Large-scale Multi-view Tensor Clustering with Implicit Linear Kernels

Supplementary Material

7. Preliminary

In the following, five more tensor operations [19], including bcirc, bvec, bvfold, bdiag and bdfold are introduced. In specific, bcirc is to form the block circular matrix as

$$\mathrm{bcirc}(\mathcal{A}) = \begin{bmatrix} \mathbf{A}^{(1)} & \mathbf{A}^{(N_3)} & \cdots & \mathbf{A}^{(2)} \\ \mathbf{A}^{(2)} & \mathbf{A}^{(N_1)} & \cdots & \mathbf{A}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}^{(N_3)} & \mathbf{A}^{(N_3-1)} & \cdots & \mathbf{A}^{(1)} \end{bmatrix} \tag{22}$$

bvec and bvfold are the block vectorizing and its opposite operation, i.e.

$$\mathrm{bvec}(\mathcal{A}) = \begin{bmatrix} \mathbf{A}^{(1)} \\ \mathbf{A}^{(2)} \\ \vdots \\ \mathbf{A}^{(N_3)} \end{bmatrix}, \quad \mathrm{bvfold}(\mathrm{bvec}(\mathcal{A})) = \mathcal{A}. \tag{23}$$

bdiag and bdfold are the block diagonal matrix and its opposite operation, i.e.

$$\mathrm{bdiag}(\mathcal{A}) = \begin{bmatrix} \mathbf{A}^{(1)} & & \\ & \ddots & \\ & & \mathbf{A}^{(N_3)} \end{bmatrix}, \ \mathrm{bdfold}(\mathrm{bdiag}(\mathcal{A})) = \mathcal{A}. \tag{24}$$

Nevertheless, the tensor unfolding operation along l-th mode is defined as transforming $\mathcal{A}$ to a matrix whose columns are mode-l fibers, i.e. $\mathrm{unfold}_l(\mathcal{A}) = \mathcal{A}_{(l)} \in \mathbb{R}^{N_l \times \prod_{l' \neq l} N_{l'}}$, while the opposite folding operation as $\mathrm{fold}_l(\mathcal{A}_{(l)}) = \mathcal{A}$.

To narrow the gap between Definition 4 and Alg. 1, the following derivations are introduced. Specifically, according to [18, 50], Eq. (3) can be written as

$$\mathcal{A} = \sum_{i=1}^{\min(N_1,N_2)} \mathcal{U}(:,i,:) * \mathcal{S}(i,i,:) * \mathcal{V}(:,i,:)^\top. \tag{25}$$

By considering the t-SVD in Fourier domain, the above decomposition is always transformed to

$$\begin{aligned} \begin{bmatrix} \mathbf{A}_f^{(1)} & & \\ & \ddots & \\ & & \mathbf{A}_f^{(N_3)} \end{bmatrix} &= \begin{bmatrix} \mathbf{U}_f^{(1)} & & \\ & \ddots & \\ & & \mathbf{U}_f^{(N_3)} \end{bmatrix} \cdot \\ \begin{bmatrix} \mathbf{S}_f^{(1)} & & \\ & \ddots & \\ & & \mathbf{S}_f^{(N_3)} \end{bmatrix} &\cdot \begin{bmatrix} \mathbf{V}_f^{(1)} & & \\ & \ddots & \\ & & \mathbf{V}_f^{(N_3)} \end{bmatrix} \end{aligned} \tag{26}$$

in which "$\cdot$"refers to the matrix product and $\mathcal{A}_f = \mathrm{fft}(\mathcal{A}, 3)$ refers to the fast Fourier transform of tensor $\mathcal{A}$ along the 3-rd dimension, while $\mathcal{A} = \mathrm{ifft}(\mathcal{A}_f, 3)$ is the inverse operation. Therefore, the t-SVD of a tensor can be efficiently computed via performing SVD on its fast Fourier transform then transforming the resultant component back to original domain, which is formatted to the pseudocode of Alg. 1.

8. Optimization

Respect to the solution of $\mathcal{G}$-subproblem of Section 4.3, the detailed derivations are provided in the following. With fixing the latent representations $\{\mathbf{H}_v\}_{v=1}^V$, tensor $\mathcal{G}$ and Lagrange multiplier $\mathcal{W}$, the problem of Eq. (14) can be formulated into

$$\min_{\mathcal{G}} \ \frac{\lambda}{\rho}\|\mathcal{G}\|_{\circledast} + \frac{1}{2}\|\mathcal{G} - \mathcal{P}\|_F^2. \tag{27}$$

where $\mathcal{P} = \mathcal{H} - \mathcal{W}/\rho$. In Fourier domain, the optimization problem of Eq. (27) can be transformed into

$$\min_{\mathcal{G}_f} \frac{\lambda}{\rho}\|\mathrm{bdiag}(\mathcal{G}_f)\|_{\circledast} + \frac{1}{2V}\|\mathcal{G}_f - \mathcal{P}_f\|_F^2, \tag{28}$$

which can be decomposed into V sub-optimizations by considering one frontal slice a time, i.e.

$$\min_{\mathcal{G}_f^{(v)}} \frac{\lambda V}{\rho}\|\mathcal{G}_f^{(v)}\|_* + \frac{1}{2}\|\mathcal{G}_f^{(v)} - \mathcal{P}_f^{(v)}\|_F^2. \tag{29}$$

Theorem 1 *[1] For each $\tau \geq 0$ and $\mathbf{Y} \in \mathbb{R}^{n_1 \times n_2}$, the singular value shrinkage operator obeys*

$$\mathcal{D}_\tau(\mathbf{Y}) = \arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{X} - \mathbf{Y}\|_F^2 + \tau\|\mathbf{X}\|_*. \tag{30}$$

Here, the operator can be computed via

$$\mathcal{D}_\tau(\mathbf{Y}) = \mathbf{U}\, \mathcal{C}_\tau(\mathbf{S})\, \mathbf{V}^\top, \tag{31}$$

in which $\mathbf{U}\mathbf{S}\mathbf{V}^\top = \mathbf{Y}$ is the Singular Value Decomposition (SVD) and $\mathcal{C}_\tau(\mathbf{S}) = \max\{0,\ \mathbf{S} - \tau\}$.

It is obvious that Eq. (29) is a Singular Value Thresholding (SVT) problem whose solution is given in Theorem 1. Thus, the solution of $\mathcal{G}_f^{(v)}$ can be obtained as

$$\mathcal{G}_f^{(v)} = \mathcal{U}_f^{(v)} \mathcal{C}_{\rho,\lambda}(\mathcal{S}_f^{(v)}) \mathcal{V}_f^{(v)\top} \tag{32}$$

where

$$\begin{aligned} &\mathcal{U}_f^{(v)} \mathcal{S}_f^{(v)} \mathcal{V}_f^{(v)\top} = \mathcal{P}_f^{(v)} \\ &\mathcal{C}_{\rho,\lambda}(\mathcal{S}_f^{(v)}) = \max\{0,\ \mathcal{S}_f^{(v)} - \lambda V/\rho\}. \end{aligned} \tag{33}$$

Finally, the viarable $\mathcal{G}$ is computed by

$$\mathcal{G} = \mathrm{ifft}(\mathcal{G}_f,\ 3). \tag{34}$$

Table 5. Validation (**NMI**) on inadvertent label use by applying the tensor rotation trick. Note that, the arrows ↓ and ↑ represent performance decrease and increase, respectively.

Dataset	ASR-ETR			S^2MVTC			TBGL			Orth-NTF		
	Sort	Shuffle	Gap	Sort	Shuffle	Gap	Sort	Shuffle	Gap	Sort	Shuffle	Gap
ORL	95.30	90.98	4.32 ↓	*Error*	*Error*	*Error*	79.50	70.15	9.36 ↓	84.45	57.40	27.05 ↓
HW	99.62	76.75	22.87 ↓	92.47	64.40	28.07 ↓	72.15	70.39	1.76 ↓	82.55	58.15	24.40 ↓
BDGP	96.57	23.85	72.72 ↓	97.97	7.44	90.54 ↓	*OT*	*OT*	*OT*	*Error*	*Error*	*Error*
ALOI	91.53	76.08	15.45 ↓	78.87	67.86	11.01 ↓	*OT*	*OT*	*OT*	83.25	51.83	31.42 ↓
DryBean	86.17	61.20	24.96 ↓	76.87	38.16	38.71 ↓	*OT*	*OT*	*OT*	54.08	8.58	45.50 ↓
AwA	85.76	10.91	74.85 ↓	72.46	5.88	66.57 ↓	*OT*	*OT*	*OT*	81.52	2.93	78.59 ↓
YtFace	89.24	1.65	87.59 ↓	74.23	8.84	65.39 ↓	*OT*	*OT*	*OT*	*OT*	*OT*	*OT*

Table 6. Validation (**Purity**) on inadvertent label use by applying the tensor rotation trick. Note that, the arrows ↓ and ↑ represent performance decrease and increase, respectively.

Dataset	ASR-ETR			S^2MVTC			TBGL			Orth-NTF		
	Sort	Shuffle	Gap	Sort	Shuffle	Gap	Sort	Shuffle	Gap	Sort	Shuffle	Gap
ORL	91.25	84.50	6.75 ↓	*Error*	*Error*	*Error*	74.50	63.45	11.05 ↓	69.75	43.15	26.60 ↓
HW	99.85	87.15	12.70 ↓	89.30	63.85	25.45 ↓	77.60	76.00	1.60 ↓	82.55	68.15	14.40 ↓
BDGP	99.04	50.08	48.96 ↓	99.44	33.40	66.04 ↓	*OT*	*OT*	*OT*	*Error*	*Error*	*Error*
ALOI	83.43	61.13	22.30 ↓	55.16	56.92	1.76 ↑	*OT*	*OT*	*OT*	66.82	43.00	23.83 ↓
DryBean	91.17	74.23	16.94 ↓	84.62	56.26	28.37 ↓	*OT*	*OT*	*OT*	73.59	34.02	39.57 ↓
AwA	76.42	11.55	64.87 ↓	57.01	7.29	49.72 ↓	*OT*	*OT*	*OT*	67.03	5.99	61.04 ↓
YtFace	88.36	26.64	61.72 ↓	75.61	27.01	48.59 ↓	*OT*	*OT*	*OT*	*OT*	*OT*	*OT*

9. Dataset and completing methods

In the experiments, we test the proposed LMTC method on seven popular benchmark datasets which are briefly introduced as follows:

1. ORL[2] [39] contains 400 face images from 40 categories, where three types of features, including 4096-D Intensity, 3304-D Local Binary Patterns (LBP) and 6750-D Gabor, are used as different views.
2. HW[3] [42] collects 2000 digits, where six features are extracted, including 76-D Fourier Coefficient, 216-D Profile Correlation, 64-D Karhunen-Love Coefficient, 240-D Pixel Average, 47-D Zernike Moment and 6-D Morphological features.
3. BDGP[4] [40] is a dataset of images of drosophila embryos with 2500 samples of five classes, where each sample has 1000-D, 500-D and 250-D visual features.
4. ALOI[5] [12] consists of 11025 images of 100 small objects where four types of features, including RGB, HSV, Color Similiarity and Haralick features are used.
5. DryBean[6] [20] collects 13611 dry beans of seven different types where two features, including 12-D and 4-D shape forms, are extracted.
6. AwA[7] [21] contains 30475 images of 50 animals classes with six extracted features, including 2688-D Color Histogram, 2000-D Local Self-Similarity, 252-D Pyramid Histogram of Oriented Gradient (PHOG), 2000-D Scale Invariant Feature Transform (SIFT), 2000-D color SIFT and 2000-D Speeded Up Robust Features (SURF) features.
7. YtVideo[8] [34] consists of 101499 Youtube videos with five types of features, including 64-D audio volume, 512-D vision Cuboids Histogram, 64-D vision Histogram (HIST), 647-D vision Histogram of Oriented Gradient (HOG), 838-D vision MISC features.

Meanwhile, the proposed LMTC method is also compared with ten classic and novel large-scale multi-view clustering approaches, including

1. RMKC [2] extends the standard k-means into multi-view setting, as well, employing the structured sparsity-inducing norm to enhance its robustness to data outliers.

[2] https://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
[3] https://archive.ics.uci.edu/ml/datasets/Multiple+Features/
[4] https://www.fruitfly.org/
[5] https://aloi.science.uva.nl/
[6] https://archive.ics.uci.edu/dataset/602/dry+bean+dataset
[7] https://cvml.ist.ac.at/AwA/
[8] http://archive.ics.uci.edu/ml/datasets/YouTube+Multiview+Video+Games+Dataset

2. BMVC [53] collaboratively encodes multi-view data into compact binary representations, then clusters them with binary matrix factorization.
3. LMSC [17] employs anchor technique to approximate the self-representation matrix of subspace clustering algorithm, making it feasible on large-scale multi-view data.
4. OPMC [27] proposes a matrix tri-factorization method to integrate the complementary information of different views by utilizing the discrete label matrix into its objective function.
5. EOMSC [32] combines anchor learning and graph construction into a uniform framework and imposes a graph connectivity constraint, not only boosting the clustering perforamnce but also able to compute the labels directly without any post-processing procedures.
6. MCHBG [54] employs the high-order bipartite graph to reveal richer clustering structures while keep the overall computational complexity in linear to the number of data samples.
7. ASR-ETR [16] constructs the anchor-representation tensor rather than the self-representation strategy to reduce the time complexity and adopts the Anchor Structure Regularization (ASR) and Enhanced Tensor Rank (ETR) to capture the multi-view highorder correlation.
8. S^2MVTC [33] constructs the embedding feature tensor by stacking the embedding features of different views and adopts a novel tensor low-frequency approximation (TLFA) operator to incorporates graph similarity into embedding feature learning.
9. TBGL [49] constructs the bipartite with a variance-based de-correlation anchor selection strategy and exploits the similarity of inter-view by minimizing the tensor Schatten p-norm while that of intra-view by using the $L_{1,2}$-norm minimization regularization and connectivity constraint.
10. Orth-NTF [22] develops a novel multi-view clustering based on orthogonal nonnegative tensor factorization with one-side orthogonal constraint.

10. Result

In Section 4.1 of the main body, only the ACC results are provided. Here we present the NMI and Purity results as supplementary. It can be observed that they follow the same trend with ACC results, further validating the fact that data labels are inadvertently used in existing multi-view tensor clustering approaches.