

Tensor-SVD Based Graph Learning for Multi-View Subspace Clustering

Quanxue Gao,¹ Wei Xia,^{1*} Zhizhen Wan,¹ Deyan Xie,¹ Pu Zhang¹

¹State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China.
qxgao@xidian.edu.cn, xd.weixia@gmail.com, {949414159, 287784562, 845037828}@qq.com

Abstract

Low-rank representation based on tensor-Singular Value Decomposition (t-SVD) has achieved impressive results for multi-view subspace clustering, but it does not well deal with noise and illumination changes embedded in multi-view data. The major reason is that all the singular values have the same contribution in tensor-nuclear norm based on t-SVD, which does not make sense in the existence of noise and illumination change. To improve the robustness and clustering performance, we study the weighted tensor-nuclear norm based on t-SVD and develop an efficient algorithm to optimize the weighted tensor-nuclear norm minimization (WTNNM) problem. We further apply the WTNNM algorithm to multi-view subspace clustering by exploiting the high order correlations embedded in different views. Extensive experimental results reveal that our WTNNM method is superior to several state-of-the-art multi-view subspace clustering methods in terms of performance.

Introduction

Multi-view learning has been an active topic in artificial intelligence and data analysis due to the facts that multi-view data are ubiquitous in real applications and help provide more complementary and discriminative information, compared with single view data (Liu and Tsang 2017; Liu, Tsang, and Müller 2017). Multi-view learning, mainly including multi-view clustering and classification, aims to find meaningful groups of data, are of great importance to various real-world applications (Wang, Nie, and Huang 2013; Cao et al. 2015b; Wang et al. 2018). For a comprehensive review of multi-view learning, please refer to (Yang and Wang 2018). We herein focus on multi-view subspace clustering which achieves impressive results by exploiting the high order correlation embedded in different views.

Multi-view subspace clustering learns affinity matrix or coefficient representation, which is used to adaptively construct graph, by exploiting high order information embedded in data, and then performs spectral clustering on the learned graph. Low-rank representation (LRR) is one of the most

widely used techniques to learn graph for subspace clustering (Guangcan et al. 2013). Inspired by the impressive results of LRR for subspace clustering, Zhang et al. (Zhang et al. 2015) adaptively learned graph, which is shared by different views, by minimizing the nuclear norm of tensor-unfolding matrices that are constructed by affinity matrices, and developed low-rank tensor constrained multi-view subspace clustering (LT-MSc) method. Compared with the aforementioned nuclear norm, tensor nuclear norm based on t-SVD has been proven to be an effective convex relaxation of ℓ_1 -norm (Zhang et al. 2014) and achieved impressive performance for image denoising, video completion, and multi-view subspace clustering (Lu et al. 2016; Yuan et al. 2018; Wu, Lin, and Zha 2019; Hu et al. 2017).

Although t-SVD based tensor nuclear norm has the promising preliminary results for multi-view clustering, it still has some limitations to learn graph. First, it regularizes each singular value equally and employs the soft-thresholding function to shrink all singular values with the same parameter. This, however, ignores the prior knowledge of singular values of the matrix. For example, given an arbitrary image, there is a large difference between its non-zero singular values, especially between the first several large singular values and the last several small singular values, and the larger singular values are generally associated with some salient parts (main information) in the image. Thus, to preserve the salient parts, we should make the large singular values shrink less, which was unfortunately not taken into account in the existing tensor nuclear norm minimization. Second, the existing TNNM tends to preserve some large singular values but discard small singular values. Doing so implies that the larger singular values are considered to be important while the smaller singular values are unimportant. However, such an assumption may not always be true in reality and the larger singular values may carry undesirable information. For example, given an image having large illumination variation in contents, the first a few large singular values contain somewhat illumination information (Belhumeur, Hespanha, and Kriegman 1997), which actually has nothing to do with the content of the image.

The aforementioned facts motivate us to investigate how to leverage the prior information of singular values when us-

*Corresponding author: Wei Xia (xd.weixia@gmail.com)

Copyright © 2020, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

ing tensor nuclear norm, such that the learned graph well exploits high order information embedded in multi-view data. Inspired by t-SVD based nuclear norm, we develop the t-SVD based weighted tensor nuclear norm (WTNN) and study the minimization problem of WTNN (WTNNM). After that, we apply WTNNM to learn graph, which is shared by different views, and develop an enhanced t-SVD based multi-view subspace clustering method. the proposed method well characterizes both the high order information and complementary information embedded in multi-view data. The main contributions of our work are summarized as follows:

- We study the t-SVD based WTNNM which explicitly considers the prior information of singular values, and solve WTNNM with an efficient algorithm which has good convergence. Given the generalization of our model, the existing t-SVD based tensor nuclear norm is a special case of our model. Moreover, our proposed algorithm can also be used to solve the standard matrix nuclear norm minimization.
- We apply our WTNNM to multi-view subspace clustering and develop an enhanced t-SVD based multi-view subspace clustering model. Our method well exploits the high order correlation embedded in multi-view data due to the fact that it well preserves major information of data.
- Our proposed WTNNM can also be applied to other areas in machine learning such as image denoising, matrix completion, and so on.

Multi-view subspace clustering

Multi-view clustering has attracted tremendous attention in machine learning and artificial intelligence (Cao et al. 2015a; Wang et al. 2015; Nie, Cai, and Li 2017). Being the most successful multi-view technique, multi-view subspace clustering aims to learn a new coefficient representation or affinity matrix which is shared by different views. The most representative multi-view subspace model is

$$\min_{\mathbf{Z} \in \mathbb{C}} \sum_{k=1}^V \|\mathbf{X}^k - \mathbf{X}^k \mathbf{Z}^k\|_l + \lambda \Omega(\mathbf{Z}^1, \mathbf{Z}^2, \dots, \mathbf{Z}^V), \quad (1)$$

where \mathbf{X}^k ($k = 1 \dots V$) denotes the data matrix from the k th view. λ is trade-off between the loss function (the first term) and the regularization for coefficient representation (the second term), \mathbb{C} denotes the constraint set on \mathbf{Z}^k , and $\|\bullet\|_l$ represents a proper norm. By employing different norms for loss function and regularization term $\Omega(\mathbf{Z})$, many subspace clustering methods have been developed. For example, using squared F-norm to measure loss function, Gao et al. (Gao et al. 2015) developed a novel multi-view subspace clustering (MVSC) method. MVSC performed spectral clustering on each view coefficient matrix to achieve a common indicator which characterizes the common cluster structure among different views; To well exploit the correlation consensus among multi-views, Wang et al. (Wang et al. 2015) leveraged nuclear and angular regularization to enhance the correlation between the cross-view coefficients within the same subspace to replace the second term in the model (1). To well

exploit both the sparseness of coefficients and complementary information among different views, Wang et al. (Wang et al. 2017) proposed exclusivity-consistency multi-view subspace clustering by imposing sparse constraint on both coefficients and loss function with ℓ_1 -norm.

Apart from the squared F-norm and ℓ_1 -norm, nuclear norm is also a useful measurement and has been proven to be useful in characterizing the spatial structure of data for multi-view subspace clustering. Wang et al. (Wang et al. 2016) proposed a multi-graph laplacian regularized low-rank representation to characterize its local manifold structure, and minimized the divergence between coefficient matrices of different views. Luo et al. (Luo et al. 2018) divided self-representation coefficient matrix of each view into consistency, which has a low-rank structure and is shared different views, and specificity which characterizes the inherent difference in each view, and then proposed consistency-specificity multi-view subspace clustering (CSMSC).

Although the aforementioned multi-view subspace clustering methods have achieved good performance, they ignore the high order correlation underlying multi-view data. Thus, they cannot well exploit the complementary information which is important for clustering. To handle this problem, Zhang et al. (Zhang et al. 2015) proposed a low-rank tensor constrained multi-view subspace clustering (LT-MSC) method which is based on the Tucker tensor decomposition. However, the obtained solution is not a tight convex relaxation of the Tucker rank. To handle this disadvantage, motivated by t-SVD which is an effective convex relaxation, Xie et al. (Yuan et al. 2018) proposed t-SVD based multi-view subspace clustering (t-SVD-MSC). Wu et al. (Wu, Lin, and Zha 2019) employed transition probability matrices corresponding to different views as the tensor input and developed a new method for multi-view clustering (ETLMSC) which is an extension of robust multi-view spectral clustering (RMSC) (Xia et al. 2014). Although the aforementioned tensor nuclear norm based multi-view subspace methods have achieved impressive results, all of them leverage the soft-thresholding function to shrink each singular values with the same parameter. Thus, the prior knowledge of matrix singular values, which is important for clustering, is not in use. To handle this problem, we study the t-SVD based weighted tensor nuclear norm minimization which shrinks singular values with different parameters and describe an efficient algorithm for multi-view subspace clustering which well captures the high order correlation underlying multi-view data.

Notations and preliminaries

For convenience, we first introduce the notations and definitions used throughout the paper. We use bold calligraphy letters for tensors, e.g., $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, bold upper case letters for matrices, e.g., \mathbf{A} , bold lower case letters for vectors, e.g., \mathbf{a} , and lower case letters such as a_{ijk} for the entries of \mathcal{A} . Moreover, we denote $\mathbf{A}^{(i)}$ by the i -th frontal slice of \mathcal{A} and $\bar{\mathcal{A}}$ the discrete Fast Fourier transform (FFT) of \mathcal{A} along the third dimension i.e., $\bar{\mathcal{A}} = \text{fft}(\mathcal{A}, [], 3)$. Similarly, thus \mathcal{A} can be obtained by inverse FFT (IFFT) of

$\bar{\mathcal{A}}$ along the third dimension, i.e., $\mathcal{A} = \text{ifft}(\bar{\mathcal{A}}, [], 3)$.

Definition 1 (Kilmer and Martin 2011) For a 3-way tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the Frobenius norm of \mathcal{A} is $\|\mathcal{A}\|_F = \sqrt{\sum_{ijk} |a_{ijk}|^2}$, and conjugate transpose of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is $\mathcal{A}^T \in \mathbb{R}^{n_2 \times n_1 \times n_3}$.

Definition 2 (Kilmer et al. 2013) For a 3-way tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we denote $\bar{\mathbf{A}}$ as a block diagonal matrix with each block on diagonal as the frontal slice $\bar{\mathbf{A}}^{(i)}$ of $\bar{\mathcal{A}}$. $\bar{\mathbf{A}}$ has the following form:

$$\bar{\mathbf{A}} = \text{bdiag}(\bar{\mathcal{A}}) = \begin{bmatrix} \bar{\mathbf{A}}^{(1)} & & & \\ & \bar{\mathbf{A}}^{(2)} & & \\ & & \ddots & \\ & & & \bar{\mathbf{A}}^{(n_3)} \end{bmatrix}. \quad (2)$$

Definition 3 (Kilmer et al. 2013) For a 3-way tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, its block circulant matrix is a matrix of $n_1 n_3 \times n_2 n_3$ having the following form:

$$\text{bcirc}(\mathcal{A}) = \begin{bmatrix} \mathbf{A}^{(1)} & \mathbf{A}^{(n_3)} & \dots & \mathbf{A}^{(2)} \\ \mathbf{A}^{(2)} & \mathbf{A}^{(1)} & \dots & \mathbf{A}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}^{(n_3)} & \mathbf{A}^{(n_3-1)} & \dots & \mathbf{A}^{(1)} \end{bmatrix}. \quad (3)$$

Definition 4 (Kilmer et al. 2013) For a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we have

$$\text{unfold}(\mathcal{A}) = [\mathbf{A}^{(1)}; \mathbf{A}^{(2)}; \dots; \mathbf{A}^{(n_3)}], \quad (4)$$

$$\text{fold}(\text{unfold}(\mathcal{A})) = \mathcal{A}.$$

Definition 5 (Kilmer et al. 2013) (t-product) Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_2 \times l \times n_3}$, then the t-product between them is $\mathcal{E} \in \mathbb{R}^{n_1 \times l \times n_3}$, i.e.,

$$\mathcal{E} = \mathcal{A} * \mathcal{B} = \text{fold}(\text{bcirc}(\mathcal{A}) \cdot \text{unfold}(\mathcal{B})), \quad (5)$$

t-product between \mathcal{A} and \mathcal{B} can be computed efficiently by

1. Calculate $\bar{\mathcal{A}} = \text{fft}(\mathcal{A}, [], 3)$ and $\bar{\mathcal{B}} = \text{fft}(\mathcal{B}, [], 3)$;
2. Multiply the each pair of the frontal slices of $\bar{\mathcal{A}}$ and $\bar{\mathcal{B}}$ to obtain $\bar{\mathcal{E}}$;
3. Calculate $\mathcal{E} = \text{ifft}(\bar{\mathcal{E}}, [], 3)$;

Definition 6 (Kilmer et al. 2013) A tensor is called *f-diagonal* if each of its frontal slices is diagonal matrix.

Theorem 1 (Kilmer et al. 2013). Block-circulant matrix can be block-diagonalized by

$$(\mathbf{F}_{n_3} \otimes \mathbf{I}_{n_1}) \cdot \text{bcirc}(\mathcal{A}) \cdot (\mathbf{F}_{n_3}^{-1} \otimes \mathbf{I}_{n_2}) = \bar{\mathbf{A}}, \quad (6)$$

where \otimes denotes the Kronecker product, \mathbf{F}_{n_3} is the $n_3 \times n_3$ Discrete Fourier Transform (DFT) matrix, \mathbf{I}_{n_1} and \mathbf{I}_{n_2} denote $n_1 \times n_1$ and $n_2 \times n_2$ identity matrices, respectively.

Theorem 2 (Kilmer et al. 2013) (T-SVD). Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, then \mathcal{A} can be factored as

$$\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T, \quad (7)$$

where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal, $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a f-diagonal tensor.

Definition 7 (Zhang et al. 2014) (tensor nuclear norm). Given $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, its nuclear norm is

$$\|\mathcal{A}\|_{\otimes} = \sum_{i=1}^{n_3} \|\bar{\mathbf{A}}^{(i)}\|_*, \quad (8)$$

where $\|\bar{\mathbf{A}}^{(i)}\|_*$ is nuclear norm of $\bar{\mathbf{A}}^{(i)}$, which is the sum of all singular values of $\bar{\mathbf{A}}^{(i)}$.

WTNNM

Problem Formulation and objective

In real applications (Lu et al. 2016; Yuan et al. 2018; Wu, Lin, and Zha 2019; Hu et al. 2017), most existing t-SVD based tensor low-rank approximation methods involve to solve the following tensor nuclear norm minimization:

$$\argmin_{\mathcal{X}} \frac{1}{2} \|\mathcal{X} - \mathcal{A}\|_F^2 + \tau \|\mathcal{X}\|_{\otimes}. \quad (9)$$

The optimization solution of model (9) can be obtained by solving n_3 independent optimization problems. Denote by r the number of singular values of $\bar{\mathbf{A}}^{(i)}$, the i -th ($i = 1, 2, \dots, n_3$) optimization problem is

$$\argmin_{\bar{\mathbf{X}}^{(i)}} \frac{1}{2} \|\bar{\mathbf{X}}^{(i)} - \bar{\mathbf{A}}^{(i)}\|_F^2 + \sum_{j=1}^r \tau * \sigma_j(\bar{\mathbf{X}}^{(i)}). \quad (10)$$

The most low rank matrix $\bar{\mathbf{X}}^{(i)*}$ in the model (10) can well be recovered by soft-thresholding of singular values of $\bar{\mathbf{A}}^{(i)}$ with the same parameter τ . As the aforementioned in Section 1, this is not very reasonable since different singular values may have different importance and hence they should be treated differently. It indicates that the model (9) may not work well in real applications. To this end, we propose the following weighted tensor nuclear norm minimization (WTNNM) problem

$$\argmin_{\mathcal{X}} \frac{1}{2} \|\mathcal{X} - \mathcal{A}\|_F^2 + \tau \|\mathcal{X}\|_{\omega, \otimes}, \quad (11)$$

where $\|\mathcal{X}\|_{\omega, \otimes}$ is called the weighted nuclear norm of tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, which is defined

$$\|\mathcal{X}\|_{\omega, \otimes} = \sum_{i=1}^{n_3} \|\bar{\mathbf{X}}^{(i)}\|_{\omega, *} = \sum_{i=1}^{n_3} \sum_{j=1}^{\min(n_1, n_2)} \omega_j * \sigma_j(\bar{\mathbf{X}}^{(i)}), \quad (12)$$

where, $\sigma_j(\bar{\mathbf{X}}^{(i)})$ denotes the j largest singular value of $\bar{\mathbf{X}}^{(i)}$, ω_j denotes the j element of vector ω .

It can be seen that, when tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ becomes a 2-order matrix \mathbf{X} ($n_3 = 1$), we have $\mathbf{X} = \bar{\mathbf{X}}^{(1)}$, then the model (12) becomes

$$\|\mathbf{X}\|_{\omega, *} = \sum_{i=1}^{\min(n_1, n_2)} \omega_i \sigma_i(\mathbf{X}),$$

which is called weighted nuclear norm. It indicates that our WTNNM can also be applied to matrix nuclear norm minimization. Moreover, when $\forall i, \omega_i = 1$, WTNNM becomes the existing tensor nuclear norm minimization which has been widely used in multi-view subspace clustering, image denoising, and matrix completion. Thus, WTNNM can also be used in the aforementioned areas.

Optimization

For solving the weighted tensor nuclear norm minimization, i.e., the model (11), we first introduce the following two Theorems.

Theorem 3 Let $\mathbf{Y} = \mathbf{U}_Y * \mathbf{D}_Y * \mathbf{V}_Y^T$ be the SVD of $\mathbf{Y} \in \mathbb{R}^{m \times n}$, $\tau > 0$, $l = \min(m, n)$, the standard weighted nuclear norm minimization is

$$\arg \min_{\mathbf{X}} \frac{1}{2} \|\mathbf{X} - \mathbf{Y}\|_F^2 + \tau \|\mathbf{X}\|_{\omega, *}. \quad (13)$$

Then, the optimal solution of the model (13) is (Chen, Dong, and Chan 2013)

$$\Gamma_{\tau * \omega}[\mathbf{Y}] = \mathbf{U}_Y \mathbf{P}_{\tau * \omega}(\mathbf{Y}) \mathbf{V}_Y^T, \quad (14)$$

where,

$$\begin{aligned} \mathbf{P}_{\tau * \omega}(\mathbf{Y}) &= \text{diag}(\gamma_1, \gamma_2, \dots, \gamma_l), \\ \gamma_i &= \text{sign}(\sigma_i(\mathbf{Y})) \max(\sigma_i(\mathbf{Y}) - \tau * \omega_i, 0). \end{aligned}$$

Theorem 4 For $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $l = \min(n_1, n_2)$, let $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$. For

$$\arg \min_{\mathcal{X}} \frac{1}{2} \|\mathcal{X} - \mathcal{A}\|_F^2 + \tau \|\mathcal{X}\|_{\omega, \otimes}, \quad (15)$$

then, the optimal solution is

$$\mathcal{X}^* = \Gamma_{\tau * \omega}(\mathcal{A}) = \mathcal{U} * \text{fft}(\mathbf{P}_{\tau * \omega}(\bar{\mathcal{A}})) * \mathcal{V}^T, \quad (16)$$

where, $\mathbf{P}_{\tau * \omega}(\bar{\mathcal{A}})$ is a tensor; $\mathbf{P}_{\tau * \omega}(\bar{\mathcal{A}}^{(i)})$ is the i -th frontal slice of $\mathbf{P}_{\tau * \omega}(\bar{\mathcal{A}})$.

Proof: In Fourier domain, the optimization problem can be reformulated as:

$$\bar{\mathcal{X}}^* = \arg \min_{\bar{\mathcal{X}}} \frac{1}{2} \|\bar{\mathcal{X}} - \bar{\mathcal{A}}\|_F^2 + \sum_{i=1}^{n_3} \sum_{j=1}^l \tau * \omega_j * \sigma_j(\bar{\mathcal{X}}^{(i)}). \quad (17)$$

According to Definition 1, we have

$$\arg \min_{\bar{\mathcal{X}}} \sum_{i=1}^{n_3} \left(\frac{1}{2} \|\bar{\mathcal{X}}^{(i)} - \bar{\mathcal{A}}^{(i)}\|_F^2 + \sum_{j=1}^l \tau * \omega_j * \sigma_j(\bar{\mathcal{X}}^{(i)}) \right), \quad (18)$$

where, $\bar{\mathcal{X}}^{(i)}$ is the i -th frontal slice of $\bar{\mathcal{X}}$, $\sigma_j(\bar{\mathcal{X}}^{(i)})$ denotes the j largest singular value of $\bar{\mathcal{X}}^{(i)}$, ω_j is the weighted coefficient of $\sigma_j(\bar{\mathcal{X}}^{(i)})$.

In Eq. (18), each variable $\bar{\mathcal{X}}^{(i)}$ is independent. Thus, it can be divided into n_3 independent subproblems. For the i -th ($i = 1, 2, \dots, n_3$) subproblem, we have

$$\begin{aligned} \bar{\mathcal{X}}^{(i)*} &= \arg \min_{\bar{\mathcal{X}}^{(i)}} \frac{1}{2} \|\bar{\mathcal{X}}^{(i)} - \bar{\mathcal{A}}^{(i)}\|_F^2 + \\ &\quad \sum_{j=1}^l \tau * \omega_j * \sigma_j(\bar{\mathcal{X}}^{(i)}). \end{aligned} \quad (19)$$

According to Theorem 3, the solution of Eq. (19) is $\bar{\mathcal{X}}^{(i)*} = \Gamma_{\tau * \omega}[\bar{\mathcal{A}}^{(i)}] = \bar{\mathbf{U}}^{(i)} \mathbf{P}_{\tau * \omega}(\bar{\mathcal{A}}^{(i)}) \bar{\mathbf{V}}^{(i)T}$, which is the i -th frontal slice of $\bar{\mathcal{X}}^*$. According to Definition 5, we can easily get

$$\mathcal{X}^* = \Gamma_{\tau * \omega}[\mathcal{A}] = \mathcal{U} * \text{fft}(\mathbf{P}_{\tau * \omega}(\bar{\mathcal{A}})) * \mathcal{V}^T, \quad (20)$$

where $\mathcal{U} = \text{fft}(\bar{\mathcal{U}}, [], 3)$ and $\bar{\mathcal{U}}^{(i)}$ is the i -th frontal slice of $\bar{\mathcal{U}}$, $\mathcal{V} = \text{fft}(\bar{\mathcal{V}}, [], 3)$ and $\bar{\mathcal{V}}^{(i)}$ is the i -th frontal slice of $\bar{\mathcal{V}}$. ■

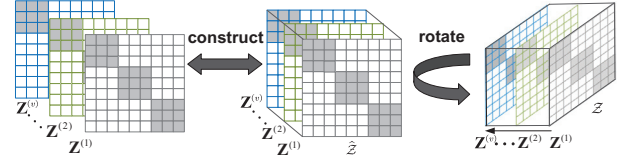


Figure 1: Tensor \mathcal{Z} Construction.

Multi-view Clustering based on WTNNM

Tensor nuclear norm minimization has been widely used in multi-view clustering and obtains impressive experimental results (Yuan et al. 2018; Wu, Lin, and Zha 2019; Hu et al. 2017). However, none of the existing multi-view clustering methods, which are based on tensor low-rank constraint, takes into account the prior knowledge of matrix singular values, which is important for clustering. To handle this limitation, we proposed a new Multi-view Subspace Clustering by WTNNM. The objective function is

$$\begin{aligned} \min_{\mathbf{Z}^{(v)}, \mathbf{E}^{(v)}} & \lambda \|\mathbf{E}\|_{2,1} + \|\mathcal{Z}\|_{\omega, \otimes} \\ \text{s.t. } & \mathbf{X}^{(v)} = \mathbf{X}^{(v)} \mathbf{Z}^{(v)} + \mathbf{E}^{(v)}, v = 1, \dots, V \\ & \mathcal{Z} = \Phi(\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}, \dots, \mathbf{Z}^{(V)}), \\ & \mathbf{E} = [\mathbf{E}^{(1)}; \mathbf{E}^{(2)}; \dots; \mathbf{E}^{(V)}], \end{aligned} \quad (21)$$

where the function $\Phi(\bullet)$ constructs the tensor \mathcal{Z} by merging different representation $\mathbf{Z}^{(v)}$ to a 3-way tensor, and then rotate its dimensionality to $N \times V \times N$ (See Figure 1). Thus, we have

$$\Phi_{(v)}^{-1}(\mathcal{Z}) = \mathbf{Z}^{(v)}, \quad (22)$$

where $\Phi_{(v)}^{-1}(\bullet)$ denotes the inverse function of $\Phi(\bullet)$, and its subscript (v) means to extract the v -th frontal slice.

The above optimization problem can be solved by using the Augmented Lagrange Multiplier (ALM). To adopt alternating direction minimizing strategy to the model (21), we introduce an auxiliary tensor variable \mathcal{J} and rewrite it as minimizing the following unconstrained problem.

$$\begin{aligned} L(\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}, \dots, \mathbf{Z}^{(V)}; \mathbf{E}^{(1)}; \mathbf{E}^{(2)}; \dots; \mathbf{E}^{(V)}; \mathcal{J}) &= \lambda \|\mathbf{E}\|_{2,1} + \|\mathcal{J}\|_{\omega, \otimes} \\ &+ \langle \mathcal{W}, \mathcal{Z} - \mathcal{J} \rangle + \frac{\rho}{2} \|\mathcal{Z} - \mathcal{J}\|_F^2 \\ &+ \sum_{v=1}^V \langle \mathbf{Y}_v, \mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} - \mathbf{E}^{(v)} \rangle \\ &+ \sum_{v=1}^V \frac{\mu}{2} \|\mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} - \mathbf{E}^{(v)}\|_F^2. \end{aligned} \quad (23)$$

where the matrix \mathbf{Y}_v and tensor \mathcal{W} represent two Lagrange multipliers, μ and ρ are actually the penalty parameters.

The alternative minimization scheme is adopted for updating $\mathbf{Z}^{(v)}$, $\mathbf{E}^{(v)}$ and \mathcal{J} , respectively. The main procedure can be partitioned into the following three steps.

$\mathbf{Z}^{(v)}$ -subproblem (variables \mathbf{E} and \mathcal{J} are fixed):

Since $\Phi_{(v)}^{-1}(\mathcal{J}) = \mathbf{J}^{(v)}$, $\Phi_{(v)}^{-1}(\mathcal{W}) = \mathbf{W}^{(v)}$, the model (23) becomes

$$\begin{aligned} \min_{\mathbf{Z}^{(v)}} & \langle \mathbf{Y}_v, \mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} - \mathbf{E}^{(v)} \rangle \\ &+ \frac{\rho}{2} \|\mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} - \mathbf{E}^{(v)}\|_F^2 \\ &+ \langle \mathbf{W}^{(v)}, \mathbf{Z}^{(v)} - \mathbf{J}^{(v)} \rangle + \frac{\rho}{2} \|\mathbf{Z}^{(v)} - \mathbf{J}^{(v)}\|_F^2. \end{aligned} \quad (24)$$

By setting the derivative of Eq. (24) to zero, we have

$$\mathbf{Z}^{(v)*} = (\mathbf{I} + \frac{\mu}{\rho} \mathbf{X}^{(v)T} \mathbf{X}^{(v)})^{-1} \left((\mathbf{X}^{(v)T} \mathbf{Y}_v + \mu \mathbf{X}^{(v)T} \mathbf{X}^{(v)} - \mu \mathbf{X}^{(v)T} \mathbf{E}^{(v)} - \mathbf{W}^{(v)}) / \rho + \mathbf{J}^{(v)} \right). \quad (25)$$

E^(v)-subproblem (other variables $\mathbf{Z}^{(v)}$ and \mathcal{J} are fixed):
In this case, the model (23) becomes

$$\begin{aligned} \arg \min_{\mathbf{E}} \lambda \|\mathbf{E}\|_{2,1} + \sum_{v=1}^V \langle \mathbf{Y}_v, \mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} - \mathbf{E}^{(v)} \rangle \\ + \sum_{v=1}^V \frac{\mu}{2} \|\mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} - \mathbf{E}^{(v)}\|_F^2 \\ = \arg \min_{\mathbf{E}} \frac{\lambda}{\mu} \|\mathbf{E}\|_{2,1} + \frac{1}{2} \|\mathbf{E} - \mathbf{D}\|_F^2. \end{aligned} \quad (26)$$

The optimal solution is (Guangcan et al. 2013)

$$\mathbf{E}_{:,i}^* = \begin{cases} \frac{\|\mathbf{D}_{:,i}\|_2 - \frac{\lambda}{\mu}}{\|\mathbf{D}_{:,i}\|_2} \mathbf{D}_{:,i} & \|\mathbf{D}_{:,i}\|_2 > \frac{\lambda}{\mu} \\ \mathbf{0} & \text{otherwise} \end{cases} \quad (27)$$

where $\mathbf{D}_{:,i}$ denotes the i -th column of $\mathbf{D} = [\mathbf{D}^1; \dots; \mathbf{D}^V]$, $\mathbf{D}^j = \mathbf{X}^{(j)} - \mathbf{X}^{(j)} \mathbf{Z}^{(j)} + \frac{1}{\mu} \mathbf{Y}_j$, $j = 1, \dots, V$.

J-subproblem (Variables $\mathbf{Z}^{(v)}$ and \mathbf{E} are fixed):
In this case, the model (23) becomes

$$\mathcal{J}^* = \arg \min_{\mathcal{J}} \frac{1}{\rho} \|\mathcal{J}\|_{\omega, \oplus} + \frac{1}{2} \left\| \mathcal{J} - \left(\mathcal{Z} + \frac{1}{\rho} \mathcal{W} \right) \right\|_F^2. \quad (28)$$

According to Theorem 4, the solution is

$$\mathcal{J}^* = \Gamma_{\tau^* \omega} \left(\mathcal{Z} + \frac{1}{\rho} \mathcal{W} \right). \quad (29)$$

Additionally, Lagrange multipliers \mathbf{Y}_v and \mathcal{W} need to be updated as follows:

$$\mathbf{Y}_v^* = \mathbf{Y}_v + \mu (\mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} - \mathbf{E}^{(v)}), \quad (30)$$

$$\mathcal{W}^* = \mathcal{W} + \rho (\mathcal{Z} - \mathcal{G}). \quad (31)$$

Finally, we summarize the pseudo code in Algorithm 1.

Experimental Results and Analysis

Database and Competitors

Database We use five common databases, which involve different clustering tasks, including digit clustering, face clustering, scene clustering. We briefly introduce these databases as follows. (1) **Yale**¹ face database has 165 grayscale images of 15 persons. Each subject includes 11 different images about expression and configuration. Just as (Luo et al. 2018), we extract three types of features: 4096 dimensions intensity feature, 3304 LBP feature, and 6750 dimensions Gabor feature. (2) **Notting-Hill** video based face dataset (Zhang et al. 2009) is constructed from the movie 'Notting Hill', where the faces of 5 main casts are collected, including 4660 face in 76 tracks. We randomly choose 110 images of each cast, and extract LBP, Gabor and

¹<http://vision.ucsd.edu/content/yale-face-database>

Algorithm 1 WTNNM for Multi-view clustering

Input: Multi-view feature matrices: $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(V)}$, λ, ω , and cluster number K
Output: Clustering result C
Initialized: $\mathbf{Z}^{(v)} = \mathbf{0}, \mathbf{E}^{(v)} = \mathbf{0}, \mathbf{Y}_v = \mathbf{0}, i = 1, \dots, V, \mathcal{J} = \mathcal{W} = \mathbf{0}, \mu = 10^{-5}, \rho = 10^{-4}, \eta = 2, \mu_{\max} = \rho_{\max} = 10^{10}, \varepsilon = 10^{-7}$
while not converge **do**
(1) Update \mathbf{Z}_{k+1} by Eq. (25)
(2) Update \mathbf{E} by solving Eq. (26)
(3) Update \mathbf{Y}_{k+1} by Eq. (30)
(4) Obtain $\mathcal{Z}_{k+1} = \Phi(\mathbf{Z}_{k+1}^{(1)}, \mathbf{Z}_{k+1}^{(2)}, \dots, \mathbf{Z}_{k+1}^{(V)})$
(5) Update \mathcal{J}_{k+1} by solving Eq. (29)
(6) Update \mathcal{W}_{k+1} by solving Eq. (31)
(7) Update parameters μ_{k+1} and ρ_{k+1} : $\mu_{k+1} = \min(\eta \mu_k, \mu_{\max}), \rho_{k+1} = \min(\eta \rho_k, \rho_{\max})$
(8) Obtain $(\mathbf{J}_{k+1}^{(1)}, \mathbf{J}_{k+1}^{(2)}, \dots, \mathbf{J}_{k+1}^{(V)}) = \Phi^{-1}(\mathcal{J}_{k+1})$
(9) Check the convergence conditions:
 $\left\| \mathbf{X}_{k+1}^{(v)} - \mathbf{X}_{k+1}^{(v)} \mathbf{Z}_{k+1}^{(v)} - \mathbf{E}_{k+1}^{(v)} \right\|_{\infty} < \varepsilon$
 $\left\| \mathbf{Z}_{k+1}^{(v)} - \mathbf{J}_{k+1}^{(v)} \right\|_{\infty} < \varepsilon$
end
(10) Obtain the affinity matrix by
 $\mathbf{S} = \frac{1}{V} \sum_{v=1}^V |\mathbf{Z}^{(v)}| + |\mathbf{Z}^{(v)T}|$
(11) Obtain the clustering result C using the spectral clustering method with the affinity matrix \mathbf{S} .

intensity features following the work in (Luo et al. 2018). (3) **Caltech-101** database² contains 8677 images of objects belonging to 101 categories, with about 40 to 800 images per category. In the experiments, we select the widely used 7 classes, i.e. Face, Motorbikes, Dolla-Bill, Garfield, Snoopy, Stop-sign and Windsor-Chair, leading to 1474 images, and extract 620 dimension HOG feature, 1160 dimension LBP feature and 2560 dimension Sift feature. (4) **UCI-Digits** (Asuncion and Newman 2007) consists of 2000 digits images corresponding to 10 classes. There are three different views. (5) **Scene-15** has 15 natural scene categories with both indoor and outdoor environments, including industrial, store, bedroom, kitchen and so on. There are 4485 images features and three different views (Yuan et al. 2018).

Competitors To evaluate and compare the performance of our work on the mentioned five databases, we choose the following clustering algorithms as competitors: (1) Best Single View (termed as SC): Choosing the best result as final result among all results which can be obtained by using spectral clustering with a single view of data respectively. (2) Feature Concatenation (Feature): Performing the spectral clustering on the new feature representation constructed by concatenating the feature of each view. (3) RMSC (Xia et al. 2014). (4) CSMSC (Luo et al. 2018). (5) MLAN (Nie, Cai, and Li 2017). (6) t-SVD-MS (Yuan et al. 2018). (7)

²http://www.vision.caltech.edu/Image_Datasets/Caltech101/

Table 1: Experimental results on the Yale, the Notting-Hill, the Caltech-101, the UCI-Digits and Scene-15.

Dataset	Metric	SC	Feature	MLAN	RMSC	CSMSC	t-SVD-MSC	ETLMSC	Our work
Yale	ACC	0.556±0.044	0.539±0.043	0.498±0.003	0.703±0.037	0.750±0.002	0.874±0.013	0.659±0.043	0.979±0.000
	NMI	0.592±0.034	0.588±0.026	0.567±0.003	0.723±0.020	0.783±0.002	0.918±0.010	0.697±0.038	0.975±0.000
	Purity	0.567±0.044	0.551±0.038	0.521±0.004	0.709±0.035	0.750±0.002	0.883±0.012	0.659±0.043	0.979±0.000
	F-Score	0.403±0.042	0.391±0.035	0.312±0.003	0.563±0.029	0.639±0.001	0.834±0.020	0.533±0.049	0.958±0.000
	Recall	0.427±0.038	0.413±0.034	0.438±0.016	0.592±0.029	0.671±0.004	0.865±0.018	0.550±0.048	0.959±0.000
	AR	0.362±0.045	0.349±0.038	0.253±0.004	0.533±0.031	0.615±0.001	0.823±0.022	0.501±0.053	0.955±0.000
Notting-Hill	ACC	0.837±0.057	0.725±0.080	0.720±0.064	0.818±0.027	0.927±0.000	0.965±0.000	0.942±0.000	0.984±0.000
	NMI	0.715±0.016	0.615±0.065	0.707±0.058	0.763±0.034	0.832±0.000	0.919±0.000	0.902±0.000	0.956±0.000
	Purity	0.852±0.025	0.756±0.049	0.768±0.064	0.842±0.014	0.927±0.000	0.965±0.000	0.942±0.000	0.984±0.000
	F-Score	0.770±0.047	0.675±0.076	0.720±0.059	0.807±0.046	0.883±0.000	0.935±0.000	0.905±0.000	0.973±0.000
	Recall	0.766±0.037	0.673±0.072	0.835±0.012	0.805±0.054	0.880±0.000	0.922±0.000	0.887±0.000	0.968±0.000
	AR	0.706±0.061	0.585±0.099	0.626±0.085	0.754±0.058	0.850±0.000	0.918±0.000	0.879±0.000	0.966±0.000
Caltech-101	ACC	0.545±0.030	0.558±0.004	0.587±0.000	0.529±0.029	0.567±0.000	0.556±0.000	0.642±0.000	0.830±0.000
	NMI	0.441±0.019	0.438±0.009	0.492±0.000	0.371±0.012	0.480±0.000	0.512±0.000	0.610±0.000	0.880±0.000
	Purity	0.624±0.022	0.624±0.012	0.655±0.000	0.565±0.015	0.633±0.000	0.649±0.000	0.739±0.000	0.909±0.000
	F-Score	0.498±0.017	0.484±0.007	0.494±0.000	0.432±0.024	0.495±0.000	0.498±0.000	0.617±0.000	0.861±0.000
	Recall	0.540±0.020	0.511±0.009	0.592±0.000	0.461±0.052	0.498±0.000	0.509±0.000	0.638±0.000	0.834±0.000
	AR	0.390±0.024	0.376±0.012	0.347±0.000	0.313±0.022	0.395±0.000	0.396±0.000	0.539±0.000	0.835±0.000
UCI-Digits	ACC	0.680±0.040	0.541±0.029	0.970±0.000	0.892±0.053	0.927±0.000	0.995±0.000	0.994±0.000	0.998±0.000
	NMI	0.641±0.012	0.555±0.014	0.934±0.000	0.842±0.020	0.855±0.000	0.986±0.000	0.985±0.000	0.993±0.000
	Purity	0.686±0.032	0.570±0.022	0.970±0.000	0.898±0.040	0.927±0.000	0.995±0.000	0.994±0.000	0.998±0.000
	F-Score	0.576±0.022	0.451±0.013	0.941±0.000	0.828±0.036	0.862±0.000	0.990±0.000	0.988±0.000	0.995±0.000
	Recall	0.587±0.016	0.468±0.013	0.942±0.000	0.837±0.027	0.865±0.000	0.990±0.000	0.988±0.000	0.995±0.000
	AR	0.529±0.025	0.388±0.015	0.934±0.000	0.809±0.040	0.847±0.000	0.989±0.000	0.987±0.000	0.994±0.000
Scene-15	ACC	0.680±0.040	0.312±0.016	0.340±0.031	0.519±0.000	0.509±0.000	0.892±0.000	0.871±0.000	0.904±0.000
	NMI	0.456±0.012	0.288±0.006	0.486±0.029	0.488±0.000	0.564±0.000	0.919±0.000	0.891±0.000	0.932±0.000
	Purity	0.534±0.020	0.351±0.016	0.351±0.034	0.559±0.000	0.611±0.000	0.922±0.000	0.906±0.000	0.933±0.000
	F-Score	0.374±0.015	0.214±0.005	0.262±0.033	0.402±0.000	0.433±0.000	0.883±0.000	0.853±0.000	0.901±0.000
	Recall	0.372±0.014	0.217±0.004	0.740±0.043	0.399±0.000	0.489±0.000	0.892±0.000	0.863±0.000	0.916±0.000
	AR	0.328±0.017	0.155±0.005	0.167±0.041	0.358±0.000	0.385±0.000	0.874±0.000	0.842±0.000	0.893±0.000

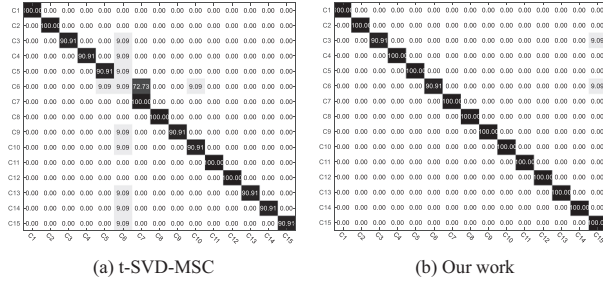


Figure 2: Confusion matrices on Yale database

ETLMSC (Wu, Lin, and Zha 2019).

Parameter Setting In our experiments, we tune the parameter λ in the range of $[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 5, 10, 100]$, and the weight $\omega_i \in (0, 100]$ to get the best results. Specifically, λ is set to 1, the weight vector ω is set to (1; 10; 100) on Yale dataset; λ is set to 0.5, the weight vector ω is set to (1; 10; 100) on Notting-Hill dataset; λ is set to 0.01, the weight vector ω is set to (0.5; 5; 10) on Caltech-101 dataset; λ is set to 0.01, the weight vector ω is set to (1; 10; 100) on UCI-Digits dataset; λ is set to 0.005, the weight vector ω is set to (0.5; 1; 5) on Scene-15 dataset. For all the compared methods, we follow the experiments in the corresponding paper. We run each experiment 10 times and report the average and standard deviation.

Experimental Results and Analysis

To adequately evaluate the performance of clustering, we use six commonly used indexes including accuracy(ACC), normalized mutual information (NMI), purity, F-score, Recall, and Adjusted Rand index (AR). We run each experiment 10 times, Table 1 lists the results on all the five datasets. From Table 1, we have the following observations. 1) Our method achieves the best performance on all the five databases. For example, on Yale dataset, our method indicates a significant increase of 10.5%, 5.7%, 9.6%, 12.4%, 9.4%, and 13.2% w.r.t ACC, NMI, Purity, F-score, Recall and AR, respectively, compared to the corresponding second best baseline. On Scene-15 dataset with 4485 images in three views, our method shows 1.2%, 1.3%, 1.1%, 1.8%, 2.4%, and 1.9% of relative improvement w.r.t ACC, NMI, Purity, F-score, Recall and AR over the corresponding second best baseline. 2) The tensor based methods, including t-SVD-MSC, ETLMSC and our method, achieve significant improvement compared with all other methods in most cases. This great improvement is attributed to effectiveness of high order correlation exploration. In addition, the complementary information among different views can be explored more efficiently and thoroughly by the tensor based methods. 3) Multi-view methods mostly achieve superior performance gain over the standard spectral clustering method based on single view, which demonstrates the necessity of combining multiple views for clustering. MLAN achieves lower performance than the best spectral clustering method on Yale database, Notting-Hill

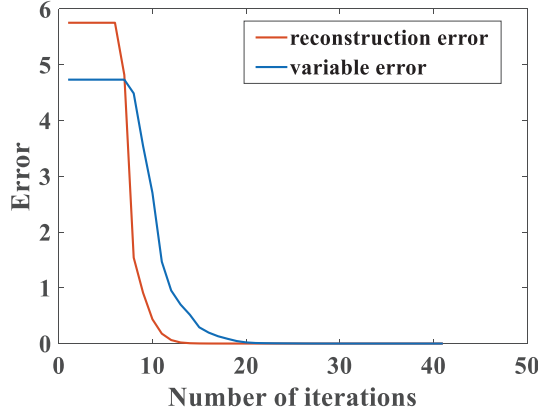


Figure 3: Convergence curves on the Caltech-101 database

database and Scene-15 database. It is probably because the heterogeneous information underlying different views leads to poor similarity matrix by MLAN. And feature concatenation performs poorly than the best single view on all the five datasets except Caltech-101. It is mainly because feature concatenation directly concatenates all the views, but ignores the relation between different views. This leads to the redundancy of view features. Moreover, confusion matrices performed by t-SVD-MSC and our method are shown in Figure 2, where row and column are true and predicted labels respectively. Here, the cluster label is predicted by the best permutation mapping function used in the metric of ACC. We can see that, compared with t-SVD-MSC, our method wins in almost all categories in terms of clustering accuracy.

Convergence analysis

It is a well-known fact that the convergence of inexact ALM with three or more blocks variables is unclear (Eckstein and Bertsekas 1992). Thus, it is difficult to prove the convergence of Algorithm 1 in theory because there are more than 3 block variables in Algorithm 1 and the objective function of (23) is not smooth. Fortunately, according to the theoretical results in (Eckstein and Bertsekas 1992), two conditions are sufficient (but may not necessary) for Algorithm 1 to converge: (1) each feature matrix $\mathbf{X}^{(v)}$ is of full column rank; (2) the optimality gap produced in each iteration step is monotonically decreasing. The first condition is easy to obey by factorizing $\mathbf{Z}^{(v)}$ into $\mathbf{H}^{(v)}\bar{\mathbf{Z}}^{(v)}$, where $\mathbf{H}^{(v)}$ can be computed in advance by orthogonalizing the columns of $\mathbf{X}^{(v)T}$ (Guangcan et al. 2013). For the second condition, the convexity of the Lagrangian function could guarantee its validity to some extent according to the work of Eckstein and Bertsekas (Eckstein and Bertsekas 1992). Therefore, the proposed WTNNM algorithm ensures good convergence properties. In the experiment, we show the convergence of reconstruction error $\|\mathbf{X}_{k+1}^{(v)} - \mathbf{X}_{k+1}^{(v)}\mathbf{Z}_{k+1}^{(v)} - \mathbf{E}_{k+1}^{(v)}\|_\infty$ and variable error $\|\mathbf{Z}_{k+1}^{(v)} - \mathbf{J}_{k+1}^{(v)}\|_\infty$ of our algorithm on the

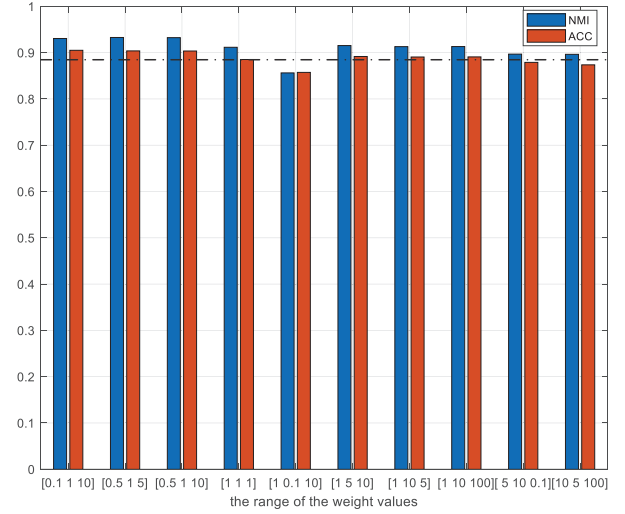


Figure 4: Performances vs. different weighted values on the Scene-15 database.

Caltech-101 dataset in Figure 3. It can be seen that our method has a good convergence and perform well in reality.

Weighted values analysis

Fig. 4 shows the results by using the different weighted values on the Scene-15 database. It can be seen that the weighted parameter ω has a large efficiency for clustering performance, which indicates the importance of prior knowledge of singular values. Our method overall has good performances (ACC and NMI) with $\omega = [0.5, 1, 5]$. This is probably because that each singular value of a matrix usually has different roles to the robustness of clustering algorithms in the existence of illumination and noise.

Conclusion

We have studied the weighted tensor nuclear norm minimization and propose an efficient iterative algorithm, which has good convergence. Our method can also be used to solve the matrix nuclear norm minimization, which have been widely used in machine learning and artificial intelligence. Moreover, applying our proposed weighted tensor nuclear norm to multi-view clustering, we develop an novel subspace clustering algorithm which well captures both the high order information and complementary information embedded in multi-view data. Extensive experimental results indicate that our method is superiority to state-of-the-art multi-view subspace clustering methods.

Acknowledgments

This work is supported by National Natural Science Foundation of China under Grant 61773302 and 61906142, Initiative Postdocs Supporting Program, China Postdoctoral Science Foundation (Grant 2019M653564) and the Fundamental Research Funds for the Central Universities.

References

- Asuncion, A., and Newman, D. 2007. Uci machine learning repository.
- Belhumeur, P. N.; Hespanha, J. P.; and Kriegman, D. J. 1997. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* 19(7):711–720.
- Cao, X.; Wei, X.; Han, Y.; and Lin, D. 2015a. Robust face clustering via tensor decomposition. *IEEE Trans. Cybernetics* 45(11):2546–2557.
- Cao, X.; Zhang, C.; Fu, H.; and Liu, S. 2015b. Diversity-induced multi-view subspace clustering. In *Proc. IEEE CVPR*, 586–594.
- Chen, K.; Dong, H.; and Chan, K.-S. 2013. Reduced rank regression via adaptive nuclear norm penalization. *Biometrika* 100(4):901–920.
- Eckstein, J., and Bertsekas, D. P. 1992. On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.* 55:293–318.
- Gao, H.; Nie, F.; Li, X.; and Huang, H. 2015. Multi-view subspace clustering. In *Proc. IEEE ICCV*, 4238–4246. IEEE.
- Guangcan, L.; Zhouchen, L.; Shuicheng, Y.; Ju, S.; Yong, Y.; and Yi, M. 2013. Robust recovery of subspace structures by low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 35(1):171–184.
- Hu, W.; Tao, D.; Zhang, W.; Xie, Y.; and Yang, Y. 2017. The twist tensor nuclear norm for video completion. *IEEE Trans. Neural Netw. Learning Syst.* 28(12):2961–2973.
- Kilmer, M. E., and Martin, C. D. 2011. Factorization strategies for third-order tensors. *Linear Algebra and Its Applications* 435(3):641–658.
- Kilmer, M. E.; Braman, K. S.; Hao, N.; and Hoover, R. C. 2013. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. *SIAM J. Matrix Analysis Applications* 34(1):148–172.
- Liu, W., and Tsang, I. W. 2017. Making decision trees feasible in ultrahigh feature and label dimensions. *Journal of Machine Learning Research* 18:81:1–81:36.
- Liu, W.; Tsang, I. W.; and Müller, K. 2017. An easy-to-hard learning paradigm for multiple classes and multiple labels. *Journal of Machine Learning Research* 18:94:1–94:38.
- Lu, C.; Feng, J.; Chen, Y.; Liu, W.; Lin, Z.; and Yan, S. 2016. Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization. In *Proc. IEEE CVPR*, 5249–5257.
- Luo, S.; Zhang, C.; Zhang, W.; and Cao, X. 2018. Consistent and specific multi-view subspace clustering. In *Proc. AAAI*, 3730–3737.
- Nie, F.; Cai, G.; and Li, X. 2017. Multi-view clustering and semi-supervised classification with adaptive neighbours. In *Proc. AAAI*, 2408–2414.
- Wang, Y.; Lin, X.; Wu, L.; Zhang, W.; Zhang, Q.; and Huang, X. 2015. Robust subspace clustering for multi-view data by exploiting correlation consensus. *IEEE Trans. Image Processing* 24(11):3939–49.
- Wang, Y.; Zhang, W.; Wu, L.; Lin, X.; Fang, M.; and Pan, S. 2016. Iterative views agreement: An iterative low-rank based structured optimization method to multi-view spectral clustering. In *Proc. IJCAI*, 2153–2159.
- Wang, X.; Guo, X.; Lei, Z.; Zhang, C.; and Li, S. Z. 2017. Exclusivity-consistency regularized multi-view subspace clustering. In *Proc. IEEE CVPR*, 1–9.
- Wang, Y.; Wu, L.; Lin, X.; and Gao, J. 2018. Multiview spectral clustering via structured low-rank matrix factorization. *IEEE Trans. Neural Netw. Learning Syst.* 29(10):4833–4843.
- Wang, H.; Nie, F.; and Huang, H. 2013. Multi-view clustering and feature learning via structured sparsity. In *Proc. ICML*.
- Wu, J.; Lin, Z.; and Zha, H. 2019. Essential tensor learning for multi-view spectral clustering. *IEEE Trans. Image Processing* 28(12):5910–5922.
- Xia, R.; Pan, Y.; Du, L.; and Yin, J. 2014. Robust multi-view spectral clustering via low-rank and sparse decomposition. In *Proc. AAAI*, 2149–2155.
- Yang, Y., and Wang, H. 2018. Multiview clustering: A survey. *Bid Data Mining and Analytics* 1(2).
- Yuan, X.; Tao, D.; Zhang, W.; Yan, L.; Lei, Z.; and Qu, Y. 2018. On unifying multi-view self-representations for clustering by tensor multi-rank minimization. *IJCV* 126(11):1157–1179.
- Zhang, Y.-F.; Xu, C.; Lu, H.; and Huang, Y.-M. 2009. Character identification in feature-length films using global face-name matching. *IEEE Trans. Multimedia* 11(7):1276–1288.
- Zhang, Z.; Ely, G.; Aeron, S.; Hao, N.; and Kilmer, M. 2014. Novel methods for multilinear data completion and de-noising based on tensor-svd. In *Proc. IEEE CVPR*, 3842–3849.
- Zhang, C.; Fu, H.; Liu, S.; Liu, G.; and Cao, X. 2015. Low-rank tensor constrained multiview subspace clustering. In *Proc. IEEE ICCV*, 1582–1590.