Supplementary Materials For Submission 228

Anonymous Author(s)

A   Appendix

A.1   Details of Algorithms

The optimization problem is rewritten as,

$$\max_{\mathbf{P}} \operatorname{Tr}(\mathbf{Z}_1^\top \mathbf{Z}_2 \mathbf{P} + \lambda \mathbf{S}_1^\top \mathbf{P}^\top \mathbf{S}_2 \mathbf{P}), \text{ s.t. } \mathbf{P}\mathbf{1} = \mathbf{1}, \mathbf{P}^\top \mathbf{1} = \mathbf{1}, \mathbf{P} \in [0,1]^{m \times m}, \tag{1}$$

We refer Eq. (1) as the multi-view anchor correspondence framework. To efficiently solve Eq. (1), we adopt the Projected Fixed-Point Algorithm to update $\mathbf{P}$ as follows,

$$\mathbf{P}^{(t+1)} = (1-\alpha)\mathbf{P}^{(t)} + \alpha \mathbf{\Gamma}\left(\nabla f\left(\mathbf{P}^{(t)}\right)\right) = (1-\alpha)\mathbf{P}^{(t)} + \alpha \mathbf{\Gamma}(\mathbf{K}^\top + 2\lambda \mathbf{S}_2 \mathbf{P}^{(t)} \mathbf{S}_1^\top), \alpha \in [0,1], \tag{2}$$

where α is the step size parameter, t denotes the number of iterations and $\mathbf{\Gamma}$ denotes the double stochastic projection operator. The convergence of the algorithm has been proved, and we set $\alpha = 0.5$ in this paper.

The $\mathbf{\Gamma}$ representing the double stochastic projection operator for given matrix $\mathbf{Q} = \left(\nabla f\left(\mathbf{P}^{(t)}\right)\right)$, where follows [1] with two-step projection as follows,

$$\mathbf{\Gamma}_1(\mathbf{Q}) = \arg\min_{\mathbf{Q}_1} \|\mathbf{Q} - \mathbf{Q}_1\|_{\mathbf{F}}, \text{ s.t. } \mathbf{Q}_1 \mathbf{1} = \mathbf{1}, \mathbf{Q}_1^\top \mathbf{1} = \mathbf{1}. \tag{3}$$

and then the second step is that,

$$\mathbf{\Gamma}_2(\mathbf{Q}) = \arg\min_{\mathbf{Q}_2} \|\mathbf{Q}_2 - \mathbf{Q}\|_{\mathbf{F}}, \quad \text{s.t. } \mathbf{Q}_2 \geq 0. \tag{4}$$

Both of the two subprojections have closed-formed solutions [1] and the von Neumann successive projection lemma [2] guarantees the successive projection converges to the optimum.

A.2   Convergence Rate Proof

Remark 1. *The algorithm to solve the alignment matrix $\mathbf{P}_i$ converges at rate $\frac{1}{2} + \lambda \|(\mathbf{S}_1 \otimes \mathbf{S}_2)\|_{\mathbf{F}}$.*

Proof. By denoting $\mathbf{R}^{(t)} = (\mathbf{K}^\top + 2\lambda \mathbf{S}_2 \mathbf{P}^{(t)} \mathbf{S}_1^\top)$, it is easy to show that $\left\|\mathbf{R}^{(t)} - \mathbf{R}^{(t-1)}\right\|_{\mathbf{F}} \geq \left\|\mathbf{\Gamma}(\mathbf{R}^{(t)}) - \mathbf{\Gamma}(\mathbf{R}^{(t-1)})\right\|_{\mathbf{F}}$. Then, we have that $\left\|\mathbf{P}^{(t+1)} - \mathbf{P}^{(t)}\right\|_{\mathbf{F}} \leq \frac{1}{2}\left\|\mathbf{P}^{(t)} - \mathbf{P}^{(t-1)}\right\|_{\mathbf{F}} + \frac{1}{2}\left\|\mathbf{R}^{(t)} - \mathbf{R}^{(t-1)}\right\|_{\mathbf{F}}$ and $\left\|\mathbf{R}^{(t)} - \mathbf{R}^{(t-1)}\right\|_{\mathbf{F}} = 2\lambda\left\|\mathbf{S}_2 \mathbf{P}^{(t)} \mathbf{S}_1^\top - \mathbf{S}_2 \mathbf{P}^{(t-1)} \mathbf{S}_1^\top\right\|_{\mathbf{F}} = 2\lambda\left\|(\mathbf{S}_1 \otimes \mathbf{S}_2) vec(\mathbf{P}^{(t)} - \mathbf{P}^{(t-1)})\right\|_2 \leq 2\lambda\left\|(\mathbf{S}_2 \otimes \mathbf{S}_1)\right\|_{\mathbf{F}}\left\|\mathbf{P}^{(t)} - \mathbf{P}^{(t-1)}\right\|_{\mathbf{F}}$. Therefore we can obtain that $\frac{\left\|\mathbf{P}^{(t+1)} - \mathbf{P}^{(t)}\right\|_{\mathbf{F}}}{\left\|\mathbf{P}^{(t)} - \mathbf{P}^{(t-1)}\right\|_{\mathbf{F}}} \leq \frac{1}{2} + \lambda\left\|(\mathbf{S}_1 \otimes \mathbf{S}_2)\right\|_{\mathbf{F}}$. □

A.3   Experimental setting

By the way, all the experimental environment are implemented on a desktop computer with an Intel Core i7-7820X CPU and 64GB RAM, MATLAB 2020b (64-bit).

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

A.4 More Experimental Results

In this section, we report in Table 1 the comparison of the clustering performance of LMVSC and our proposed FMVACC before and after using the anchor alignment module.

Effect of Alignment Module: We further verify the effect of the proposed alignment module through experiments on real-world datasets. Different from the experiments in the main text, we use two others commonly used clustering effect evaluation indicators, namely NMI and Fscore, for evaluation. From the results in Table 1, we can notice that the clustering performance of the LMVSC algorithm and our FMVACC is greatly improved by utilizing the alignment module for column alignment. Specifically, in terms of NMI, LMVSC-Aligned achieves 14.46 % (UCI-Digit) and 8.70 % (BDGP) progress compared to its own original version. In terms of Fscore, compared with the FMVACC without the alignment module, FMVACC-Aligned achieves 8.52 % (BDGP) and 6.80 % (YTF-10) progress.

Table 1: Other Clustering performance on the seven benchmarks. 'OM' indicates the out-of-memory failure.

Datasets	Samples	LMVSC-Unaligned	LMVSC-Aligned	Proposed-Unaligned	Proposed-Aligned
		NMI (%)			
3-Sources	169	26.38±0.42	**27.66±4.04**	43.95±3.17	**56.85±5.44**
UCI-Digit	2000	64.89±1.57	**79.35±1.97**	77.81±2.40	**84.33±2.37**
BDGP	2500	18.17±0.07	**26.87±0.12**	27.35±2.53	**36.81±2.69**
SUNRGBD	10335	18.76±0.14	**24.22±0.23**	17.91±0.55	**19.88±0.61**
MNIST	60000	94.06±1.49	**96.45±1.41**	95.90±0.65	**96.74±0.75**
YTF-10	38654	74.71±1.41	**75.87±1.48**	75.55±2.42	**77.13±1.87**
YTF-20	63896	74.63±0.82	**75.90±1.30**	**78.94±1.23**	78.47±1.01
		Fscore (%)			
3-Sources	169	39.17±0.54	**41.33±2.77**	47.21±4.51	**54.71±7.67**
UCI-Digit	2000	57.62±2.17	**75.56±4.59**	73.35±3.95	**83.33±4.29**
BDGP	2500	32.10±0.04	**40.11±0.18**	35.73±2.37	**44.25±2.43**
SUNRGBD	10335	8.84±0.14	**11.01±0.23**	10.74±0.32	**12.94±0.29**
MNIST	60000	95.36±2.70	**96.97±3.05**	97.03±1.52	**97.65±1.68**
YTF-10	38654	65.26±2.06	**66.88±2.47**	64.49±3.72	**71.29±4.22**
YTF-20	63896	57.87±1.37	**60.80±2.52**	**66.51±3.24**	66.40±2.16

A.5 Parameter sensitivity

There are two parameters to tune, including the number of anchors m and the balanced parameter λ. To illustrate the impact of these two parameters on performance, we conduct a comparative experiment on six datasets shown in Fig. 1. For λ, we select 0.0001, 1, 10000 and positive infinity from small to large, indicating the effect of structure correspondence. For the number of anchors, we respectively let m be k, $2k$, and $5k$, where k is the number of clusters. It can be seen from (a) and (b) that when the number of anchors is constant, the clustering performance increases with the decrease of the effect of structure correspondence. As can be seen from (d) and (e), when m is fixed, for the MNIST dataset, $m = 2k$ works best, while on YTF-10, $m = 2k$ achieves the best performance.

References

[1] Yao Lu, Kaizhu Huang, and Cheng-Lin Liu. A fast projected fixed-point algorithm for large graph matching. *Pattern Recognition*, 60:971–982, 2016.

[2] John Von Neumann. *Functional operators: The geometry of orthogonal spaces*, volume 2. Princeton University Press, 1951.

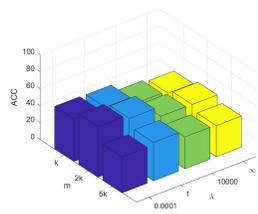
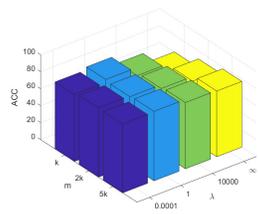
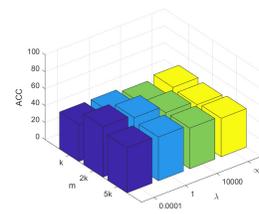
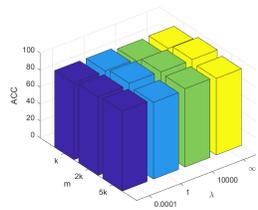
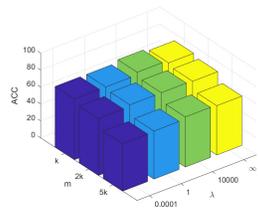
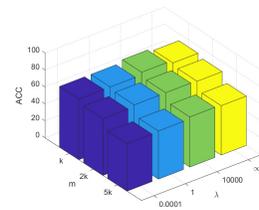
2

(a) 3-Sources   (b) UCI-Digit   (c) BDGP

(d) MNIST   (e) YTF-10   (f) YTF-20

Figure 1: The sensitivity of our proposed method on benchmark datasets.

3