Sparse Embedding-based Domain Adaptation for Object Recognition

Jingjing Zheng, Rama Chellappa
Center for Automatic Research
University of Maryland, College Park MD USA
zjngjng,rama@umiacs.umd.edu

P. Jonathon Phillips*
National Institute of Standards and Technology
Gaithersburg, MD USA
jonathon.phillips@nist.gov

Domain adaptation algorithms aim at handling the shift between source and target domains. A classifier is trained on images from the source domain; and the classifier recognizes objects in images from the target domain. In this paper, we present a joint subspace and dictionary learning framework for domain adaptation. Our approach simultaneously exploits the low-dimensional structures in two domains and the sparsity of features in the projected subspace. Specifically, we first learn domain-specific subspaces from the source and target domains respectively that can decrease the mismatch between source and target domains. Then we project features from each domain onto their domain-specific subspaces. From the projected features, a common domain-invariant dictionary for both domains is learned. Our approach handles domain shift caused by different classes of features; e.g., SURF and SIFT. In addition, the features can have different dimensions. Our framework applies to both cross-domain adaptation (cross-DA) and multiple source domain adaptation (multi-DA). Our experimental results on the benchmark dataset show that our algorithm outperforms the state of the art.

Suppose that we have P domains, and the first $P-1$ domains are source domains while the last domain is the target domain. Let $X_\pi \in \mathbf{R}^{d_\pi \times N_\pi}$ denote the feature representation of N_π samples in the π-th domain, where each column in X_π denotes a sample and d_π is the feature dimension in the π-th domain. Our goal is to jointly learn domain-specific subspaces $W_\pi \in \mathbf{R}^{m \times d_\pi}$ for each domain and a common domain-invariant dictionary $D \in \mathbf{R}^{m \times J}$. The objective function is formulated as follows:

$$\begin{aligned}
&\min_{W_\pi, D, Z_\pi} \sum_{\pi=1}^{P} \{||W_\pi X_\pi - DZ_\pi||_F^2 + \lambda_1 ||Z_\pi||_1\} + \lambda_2 \Psi(Z) \\
&- \sum_{\pi=1}^{P} tr(W_\pi X_\pi L_\pi X_\pi^T W_\pi^T) - \gamma \sum_{\pi}^{P-1} tr(W_\pi M_\pi M_P^T W_P^T) \\
&\quad s.t. \qquad W_\pi W_\pi^T = I, \pi = 1, 2, ..., P
\end{aligned} \tag{1}$$

where $Z = [Z_1, ..., Z_P], L_\pi = I_\pi - \frac{1}{N_\pi} e_\pi e_\pi^T$, $I_\pi \in \mathbf{R}^{N_\pi \times N_\pi}$ is the identity matrix for the π-th domain and all entries of vectors $e_\pi \in \mathbf{R}^{N_\pi \times 1}$ are one. The objective function consists of two parts:

- The first part corresponds to discriminative dictionary learning using the projected features within each domain as shown in the first line of (1). A discriminative regularization of sparse codes $\psi(Z) = tr(S_w(Z) - S_b(Z))$ is used where the within-class scatter of sparse codes $S_w(Z)$ is minimized and the between-class scatter of sparse codes $S_b(Z)$ is maximized.
- The second part consists of two types of regularization of W_π as shown in the second line of (1). The first type of regularization encourages to maximizing the variance of projected features in the π-th domain, whereas the second type of regularization encourages to maximizing the correlation of projected features from the π-th source domain and target domain (*i.e.* the P-th domain). Note that negative trace is used to be consistent with the overall minimization problem. The i^{th} column in $M_\pi \in \mathbb{R}^{d_\pi \times C}$ can be either the class mean from class i or a randomly selected sample of the same class from two domains respectively.

The objective function in (1) can be solved iteratively by updating one variable while fixing the other two variables. First, given a fixed W_π, D, the objective function in (1) is reduced to

$$\min_{Z=[Z_1,...,Z_P]} f(Z) + \lambda_1 ||Z||_1 \tag{2}$$

where $f(Z) = \sum_{\pi=1}^{P} \{||W_\pi X_\pi - DZ_\pi||_F^2 + \lambda_2 \psi(Z)$. Because $f(Z)$ is strictly convex to Z, we solve the sparse coding problem using FISTA [1]. Second, after the update of sparse codes Z, we update D by solving a quadratic programming problem:$\min_D \sum_{\pi=1}^{P} \{||W_\pi X_\pi - DZ_\pi||_F^2$. Note that each column in D is required to be a unit vector. As in [9], the analytical solution of D can be computed as follows: $D^* = (\sum_{\pi=1}^{P} W_\pi X_\pi Z_\pi^T) U^T (UU^T + \Lambda)^{-1}$, $U = \sum_{\pi=1}^{P} Z_\pi Z_\pi^T$. Finally, given fixed D and Z, the objective function in (1) is reduced to

$$\begin{aligned}
&\min_{W_\pi} \sum_{\pi=1}^{P} \{||W_\pi X_\pi - DZ_\pi||_F^2 - \sum_{\pi=1}^{P} tr(W_\pi X_\pi L_\pi X_\pi^T W_\pi^T) \\
&- \gamma \sum_{\pi}^{P-1} tr(W_\pi M_\pi M_P^T W_P^T) \;\; s.t.\, W_\pi W_\pi^T = I, \pi = 1, ..., P
\end{aligned}$$

We couple the constraints with $u_\pi = \frac{tr(I_P)}{tr(I_\pi)}$ as in [6] to obtain a relaxed version of the remaining optimization prob-

*The identification of any commercially product or trade name does not imply endorsement or recommendation by NIST.

Methods	W → D	D → W	A → D
symm [5]	65.2 ± 0.3	61.5 ± 0.2	56.2 ± 0.2
asymm [4]	63.9 ± 0.3	61.7 ± 0.1	58.0 ± 0.2
SGF [3]	63.4 ± 0.5	61.4 ± 0.4	44.4 ± 0.2
GFK [2]	66.3 ± 0.4	61.4 ± 0.4	44.9 ± 0.4
SDDL [7]	57.4 ± 0.3	57.1 ± 0.3	49.6 ± 0.3
JDSDDL	**68.0 ± 0.2**	**65.9 ± 0.2**	**58.1 ± 0.2**

Methods	D + A → W	A + W → D	W + D → A
SGF [3]	52.0 ± 2.5	39.0 ± 1.1	28.0 ± 0.8
FDDL [9]	41.0 ± 2.4	38.4 ± 3.4	19.0 ±1.2
SDDL [7]	57.8 ± 2.4	56.7 ± 2.3	24.1 ± 1.6
JDSDDL	**68.9 ± 1.8**	**69.5 ± 2.0**	**25.2 ± 1.0**

Table 1. **Recognition accuracies under both cross-DA (left table) and multi-DA (right table) settings with the same feature representation.**

Methods	W → D-600	W → D-sift	D-600 → W	D-sift → W	D-600 → A	D-sift→ A
asymm [4]	**60.0 ± 0.3**	51.8 ± 0.1	50.5 ± 0.3	53.0 ± 0.2	19.6 ± 0.2	20.6 ± 0.1
SDDL [7]	51.2 ± 0.3	43.6 ± 0.2	52.1 ± 0.3	52.5 ± 0.2	22.4 ± 0.1	22.2 ± 0.2
JDSDDL	59.5 ± 0.3	**53.3 ± 0.2**	**52.5 ± 0.3**	**54.9 ± 0.3**	**23.5 ± 0.1**	**23.1 ± 0.1**

Table 2. **Recognition accuracies under cross-DA setting with different feature representations.**

Methods	D-600 + A→ W	D-sift + A→ W	A + W → D-600	A + W → D-sift	D-600 + W → A	D-sift + W → A
SDDL [7]	42.9 ± 0.3	44.2 ± 0.4	45.6 ± 0.3	34.7 ± 0.3	17.6 ± 0.3	19.45 ± 0.2
JDSDDL	**54.5 ± 0.2**	**55.2 ± 0.3**	**55.5 ± 0.3**	**54.8 ± 0.4**	**24.7 ± 0.3**	**25.7 ± 0.2**

Table 3. **Recognition accuracies under multi-DA setting with different feature representations.** No extension and extension '-600' correspond to the feature representations using the codebooks of size 800 and 600 for SURF descriptors constructed on *amazon* and *dslr* respectively. Extension '-sift' corresponds to the feature representation using the codebook of size 900 for SIFT descriptors constructed on *dslr*.

lem:

$$\min_{W=[W_1,\ldots,W_P]} J(W) = tr[W(P-C)W^T - 2WQ]$$
$$s.t.\ W \begin{bmatrix} u_1 I_1 & O & O \\ O & \ddots & O \\ O & O & u_P I_P \end{bmatrix} W^T = I \tag{3}$$

where

$$P = \tilde{X}\tilde{X}^T, Q = \tilde{X}(DZ)^T, \tilde{X} = diag(X_1, ..., X_P)$$

$$C = \begin{bmatrix} X_1 L_1 X_1^T & \gamma M_1 M_2^T & \cdots & \gamma M_1 M_P^T \\ \gamma M_2 M_1^T & X_2 L_2 X_2^T & \cdots & \gamma M_2 M_P^T \\ \vdots & \vdots & \ddots & \vdots \\ \gamma M_P M_1^T & \gamma M_P M_2^T & \cdots & X_P L_P X_P^T \end{bmatrix}.$$

As the objective function $J(W)$ in (3) is a differential function, we use the Cayley transform and the corresponding algorithm in [8] to update W while preserving the orthogonality constraint.

We conducted experiments on object recognition using the dataset introduced in [5]. In order to investigate the domain shift caused by different feature types, we extract both SURF and SIFT descriptors for each image. For the SURF descriptors, two codebooks of size 800 and 600 were constructed on the *amazon* and *dslr* domains respectively to encode the SURF descriptors in each image. For the SIFT descriptors, only one codebook of size 900 was constructed on the *dslr* domain. Thus, we have three different feature representations for each image by using the three different codebooks to encode the corresponding descriptors.

We compare our joint domain-specific subspace and domain-invariant dictionary learning method (JDSDDL) with three groups of state-of-the art approaches: (1) transformation-based approaches: symm [5] and asym [4]; (2) manifold-based approaches: SGF [3] and GFK [2]; (3) dictionary-based approaches: SDDL [7]. Note that manifold-based approaches cannot be applied to handle domain shift caused by different feature representations. In addition, transformation-based approaches are not applicable for multiple source domain adaptation. Tables 1, 2 and 3 show the recognition accuracies by different methods under both cross-DA and multi-DA settings with the same or different feature representations. It can be seen that our approach consistently outperforms other approaches. It is interesting to compare the multi-DA combination (D + A → W) and other two cross-DA combinations (D → W, A → W) in Table 1. Our method yields a better performance for multi-DA combination. However, no such improvement can be seen for SDDL.

References

[1] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sciences*, 2009. 1

[2] B. Gong, Y. Shi, F. Sha, and K. Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *CVPR*, 2012. 2

[3] R. Gopalan, R. Li, and R. Chellappa. Domain adaptation for object recognition: An unsupervised approach. In *ICCV*, 2011. 2

[4] B. Kulis, K. Saenko, and T. Darrell. What you saw is not what you get: Domain adaptation using asymmetric kernel transforms. In *CVPR*, 2011. 2

[5] K. Saenko, B. Kulis, M. Fritz, and T. Darrell. Adapting visual category models to new domains. In *ECCV*, 2010. 2

[6] A. Sharma, A. Kumar, H. D. III, and D. W. Jacobs. Generalized multiview analysis: A discriminative latent space. In *CVPR*, 2012. 1

[7] S. Shekhar, V. Patel, H. Nguyen, and R. Chellappa. Generalized domain-adaptive dictionaries. In *CVPR*, 2013. 2

[8] Z. Wen and W. Yin. A feasible method for optimization with orthogonality constraints. *Mathematical Programming*, 2012. 2

[9] M. Yang, L. Zhang, X. Feng, and D. Zhang. Fisher discrimination dictionary learning for sparse representation. In *ICCV*, 2011. 1, 2