Domain Adaptive Subspace Clustering

Mahdi Abavisani
mahdi.abavisani@rutgers.edu
Vishal M. Patel
http://www.rci.rutgers.edu/~vmp93

Electrical and Computer Engineering Department
Rutegrs, The State University of New Jersey
New Brunswick, NJ, USA

Many practical applications in image processing and computer vision require one to analyze and process high-dimensional data. It has been observed that these high-dimensional data can be represented by a low-dimensional subspace. As a result, the collection of data from different classes can be viewed as samples from a union of low-dimensional subspaces. In subspace clustering, given the data from a union of subspaces, the objective is to find the number of subspaces, their dimensions, and the segmentation of the data and a basis for each subspace. In many applications, one has to deal with heterogeneous data. For example, when clustering digits, one may have to process both computer generated as well as handwritten digits. Similarly, when clustering face images collected in the wild, one may have to cluster images of the same individual collected using different cameras and possibly under different resolution and lighting conditions. Clustering of heterogeneous data is difficult because it is not meaningful to directly compare the heterogenous samples with different distributions which may span different feature spaces. In recent years, various domain adaptation methods have been developed to deal with the distributional changes that occur after learning a classifier for supervised and semi-supervised learning [3]. However, to the best of our knowledge, these methods have not been developed for clustering heterogeneous data that lie in a union of low-dimensional subspaces.

In this paper, we present domain adaptive versions of the sparse and low-rank subspace clustering methods (i.e. SSC [1] and LRR [2]). Figure 1 gives an overview of the proposed method. Given data from K different domains, we simultaneously learn the projections and find the sparse or low-rank representation in the projected common subspace. Once the projection matrices and the sparse or low-rank coefficient matrix is found, it can be used for subspace clustering.

Suppose that we are given N_s samples, $\{\mathbf{y}_i^{d_s}\}_{i=1}^{N_s}$, from domain D_s, and N_t samples, $\{\mathbf{y}_i^{d_t}\}_{i=1}^{N_t}$, from domain D_t. Assuming that each sample from domain D_s has the dimension of M_s, let $\mathbf{Y}_s = [\mathbf{y}_1^{d_s}, \ldots, \mathbf{y}_{N_s}^{d_s}] \in \mathbb{R}^{M_s \times N_s}$ denote the matrix of samples from domain D_s. Similarly, let $\mathbf{Y}_t \in \mathbb{R}^{M_t \times N_t}$ denote the matrix containing N_t samples each of dimension M_t from domain D_t. Note that the dimensions of features in D_s and D_t are not required to be the same, i.e., $M_s \neq M_t$. The task of domain adaptive subspace clustering is to cluster the data according to their original subspaces even though they might lie in different domains.

Let $\mathbf{P}_s \in \mathbb{R}^{m \times N_s}$ and $\mathbf{P}_t \in \mathbb{R}^{m \times N_t}$ be mappings represented as matrices that project the data from D_s and D_t to a latent m-dimensional space, respectively. As a result, $\mathbf{P}_s\mathbf{Y}_s$ and $\mathbf{P}_t\mathbf{Y}_t$ lie on an m-dimensional space. Let $\mathbf{G} = [\mathbf{P}_s\mathbf{Y}_s, \mathbf{P}_t\mathbf{Y}_t] = [\mathbf{g}_1, \cdots, \mathbf{g}_{N_s+N_t}] \in \mathbb{R}^{m \times (N_s+N_t)}$ denote the concatenation of the projected samples in the m-dimensional space from both source and target domains. The proposed method takes advantage of the *self-expressiveness property* of the data in the low-dimensional space. Assuming the presence of noise in the projected samples, the sparse ($p = 1$) or low-rank ($p = *$) representation matrix can be found by solving the following optimization problem

	Method	$\{1\} \rightarrow \{2\}$	$\{1\} \rightarrow \{3\}$	$\{2\} \rightarrow \{1\}$	$\{2\} \rightarrow \{3\}$	$\{3\} \rightarrow \{1\}$	$\{3\} \rightarrow \{2\}$	Avg. $\pm$ std.
50-subjects	SSC	56.93	57.14	58.16	57.14	54.69	54.08	56.36 ± 1.60
	CO-SSC	55.91	57.14	58.16	57.75	**52.65**	53.87	55.91 ± 2.22
	DA-SSC	**52.86**	**54.29**	**55.71**	**57.55**	53.67	**50.81**	**54.15** $\pm$ 2.33
	ED-SSC	57.75	58.75	59.59	60.61	54.08	53.67	57.40 ± 2.90
	GM-SSC	54.69	54.69	59.39	58.57	58.78	57.76	57.31 ± 2.09
	LRR	52.04	48.57	53.26	56.53	44.28	43.26	49.66 ± 5.23
	CO-LRR	46.73	47.35	47.96	52.45	54.49	53.27	50.37 ± 3.4
	DA-LRR	**36.76**	**36.12**	**35.51**	**34.69**	**37.55**	**36.12**	**36.13** $\pm$ 0.99
	ED-LRR	42.45	44.29	42.04	49.39	41.43	42.45	43.67 ± 2.96
	GM-LRR	47.76	45.10	47.14	37.96	47.96	49.59	45.92 ± 4.16

Table 1: Average clustering errors on the UMD-AA01 face dataset. The top performing method in each experiment is shown in boldface. Note that $\{1\},\{2\}$ and $\{3\}$ correspond to session 1, session 2 and session 3, respectively. DA-SSC and DA-LRR denote our proposed methods.

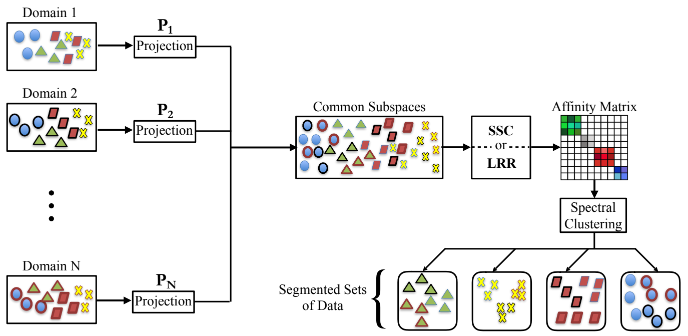

Figure 1: An overview of the proposed domain adaptive subspace clustering framework.

$$\min_{\mathbf{C}} \|\mathbf{C}\|_p + \frac{\tau}{2}\|\mathbf{G} - \mathbf{G}\mathbf{C}\|_F^2, \quad \text{s. t. } \operatorname{diag}(\mathbf{C}) = \mathbf{0}, \tag{1}$$

where the ith column of $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_{N_s+N_t}] \in \mathbb{R}^{N_s+N_t \times N_s+N_t}$ is representation coefficient for $\mathbf{g}_i$ and $\operatorname{diag}(\mathbf{C})$ is the vector of the diagonal elements of $\mathbf{C}$. We propose to learn projections $\mathbf{P}_s$ and $\mathbf{P}_t$ and the representation coefficient matrix $\mathbf{C}$ simultaneously by solving the following optimization problem

$$\min_{\mathbf{P},\mathbf{C}} \|\mathbf{C}\|_1 + \frac{\tau}{2}\|\mathbf{P}\mathbf{Y} - \mathbf{P}\mathbf{Y}\mathbf{C}\|_F^2 \ \text{ s.t.} \operatorname{diag}(\mathbf{C}) = \mathbf{0}, \ \mathbf{P}_s^T\mathbf{P}_s = \mathbf{I}_{N_s}, \mathbf{P}_t^T\mathbf{P}_t = \mathbf{I}_{N_t}$$

where $\tau > 0$ is a parameter, $\mathbf{P} = [\mathbf{P}_s, \mathbf{P}_t]$, and $\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_t \end{bmatrix}$. The constrain, $\mathbf{P}_s^T\mathbf{P}_s = \mathbf{P}_t^T\mathbf{P}_t = \mathbf{I}$, is added to avoid degenerate solutions.

Our algorithms iteratively updates $\mathbf{C}$, and $\mathbf{P}$. $\mathbf{C}$ is updated by solving (1) with ADMM similar to a regular SCC or LRR. Afterwards, we fix the $\mathbf{C}$, and re-write DA-SSC and DA-LRR problems as $\min_{\mathbf{P}} \|\mathbf{P}\mathbf{Y} - \mathbf{P}\mathbf{Y}\mathbf{C}\|_F^2$ s.t. $\mathbf{P}_s^T\mathbf{P}_s = \mathbf{I}_{N_s}, \mathbf{P}_t^T\mathbf{P}_t = \mathbf{I}_{N_t}$, which can be simplified as

$$\min_{\mathbf{P}} \operatorname{Trace}\left(\mathbf{P}[\mathbf{Y}\mathbf{Y}^T - \mathbf{Y}\mathbf{C}^T\mathbf{Y}^T - \mathbf{Y}\mathbf{C}\mathbf{Y}^T + \mathbf{Y}\mathbf{C}\mathbf{C}^T\mathbf{Y}^T]\mathbf{P}^T\right) \tag{2}$$

subject to the constraints $\mathbf{P}_s^T\mathbf{P}_s = \mathbf{I}_{N_s}, \mathbf{P}_t^T\mathbf{P}_t = \mathbf{I}_{N_t}$. This problem involves optimization on Stiefel manifold, hence, we solve it using the manifold optimization technique. Once the coefficient matrix $\mathbf{C}$ met its convergance, the affinity matrix $\mathbf{W} = |\mathbf{C}| + |\mathbf{C}|^T$ is calculated to obtain the segmentation of the heterogeneous data.

We evaluated the performance of our domain adaptive subspace clustering methods on three sets of publicly available datasets - UMD-AA01 face dataset, Amazon-DLSR-Webcam office datasets, and USPS-MNIST-Alphadigits handwritten digits datasets. Table 1 shows how the results of our methods on UMD-AA01 face dataset is compared to state-of-the-art domain adaptive subspace clustering methods.

Aknowledgement: This work was supported by the NSF grant 1618677.

[1] Ehsan Elhamifar and René Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2765–2781, 2013.

[2] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):171–184, 2013.

[3] Vishal M Patel, Raghuraman Gopalan, Ruonan Li, and Rama Chellappa. Visual domain adaptation: A survey of recent advances. *Signal Processing Magazine, IEEE*, 32(3):53–69, 2015.