Diversity-induced Multi-view Subspace Clustering

Xiaochun Cao[1,2], Changqing Zhang[1], Huazhu Fu[3], Si Liu[2], Hua Zhang[1],
[1]School of Computer Science and Technology, Tianjin University, Tianjin 300072, China.
[2]State Key Laboratory of Information Security, IIE, Chinese Academy of Sciences, Beijing, 100093, China.
[3]School of Computer Engineering, Nanyang Technological University, Nanyang Avenue 639798, Singapore.

Multi-view data are very common in many real world applications because data is often collected from diverse domains or obtained from different feature extractors. Taken alone, these views will often be deficient or incomplete because different views describe distinct perspectives of data. Therefore, a key problem for data analysis is how to integrate the multiple views and discover the underlying structures. Recently, some approaches of learning from multi-view data have been proposed. However, most of them concentrate on supervised or semi-supervised learning [7, 11], in which a validation set is required. In this work, we focus on multi-view clustering, which is much more challenging for lacking training information to guide the learning process.

The complementary principle of multi-view setting states that, each view of the data may contain some knowledge that other views do not have. Therefore, multiple views can be employed to comprehensively and accurately describe the data [10]. Furthermore, some theoretical results [1, 2, 9] have shown that the independence of different views can serve as a helpful complement to the multi-view learning. Nevertheless, the main limitation of the existing methods [4, 6, 8] is that they could not guarantee the complementarity across different similarity matrices corresponding to different views.

Figure 1(a-c) illustrates the straightforward way to combine the multi-view features, which independently constructs the similarity matrix of each feature according to some specific distance metric. By contrast, we consider the complementary information of all the different views in depth, and find that the complementary information is explored more thoroughly, while the similarity matrices of the multi-view features are more diverse. Our Diversity-induced Multi-view Subspace Clustering (DiMSC) explores the complementary information, which learns all the different subspace representations jointly with the help of the diversity constraint.

Suppose $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n] \in \mathbb{R}^{d\times n}$ is the matrix of data vectors. To cluster the data into their respective subspaces, we need to compute an similarity matrix that encodes the pairwise similarity between data pairs. Thus, the self-representation manner is written in a compact matrix form

$$\mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}, \tag{1}$$

where $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_n] \in \mathbb{R}^{n\times n}$ is the coefficient matrix with each $\mathbf{z}_i$ being the new representation of sample $\mathbf{x}_i$, and $\mathbf{E}$ is the error matrix. After obtaining the self-representation matrix $\mathbf{Z}$, the similarity matrix $\mathbf{S}$ is usually constructed as [3]:

$$\mathbf{S} = |\mathbf{Z}| + |\mathbf{Z}^T|, \tag{2}$$

where $|\cdot|$ denotes the absolute operator. Afterwards, the similarity matrix is used as the input of spectral clustering algorithm to obtain the final clustering result. The subspace based clustering technique has shown its power in many image processing fields. However, the multi-view representation is ubiquitous and, hence, extending subspace clustering into the multi-view setting is of vital importance for many applications.

Therefore, the objective function of smooth representation clustering corresponding to the v^{th} view turns out to be:

$$\min_{\mathbf{Z}^{(v)}} f(\mathbf{Z}^{(v)}) = ||\mathbf{X}^{(v)} - \mathbf{X}^{(v)}\mathbf{Z}^{(v)}||_F^2 + \alpha^{(v)}\Omega(\mathbf{Z}^{(v)}), \tag{3}$$

where $\alpha^{(v)}$ are tradeoff factors and $\Omega(\cdot)$ denotes the smooth regularized term which is defined as follow:

$$\begin{aligned}\Omega(\mathbf{Z}^{(v)}) &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}^{(v)} ||\mathbf{z}_i^{(v)} - \mathbf{z}_j^{(v)}||_2^2 \\ &= tr(\mathbf{Z}^{(v)}\mathbf{L}^{(v)}{\mathbf{Z}^{(v)}}^T),\end{aligned} \tag{4}$$

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

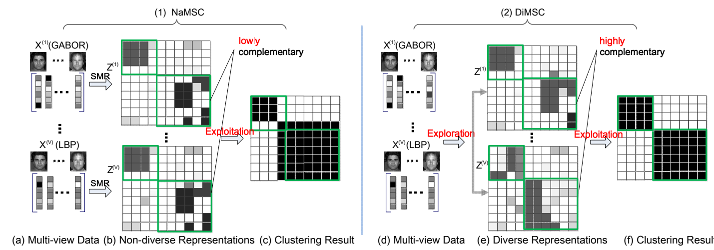

Figure 1: Comparison of naive multi-view subspace clustering (*NaMSC*) and our *DiMSC*. The green rectangle indicates the ground-truth clustering. With the multi-view input (a), NaMSC independently learns the subspace representations using SMR [5] (b), which can not ensure the complementarity across different views. In contrast, our DiMSC employs diverse subspace representations to explore the complementary information across the multiple views, and the final clustering result (f) is obtained.

To enhance the complementary information, in our approach, we encourage the new representations of different views to be of sufficient diversity. This amounts to enforcing the representations of each view to be novel to each other. Let $\mathbf{X}^{(v)}$, $\mathbf{Z}^{(v)}$ denote the features in v^{th} view and corresponding subspace representation, respectively. Then, we should minimize the following objective function:

$$\begin{aligned}\mathcal{O}(\mathbf{Z}^{(1)}, ..., \mathbf{Z}^{(V)}) &= \underbrace{\sum_{v=1}^{V} ||\mathbf{X}^{(v)} - \mathbf{X}^{(v)}\mathbf{Z}^{(v)}||_F^2}_{\text{error}} \\ &+ \underbrace{\lambda_S \sum_{v=1}^{V} tr(\mathbf{Z}^{(v)}\mathbf{L}^{(v)}{\mathbf{Z}^{(v)}}^T)}_{\text{smoothness}} + \underbrace{\lambda_V \sum_{v\neq w} \text{HSIC}(\mathbf{Z}^{(v)}, \mathbf{Z}^{(w)})}_{\text{diversity}},\end{aligned} \tag{5}$$

where λ_S and λ_V are tradeoffs corresponding to the smoothness and diversity regularization terms, respectively. Under the assumption that the data are drawn from different subspaces, the first term ensures the relationships are constructed in the same subspace. The second and third terms enforce that the learned subspace representations to meet the grouping effect independently and diversity jointly.

[1] Avrim Blum and Tom Mitchell. Combining labeled and unlabeled data with co-training. In *COLT*, pages 92–100. ACM, 1998.

[2] K. Chaudhuri, S. M. Kakade, K. Livescu, and K. Sridharan. Multi-view clustering via canonical correlation analysis. In *ICML*, pages 129–136, 2009.

[3] Ehsan Elhamifar and Rene Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *TPAMI*, 35(11):2765–2781, 2013.

[4] Tzortzis G and Likas A. Kernel-based weighted multi-view clustering. In *ICDM*, pages 675–684, 2012.

[5] Han Hu, Zhouchen Lin, Jianjiang Feng, and Jie Zhou. Smooth representation clustering. In *CVPR*, pages 3834–3841, 2014.

[6] A. Kumar and Hal Daumé III. A co-training approach for multi-view spectral clustering. In *ICML*, pages 393–400, 2011.

[7] Novi Quadrianto and Christoph H. Lampert. Learning multi-view neighborhood preserving projections. In *ICML*, pages 425–432, 2011.

[8] W. Tang, Z. Lu, and I. S. Dhillon. Clustering with multiple graphs. In *ICDM*, pages 1016–1021, 2009.

[9] Wei Wang and Zhihua Zhou. Analyzing co-training style algorithms. In *ECML*, pages 454–465, 2007.

[10] Chang Xu, Dacheng Tao, and Chao Xu. A survey on multi-view learning. *arXiv preprint arXiv:1304.5634*, 2013.

[11] Chang Xu, Dacheng Tao, and Chao Xu. Large-margin multi-view information bottleneck. *TPAMI*, 36(8):1559–1572, 2014.