative to zero, we can obtain:

$$\rho = (X^T X + \lambda I)^{-1} X^T y \tag{2}$$

where X^T denotes the transpose of X, I denotes the identity matrix with $n-$dimension. To summarize, the authors assign weights on training samples to represent the query sample $y \approx \sum_i X_i \rho_i$.

To make prediction on the query sample, the authors compute the class specific residual and take the label with the minimum residual. The detailed process is in Algorithm 1. c denotes the number of classes, X^i and ρ^i denote the samples belong to class i and their weights, respectively.

Algorithm 1 CrcPre : Predicting on the query sample

Require: X, y, ρ
1: **for** $i \in [1, 2 \cdots c]$ **do**
2: $\quad r_i = ||y - X^i \cdot \rho^i||_2$
3: **end for**
4: $label = \underset{i}{argmin}\ r_i$
5: **return** $label$

Extend from CRC, Li *et al.* proposed nearest regularized CRC which introduces a locality constraint for ρ by giving different freedom to training samples according to their Euclidean distances from the query sample [19].

$$\rho = argmin_\rho ||y - X \cdot \rho||_2^2 + \lambda ||\Gamma \rho||_2^2 \tag{3}$$

where Γ is a biasing Tikhonov matrix defined by $\Gamma_{ii} = ||x_i - y||_2, i = 1, 2 \cdots, n$. Obviously, if x_i differs greatly from the query sample, Γ_{ii} will be large, then the contribution of x_i in reconstruction (ρ_i) should be small because of the constraint $||\Gamma_{ii}\rho_i||_2$.

III. Discrepancy-Aware Collaborative Representation

In this section, we first give a clear definition to domain adaptation problem, then the proposed method is presented in detail.

A. Problem Definition and Notations

Suppose that we have source and target domain data (X_s/X_t) drawn from two different but relational domains ($\mathcal{D}_s/\mathcal{D}_t$). The feature dimensions (m) of the data from two domains are the same. Our goal is to establish a robust classifier ($\mathcal{C}$) with data from both domain ($X_s \& X_t$) and source labels (Y_s), thus make accurate predictions on target samples (X_t). It is worth noting that the target labels (Y_t) are only available when evaluating methods.

B. Methodology

Collaborative representation based classification is a generalization of the nearest neighbors (NN) and nearest subspace (NS) classifiers, its performance degrades in the scenario of domain adaptation problems without the alignment of domains [9]. According to our research, nearest regularized CRC shows trustworthy performance (which can be seen in experimental results) in domain adaptation. This discovery verifies that the Euclidean distance between samples from two domains can measure the domain discrepancy approximately. However, we need to compute the distance between any two samples in two domains, which brings an enormous computational burden.

TABLE I
Notations and descriptions used in this paper.

Notations	Description
$\mathcal{D}_s/\mathcal{D}_t$	source/target domain
X_s/X_t	original source/target domain data
Y_s/Y_t	source/target domain label
n_s/n_t	number of source/target samples
m	feature dimension
$\mathcal{C}$	learned classifier from source domain

In this paper, we introduce an efficient regularization, named Discrepancy-Aware Collaborative Representation (DACR), for unsupervised domain adaptation. The basic principle is that the importance of source samples is different and those close to target domain should contribute more in the collaborative representation. The proposed method consists of two parts: discrepancy quantization and regularization.

1) Discrepancy Quantization: Here we devote to solving the problem that how to evaluate the importance of each source sample. For domain adaptation problem, we can assume that there are both common and domain-specific features [24]. Specifically, the source samples can be represented as $X_s = \mathcal{F}_s + \mathcal{F}_c$ where $\mathcal{F}_s$ denotes source specific features and $\mathcal{F}_c$ denotes common features between domains, so the target samples can be written as $X_s = \mathcal{F}_t + \mathcal{F}_c$ where $\mathcal{F}_t$ denotes target specific features. Obviously, we have $\mathcal{F}_s \perp \mathcal{F}_t$ since they are domain-specific features. A natural idea is that these samples with less domain-specific features should be more important. Based on this, we reconstruct every source samples using target samples, and record the reconstruction error. From the perspective of linear algebra, if we want to represent an vector $v_i = \mathcal{F}_s + \mathcal{F}_c$ with the basis $B = \{\mathcal{F}_t, \mathcal{F}_c\}$, the reconstruction error should be related to $\mathcal{F}_s$ since B is incomplete. If the reconstruction error is small, the sample can be considered of higher importance.

The specific calculation is like a reversed CR. First we collaboratively represent every source sample with the whole target domain. Similarly, the objective can be written as:

$$\rho = argmin_\rho ||X_t \cdot \rho - X_s||_2^2 + \alpha ||\rho||_2^2 \tag{4}$$

Noticed that we employ the matrix form of source sample (X_s) instead of performing one by one , which can save time. By setting the partial derivatives to zero, we can obtain:

$$\rho = (X_t^T X_t + \alpha I)^{-1} X_t^T X_s \tag{5}$$

Then we can calculate the residual with $||X_s - X_t \cdot \rho||$, the specific process is shown in Algorithm 2. $w \in R^{n_s \times n_s}$

indicates the discrepancy between source samples and target domain, and if w_{ii} is large, the discrepancy between x_i and target domain is big, so x_i should contribute less.

Algorithm 2 DisQu : Discrepancy Quantization

Require: X_s, X_t

$\rho = (X_t^T X_t + \alpha I)^{-1} X_t^T X_s$

$r = ||X_s - X_t \cdot \rho||$

$w_{ii} = ||r^i||_2$

return $w = diag(w_i)$

2) Discrepancy Regularization: Since we already have the indicator to quantize the discrepancy of source samples, the regularization is similar to nearest regularization. First we learn the representation to minimize the following equation

$$\underset{\rho}{minimize} \quad J = ||X_t - X_s \cdot \rho||_2^2 + \beta ||w \cdot \rho||_2^2 \tag{6}$$

In the first item we collaboratively represent the query sample X_t by means of all training samples X_s with a small reconstruction error. And in the second item, we use discrepancy-aware sparse regularization rather than treat source samples equally. When there is only one example in target domain, then the second item can be written as:

$$\begin{pmatrix} w_1 & 0 & 0 & 0 \\ 0 & w_2 & 0 & 0 \\ & \cdots & \cdots & \\ 0 & 0 & 0 & w_n \end{pmatrix} \cdot \begin{pmatrix} \rho_1 \\ \rho_2 \\ \cdots \\ \rho_n \end{pmatrix} = \begin{pmatrix} w_1\rho_1 \\ w_2\rho_2 \\ \cdots \\ w_n\rho_n \end{pmatrix} \tag{7}$$

Here we assume that sample **1** has larger discrepancy than sample **2**, which means that $w_1 > w_2$. To minimize this objective, sample **2** is expected to have larger coefficient than sample **1** ($\rho_2 > \rho_1$) in the learned representation.

By setting $\frac{\partial J}{\partial \rho}$ equals to zero, we can solve equation 6 quickly.

$$\begin{aligned} &\frac{\partial J}{\partial \rho} = X_s^T(X_s\rho - X_t) + \beta w^T w \rho = 0 \\ &\Rightarrow (X_s^T X_s + \beta w^T w)\rho = X_s^T X_t \\ &\Rightarrow \rho = (X_s^T X_s + \beta w^T w)^{-1} X_s^T X_t \end{aligned} \tag{8}$$

Then we can make prediction as original CRC (shown in Algorithm 1) does. Algorithm 3 concludes the specific process of the proposed method for unsupervised domain adaptation.

Algorithm 3 DACR : Discrepancy-Aware Collaborative Representation

Require: X_s, Y_s, X_t

Discrepancy quantization.

$w = DisQu(X_s, X_t)$

Regularized collaborative representation.

$\rho = (X_s^T X_s + \beta w^T w)^{-1} X_s^T X_t$

Make predictions.

$\hat{Y}_t^i = CrcPre(X_s, X_t^i, \rho)$

return $\hat{Y}_t$

Compared to original CRC, we add w to the sparsity item to learn more flexible representations. For these samples which have a small value of discrepancy, they are intend to have bigger weights to reconstruct target samples. Besides, we compute just one value to measure the distance between each source sample and target domain while nearest regularized CR need to compute different distances for each target sample. As a result, our method can compute with the whole samples simultaneously while nearest regularized CR can only compute one by one. When the number of target samples is large, this will lead to a huge increase in efficiency.

IV. Experiments

In this section, we conduct extensive experiments for domain adaptation problems to evaluate the proposed approach.

A. Data Preparation

Here we introduce two widely used datasets when comparing domain adaptation methods.

ImageCLEF [1] is an online competition for domain adaptation. It consists of three domains: Caltech (C), ImageNet (I) and Pascal (P), and in each domain there are twelve classes of objects, such as aeroplane, bike and etc. There are 600 images per domain.

Office31 [1] consists of three domains: Amazon (A), Webcam (W) and DSLR (D). There are over 4,000 images from 31 categories.

Fig. 2. Objects in different domains.

By pairing the domains, we can conduct six subtasks for each dataset. For ImageCLEF, we use $C \rightarrow I$ as the short for training with Caltech while testing with ImageNet, and so do others.

B. Baseline Methods

We compare the proposed method with a wide range of state-of-the-art methods to verify its effectiveness. Here we divide existing methods into shallow and deep roughly according to whether there are deep neural networks in it. For shallow methods, we employ the RESNET50 [25] features of images, which use a 2048-dimensional vector to describe an image. Since we use the deep features, we compare to deep methods though there is no deep architecture in the proposed method. And for fair comparison, we apply the feature extractor network (RESNET50) as the backbone for deep methods. The methods are listed as follows.

[1]https://www.imageclef.org/2014/adaptation

TABLE II
ACCURACY (%) ON 12 SUBTASKS COMPARED WITH BOTH SHALLOW AND DEEP METHODS.

Method	$C \to I$	$C \to P$	$I \to C$	$I \to P$	$P \to C$	$P \to I$	$A \to D$	$A \to W$	$D \to A$	$D \to W$	$W \to A$	$W \to D$	$mean$
CRC	83.33	67.00	91.50	74.17	87.83	84.00	75.90	70.82	61.55	95.97	63.05	99.00	79.51
TCA	83.16	71.66	91.50	75.00	84.66	79.50	76.71	69.94	62.87	95.47	63.83	99.20	79.46
GFK	84.83	70.00	91.16	73.50	83.00	76.00	77.51	73.21	59.07	96.35	61.01	99.60	78.75
JDA	**92.00**	75.50	92.33	77.00	82.83	79.16	81.33	81.89	67.09	95.97	**68.41**	99.60	82.76
CORAL	83.00	71.50	88.66	73.66	72.50	72.33	75.10	76.35	50.41	93.46	57.58	99.00	76.13
NRCR	86.33	72.00	93.00	78.00	92.67	89.50	83.33	78.36	63.44	95.35	65.74	98.59	83.02
RES50	78.00	65.50	91.50	74.80	91.20	83.90	68.90	68.40	62.50	96.70	60.70	99.30	78.45
DANN	87.00	74.30	**96.20**	75.00	91.50	86.00	79.70	82.00	**68.20**	96.90	67.40	99.10	83.60
RTN	86.90	72.70	95.30	75.60	92.20	86.80	77.50	**84.50**	66.20	96.80	64.80	99.40	83.22
CAN	89.50	**75.80**	94.20	**78.20**	89.20	87.50	**85.50**	81.50	65.90	**98.20**	63.40	**99.70**	**84.05**
DAN	86.30	69.20	92.80	74.50	89.80	82.20	78.60	80.50	63.60	97.10	62.80	99.60	81.41
DACR	87.67	70.83	92.67	77.50	**92.83**	**90.67**	80.92	78.36	61.95	93.71	64.15	97.99	82.43

TABLE III
RUNNING TIME (S) OF NRCR AND DACR.

Method	$C \to I$	$C \to P$	$I \to C$	$I \to P$	$P \to C$	$P \to I$	$A \to D$	$A \to W$	$D \to A$	$D \to W$	$W \to A$	$W \to D$
NRCR	30.27	30.35	30.56	30.72	30.65	30.43	780.20	1223.93	108.23	30.21	259.25	45.71
DACR	**0.23**	**0.22**	**0.23**	**0.23**	**0.24**	**0.24**	**2.00**	**2.40**	**2.36**	**0.47**	**2.52**	**0.38**
Speed Gains	*127x*	*137x*	*129x*	*131x*	*123x*	*124x*	*388x*	*508x*	*45x*	*63x*	*102x*	*119x*
Target Number	600	600	600	600	600	600	498	795	2817	795	2817	498
Source Number	600	600	600	600	600	600	2817	2817	498	498	795	795

Shallow: CRC [23](Non-adaptation), TCA [8] (Moments based), GFK [1], JDA [9] (Moments based), CORAL [7] (Moments based), NRCR [19] (Nearest regularized and collaborative representation based).

Deep: RES50 [25](Non-adaptation), DANN [21] (Adversarial based), RTN [26] (Adversarial based), CAN [27] (Adversarial based), DAN [12] (Moments based).

C. Implementation Details

The proposed method involves two parameters, α controls the sparsity when computing the discrepancy and β indicates how we care about the regularization. In the experiment, we set $\alpha = 1$ and $\beta = 1$ for all the experiments.

D. Results

The experimental results are shown in Table II. First of all, we focus on the results compared to CRC and RES50. Comparisons with the two non-adaptation methods can verify whether our method achieves improvement when applied to domain adaptation problems. Obviously, our method is superior to both. The mean accuracy increases 3% ($79.51\% \to 82.43\%$) and 4% ($78.45\% \to 82.43\%$) respectively. More precisely, the proposed method has better performance than CRC and RES50 on most subtasks. These discoveries confirm that our method is able to learn transferable representations in different scenarios.

Intuitively, our method has comparable performance to the state-of-the-art when compared with shallow methods. DACR beats TCA, GFK and CORAL on mean accuracy, and loses slightly to JDA ($0.03\%, 82.46\% \to 82.43\%$) and NRCR ($0.6\%, 83.02\% \to 82.43\%$). It may seem strange that the non-adaptation method CRC has higher average accuracy than many DA methods. The reason is the selected shallow methods use dimensionality reduction to learn transferable representations, and the classifier they used is 1-nearest neighbor. Following the experimental setting indicated in the original paper [8], [9], we set the feature dimension to 100, which will cause information loss that cannot be ignored.

To further illustrate the performance of our approach, we use some deep methods for comparison. Generally speaking, deep methods perform better than shallow. The main reason is that end-to-end training can avoid the information loss of converting images into vectors. Nevertheless, our method also shows trustworthy performance, only 1.6% ($84.05\% \to 82.43\%$) lower than the best-performed deep method CAN.

E. Parameter Sensitivity Analysis

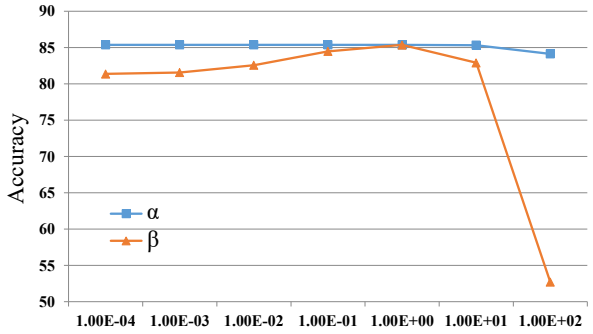

Fig. 3. Average accuracy (%) on ImageCLEF with different values of λ and β.

We conduct sensitivity analysis to validate that DACR can achieve optimal performance under a wide range of parameters, i.e. 1.00E-04 $\to$ 1.00E02. α controls the sparsity when computing reconstruction error to measure domain discrepancy. This sparsity works for controlling the complexity of target coefficients while we mainly focus on the reconstruction error. So α is not an essential parameter, and if it is too large,

the learned representation will be inaccurate due to it pay too much attention to the sparsity. The experimental results also illustrate that. As shown in Fig. 3, the mean accuracy remains unchanged while α increases from 1.00E-04 to 10, and drops slightly when it keep increasing. β is the weight of the regularized sparsity item, so it would have great effect on the accuracy. Fig. 3 presents that the average accuracy rises first and then falls sharply. The performance improves as it increases. But if it is too large, we cannot learn proper representations for query samples. This also shows the validity of the proposed regularization.

F. Running Time

To verify whether the 'sample-to-domain' distance can reduce the computational complexity, we record the running time of NRCR and DACR in Table III. The two methods are implemented via MATLAB 2017a and executed on a Windows PC with Intel Core i7 CPU at 3.6GHz and 8GB RAM. Speed gains indicates how many times the proposed DACR is faster than NRCR, besides, the sample numbers of two domains are reported. Intuitively, the proposed DACR is much faster than NRCR. According to our previous analysis, NRCR computes the distance among every target sample and whole source samples so that the running time is expected to be proportional to target number. There is a significant positive correlation between target numbers and running times for NRCR, such as in subtasks $A \rightarrow D$ and $A \rightarrow W(\frac{780.20}{1223.93} \approx \frac{498}{795})$, in subtasks $D \rightarrow A$ and $D \rightarrow W(\frac{108.23}{30.21} \approx \frac{2817}{795})$, this finding confirms that iterative predictions for target samples limit the calculation efficiency of NRCR a lot. For our method, the importance of source samples is the same for all target samples, so we can calculate the representation at once thus significantly reduce computing burden.

V. Conclusion

In this paper, we propose a solution for unsupervised domain adaptation, named Discrepancy-Aware Collaborative Representation. Unlike existing methods, we do not attempt to reduce the domain discrepancy based on statistical properties or adversarial training. By exploiting the relation of source samples and target domain, we present a novel measurement to characterize domain discrepancy. Sufficient experiments verify that the proposed method is effective. In the future, we plan to combine collaborative representation and deep neural networks to learn meaningful and explainable features.

References

[1] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *Computer Vision and Pattern Recognition*, 2012, pp. 2066–2073.

[2] H. Venkateswara, J. Eusebio, S. Chakraborty, and S. Panchanathan, "Deep hashing network for unsupervised domain adaptation," in *Computer Vision and Pattern Recognition*, 2017, pp. 5385–5394.

[3] S. Sun, B. Zhang, L. Xie, and Y. Zhang, "An unsupervised deep domain adaptation approach for robust speech recognition," *Neurocomputing*, vol. 257, pp. 79–87, 2017.

[4] H. Tang, W. Hsu, F. Grondin, and J. R. Glass, "A study of enhancement, augmentation and autoencoder methods for domain adaptation in distant speech recognition." in *Conference of the International Speech Communication Association*, 2018, pp. 2928–2932.

[5] S. Bendavid, J. Blitzer, K. Crammer, and F. Pereira, "Analysis of representations for domain adaptation," in *Neural Information Processing Systems*, 2006, pp. 137–144.

[6] A. J. Smola, A. Gretton, L. Song, and B. Scholkopf, "A hilbert space embedding for distributions," in *International Conference on Algorithmic Learning Theory*, 2007, pp. 13–31.

[7] B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," in *National Conference on Artificial Intelligence*, 2016, pp. 2058–2065.

[8] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," in *International Joint Conferences on Artificial Intelligence*, 2009, pp. 1187–1192.

[9] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *International Conference on Computer Vision*, 2013, pp. 2200–2207.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Neural Information Processing Systems*, 2012, pp. 1097–1105.

[11] E. Tzeng, J. Hoffman, N. J. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," in *arXiv:1412.3474*, 2014.

[12] M. Long, Y. Cao, Z. Cao, J. Wang, and M. I. Jordan, "Transferable representation learning with deep adaptation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, 2018.

[13] B. Sun and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *European Conference on Computer Vision*, 2016, pp. 443–450.

[14] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Computer Vision and Pattern Recognition*, 2017, pp. 2962–2971.

[15] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional adversarial domain adaptation," in *Neural Information Processing Systems*, 2018, pp. 1640–1650.

[16] Z. Pei, Z. Cao, M. Long, and J. Wang, "Multi-adversarial domain adaptation," in *National Conference on Artificial Intelligence*, 2018, pp. 3934–3941.

[17] S. Tang, Z. Chen, L. Chen, and M. Ye, "Domain adaptation of image classification exploiting target adaptive collaborative local-neighbor representation," in *Active Media Technology*, 2016, pp. 154–157.

[18] G. Zhang, Y. Zheng, and G. Xia, "Domain adaptive collaborative representation based classification," *Multimed. Tools Appl.*, pp. 1–22, 2018.

[19] W. Li, E. W. Tramel, S. Prasad, and J. E. Fowler, "Nearest regularized subspace for hyperspectral classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 52, no. 1, pp. 477–489, 2014.

[20] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.

[21] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. S. Lempitsky, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 189–209, 2016.

[22] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, 2009.

[23] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *International Conference on Computer Vision*, 2011, pp. 471–478.

[24] K. Bousmalis, G. Trigeorgis, N. Silberman, D. Krishnan, and D. Erhan, "Domain separation networks," in *Neural Information Processing Systems*, 2016, pp. 343–351.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[26] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Unsupervised domain adaptation with residual transfer networks," in *Neural Information Processing Systems*, 2016, pp. 136–144.

[27] W. Zhang, W. Ouyang, W. Li, and D. Xu, "Collaborative and adversarial network for unsupervised domain adaptation," in *Computer Vision and Pattern Recognition*, 2018, pp. 3801–3809.