MULTI-VIEW LEARNING VIA LOW-RANK TENSOR OPTIMIZATION

Lele Fu[†]*, Zhaoliang Chen*[†]*, Sujia Huang*[‡]*, Sheng Huang*[†]*, Shiping Wang*[†*]

[†]College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350116, China
[‡]School of Computer Information Engineering, Jiangxi Normal University, Nanchang 330022, China
lawrencefzu@gmail.com, chenzl23@outlook.com, sujia@jxnu.edu.cn,
{hshs0594, shipingwangphd}@163.com

2021 IEEE International Conference on Multimedia and Expo (ICME) | 978-1-6654-3864-3/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICME51207.2021.9428291

ABSTRACT

In tensor-based multi-view learning methods, the self-representation based subspace clustering is widely researched, which is effective but heavy in high computational complexity. Furthermore, most of approaches learn the low-rank tensor representation and the final affinity matrix separately and ignore the difference between views. In this paper, we construct the target tensor composed of multiple normalized similarity matrices based on the Gaussian kernel function, which is constrained with the t-SVD based tensor nuclear norm to recover the low-rank part. The final affinity matrix is simultaneously learned via weighted multi-view fusion while optimizing the low-rank tensor, which suggests that each view is distributed to an adaptive weight. Moreover, the proposed method can be extended to semi-supervised classification through the collaborative optimization of the similarity tensor and the label indicator matrix. Extensive experiments conducted on four real-world datasets demonstrate the superiority of the proposed method compared with other state-of-the-art methods.

Index Terms— Multi-view clustering, semi-supervised classification, tensor nuclear norm, low-rank tensor

1. INTRODUCTION

Multi-view data have become a crucial data format with the rapid development of multi-media services and feature engineering techniques, owing to the fact that representations of multiple views depict data more comprehensively than single view. In light of this, a growing number of clustering and semi-supervised classification methods based on multi-view learning paradigms have been developed in the last decades.

Substantial researches have revealed the encouraging performance of multi-view clustering. For instance, Xia et al. [1] was devoted to optimizing a potential low-rank transition probability matrix as the input of spectral clustering by Markov chain. Luo et al. [2] regarded multi-view data as the combination of the consistent component and the specific components, which were jointly exploited for subspace representation learning. Chen et al. [3] explored the latent embedding space from multiple views, and then the subspace representation was learned. Zhang et al. [4] performed multi-view clustering by exploring the latent representation from multiple views. In a nutshell, all of these multi-view learning methods concentrated on discovering the latent representations from multi-view data. In practical applications, data are accompanied with a small number of labeled samples. Multi-view semi-supervised classification aims to utilize multi-view data to classify samples under the supervision of partially labeled data. Tao et al. [5] proposed a regression-based loss function for each view that was combined with adaptive weights. Nie et al. [6] considered the effect of noisy entries and learned the local structure of data for better classification. Wang et al. [7] proposed an auto-weighted manifold embedding method to tackle multi-view semi-supervised classification problems.

Different from multi-view methods mentioned above, low-rank tensor learning mines the complementary and discriminant information of multi-view data from tensor aspect rather than matrix aspect, which discovers the high-order correlations between views from a global perspective. Zhang et al. [8] imposed the sum of nuclear norms (SNN) on the subspace representation tensor to recover the low-rank property. Wang et al. [9] paid attention to preserving the data structure to handle the high dimensional data while optimizing the coefficient tensor. Sun et al. [10] presented a tensor logregularizer (TLR) to better approximate the tensor rank in terms of multi-view subspace clustering. Although tensor-based methods have achieved encouraging results, there are still some problems to be addressed. Firstly, most tensor-based methods use subspace representations to construct the target tensor, which makes the optimization process time-consuming. Then, the ultimate affinity matrix and the low-rank tensor representation are learned separately, thereby neglecting the dependence between them. Finally, potential effects of low-rank tensor learning for multi-view semi-supervised classification should be further considered.

In this paper, we first assemble the Gaussian kernel-based

∗Corresponding Author

978-1-6654-3864-3/21/$31.00 ©2021 IEEE

normalized similarity matrices into the target tensor, after which the t-SVD based tensor nuclear norm is adopted to explore the high-order correlations across views. This target tensor construction method not only guarantees encouraging performance but also improves the efficiency of the algorithm. The diversity of different views is considered and each view is allocated to an adaptive weight. Finally, the affinity matrix is obtained via weighted multi-view fusion, which is combined with low-rank tensor learning into a unified optimization process. Furthermore, the proposed method can be extended to semi-supervised classification tasks by the co-optimization of the similarity tensor and label indicator matrix. The main contributions are summarized as:

- A unified framework is proposed that can be applied to both multi-view clustering and semi-supervised classification.
- The low-rank tensor representation and the weighted multi-view fusion are integrated into a joint learning process, where the weights are learned adaptively.
- Comprehensive experimental results reveal that the proposed method outperforms other state-of-the-art methods in clustering and classification tasks.

2. NOTATIONS AND PRELIMINARIES

In this section, the meanings of some notations and operations related to tensor are described in detail. Specifically, we use $\boldsymbol{\mathcal{X}} = \{\mathbf{X}^{(v)}\}_{v=1}^{m}$ to denote the multi-view data, where m is the number of views and $\mathbf{X}^{(v)} \in \mathbb{R}^{n \times d^{(v)}}$. Bold lowercase letters, bold capital letters and bold calligraphy letters (e.g., $\mathbf{a}$, $\mathbf{A}$, $\boldsymbol{\mathcal{A}}$) represent vectors, matrices and tensors, respectively. Given a third-order tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\boldsymbol{\mathcal{A}}^{(i)}$ denotes the i-th frontal slice of $\boldsymbol{\mathcal{A}}$ and $\mathbf{A}_{(i)}$ is comprised of vectors of $\boldsymbol{\mathcal{A}}$ along with the i-th dimension. $\boldsymbol{\mathcal{A}}^T \in \mathbb{R}^{n_2 \times n_1 \times n_3}$ is the transpose of $\boldsymbol{\mathcal{A}}$. $\boldsymbol{\mathcal{A}}_f = fft(\boldsymbol{\mathcal{A}}, [\,], 3)$ denotes the result of $\boldsymbol{\mathcal{A}}$ along with the third dimension after fast Fourier transformation (FFT). $l_{2,1}$ norm of a tensor is defined as $\|\boldsymbol{\mathcal{A}}\|_{2,1} = \sum_{i,j} \|\boldsymbol{\mathcal{A}}(i,j,:)\|_2$. For a better interpretation of the t-SVD based tensor nuclear norm, some definitions about tensor are clarified in advance.

Definition 1 ***(t-product)****: Given tensor $\boldsymbol{\mathcal{M}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and tensor $\boldsymbol{\mathcal{N}} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$, tensor $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{n_1 \times n_4 \times n_3}$ is the t-product result of $\boldsymbol{\mathcal{M}}$ and $\boldsymbol{\mathcal{N}}$*

$$\boldsymbol{\mathcal{B}} = \boldsymbol{\mathcal{M}} * \boldsymbol{\mathcal{N}} = bvfold(bcirc(\boldsymbol{\mathcal{M}}) \cdot bvec(\boldsymbol{\mathcal{N}})), \quad (1)$$

where $bvfold$, $bcirc$ and $bvec$ denote block-based operators defined in [11].

Definition 2 ***(t-SVD)****: Tensor singular value decomposition of $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is defined as*

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{D}} * \boldsymbol{\mathcal{V}}^T, \quad (2)$$

where $\boldsymbol{\mathcal{U}} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ and $\boldsymbol{\mathcal{V}}^T \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal tensors , $\boldsymbol{\mathcal{D}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is an f-diagonal tensor.

Definition 3 ***(t-SVD based tensor nuclear norm)****: The t-SVD based tensor nuclear norm of $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is defined as the sum of diagonal values of $\boldsymbol{\mathcal{D}}_f$ decomposed from $\boldsymbol{\mathcal{A}}_f$, that is,*

$$\|\boldsymbol{\mathcal{A}}\|_* = \sum_{i=1}^{\min\{n_1,n_2\}} \sum_{j=1}^{n_3} |\boldsymbol{\mathcal{D}}_f(i,i,j)|. \quad (3)$$

3. THE PROPOSED METHOD

3.1. Problem formulation

For convenience, we first illustrate our model with a clustering task. In the model, the initial normalized similarity matrices with the Gaussian kernel function for each view are first calculated, that is, $\mathbf{Z}^{(v)} = \mathbf{D}^{(v)^{-\frac{1}{2}}} \mathbf{K}^{(v)} \mathbf{D}^{(v)^{-\frac{1}{2}}}$, and $\mathbf{D}^{(v)}$ is a diagonal matrix defined as $\mathbf{D}_{ii}^{(v)} = \sum_{j=1}^{n} \mathbf{K}_{ij}$. Then, $\{\mathbf{Z}^{(v)}\}_{v=1}^{m}$ are combined into tensor $\boldsymbol{\mathcal{Z}}$. Inspired by [12], we assume that tensor $\boldsymbol{\mathcal{Z}}$ is composed of a low-rank tensor $\boldsymbol{\mathcal{S}}$ and an error tensor $\boldsymbol{\mathcal{E}}$. To capture complementary information across views better, the dimensions of all tensors are rotated from $n \times n \times m$ to $n \times m \times n$. Furthermore, we obtain the final affinity matrix via weighted multi-view fusion, and the fusion process is integrated with the low-rank tensor learning into a unified procedure. In light of above points, the objective function of the proposed method is written as

$$\begin{aligned} \min_{\boldsymbol{\mathcal{S}},\boldsymbol{\mathcal{E}},\mathbf{A},w^{(v)}} & \|\boldsymbol{\mathcal{S}}\|_* + \lambda\|\boldsymbol{\mathcal{E}}\|_{2,1} + \alpha \sum_{v=1}^{m} (w^{(v)})^r (\|\mathbf{S}^{(v)} - \mathbf{A}\|_F^2) \\ \text{s.t. } & \boldsymbol{\mathcal{Z}} = \boldsymbol{\mathcal{S}} + \boldsymbol{\mathcal{E}}, \mathbf{w}^T\mathbf{1} = 1, w^{(v)} \geq 0, v = 1, \ldots, m, \end{aligned} \quad (4)$$

where $\|\cdot\|_*$ denotes tensor nuclear norm, $\|\cdot\|_{2,1}$ represents $l_{2,1}$ norm and handles the noises to enhance model stability. $w^{(v)}$ is the weight of the i-th view and $\mathbf{w}$ is the vector constituted of $\{w^{(v)}\}_{v=1}^{m}$. The solved affinity matrix $\mathbf{A}$ is regarded as the input of spectral clustering. $\lambda > 0$ and $\alpha > 0$ are two penalty parameters, respectively.

For classification tasks, the proposed objective function is still applicable without introducing any extra terms. When the normalized similarity matrix $\mathbf{Z}^{(v)}$ of each view is solved, we splice it horizontally with the label matrix $\mathbf{F} = [\mathbf{F}_1; \ldots; \mathbf{F}_n] \in \mathbb{R}^{n \times c}$ to form $[\mathbf{Z}^{(v)}, \mathbf{F}] \in \mathbb{R}^{n \times (n+c)}$. For each point x_i, if x_i belongs to the j-th cluster, we define $\mathbf{F}_{ij} = 1$ and $\mathbf{F}_{ij} = 0$ otherwise. Thus, we collect $\{[\mathbf{Z}^{(v)}, \mathbf{F}]\}_{v=1}^{m}$ into a tensor as the input of the model. The rotation of tensor is also performed, that is, the dimension is adjusted from $n \times (n+c) \times m$ to $n \times m \times (n+c)$. If data points belong to the same class in the low-rank tensor,

the corresponding rows in matrices $\{[\mathbf{Z}^{(v)}, \mathbf{F}]\}_{v=1}^{m}$ tend to be consistent, and the part of unknown label information in $\mathbf{F}$ will be completed. After low-rank optimization, the final matrix $[\mathbf{A}, \mathbf{F}]$ is acquired. For each sample x_i without label, its label y_i is computed by $y_i = argmax_{j \leq c} \mathbf{F}_{ij}$.

Algorithm 1 Multi-view learning based on low-rank tensor optimization (MLLTO)

Input: $\boldsymbol{\mathcal{X}} = \{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(m)}\}$, $\mathbf{X}^{(v)} \in \mathbb{R}^{n \times d^{(v)}}$, λ, α, and number of clusters c.

Output: Matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ for clustering and $\mathbf{A} \in \mathbb{R}^{n \times (n+c)}$ for classification.

Initialize $\boldsymbol{\mathcal{G}}_0 = \boldsymbol{\mathcal{S}}_0 = \boldsymbol{\mathcal{E}}_0 = \boldsymbol{\mathcal{Y}}_0 = \boldsymbol{\mathcal{H}}_0 = \mathbf{0}$, $w_0^{(v)} = \frac{1}{m}$, $\omega = 2$, $\varepsilon = 10^{-7}$, $\mu_0 = \beta_0 = 10^{-4}$, $\mu_{max} = \beta_{max} = 10^{10}$, $k = 0$.
while not convergent **do**
  Update $\boldsymbol{\mathcal{G}}_{k+1}$ by solving Eq. (7);
  for $v = 1 : m$ **do**
    Update $\mathbf{S}_{k+1}^{(v)}$ by solving Eq. (9);
  end for
  Update $\mathbf{E}_{(3)k+1}$ by solving Eq. (11);
  Update $w_{k+1}^{(v)}$ by solving Eq. (14);
  Update $\mathbf{A}_{k+1}$ by solving Eq. (16);
  Update $\boldsymbol{\mathcal{Y}}_{k+1}$, $\boldsymbol{\mathcal{H}}_{k+1}$, μ_{k+1} and β_{k+1} by solving Eq. (17);
  Check the convergence conditions: $||\boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{S}}_{k+1} - \boldsymbol{\mathcal{E}}_{k+1}||_\infty \leq \varepsilon$, $||\boldsymbol{\mathcal{G}}_{k+1} - \boldsymbol{\mathcal{S}}_{k+1}||_\infty \leq \varepsilon$, $||\mathbf{A}_{k+1} - \mathbf{A}_k||_\infty \leq \varepsilon$.
  $k = k + 1$;
end while
return Matrix $\mathbf{A}$.

3.2. Optimization

In order to solve Eq. (4) based on the alternating direction method of multipliers (ADMM), the augmented Lagrangian function of Eq. (4) is formulated as

$$\begin{aligned}
&\mathcal{F}\left(\boldsymbol{\mathcal{S}}; \boldsymbol{\mathcal{E}}; \boldsymbol{\mathcal{G}}; \mathbf{A}; \{w^{(v)}\}_{v=1}^{m}\right) \\
&= \|\boldsymbol{\mathcal{G}}\|_* + \lambda\|\boldsymbol{\mathcal{E}}\|_{2,1} + \alpha \sum_{v=1}^{m} (w^{(v)})^r \|\mathbf{S}^{(v)} - \mathbf{A}\|_F^2 + \\
&\langle \boldsymbol{\mathcal{Y}}, \boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{E}} \rangle + \frac{\mu}{2}\|\boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{E}}\|_F^2 + \langle \boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{S}} \rangle \\
&+ \frac{\beta}{2}\|\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{S}}\|_F^2,
\end{aligned} \tag{5}$$

where $\boldsymbol{\mathcal{G}}$ is an auxiliary variable and we have $\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{S}}$, $\boldsymbol{\mathcal{Y}}$ and $\boldsymbol{\mathcal{H}}$ are two Lagrange multipliers, μ and β are two penalty parameters. Next, the updating rules of varying variables are as follows.

Update $\boldsymbol{\mathcal{G}}$: When fixing $\boldsymbol{\mathcal{S}}$, $\boldsymbol{\mathcal{E}}$, $\mathbf{A}$ and $\{w^{(v)}\}_{v=1}^{m}$, the problem becomes

$$\begin{aligned}
&\min_{\boldsymbol{\mathcal{G}}} ||\boldsymbol{\mathcal{G}}||_* + \langle \boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{S}} \rangle + \frac{\beta}{2}\|\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{S}}\|_F^2 \\
&= \min_{\boldsymbol{\mathcal{G}}} ||\boldsymbol{\mathcal{G}}||_* + \frac{\beta}{2}||\boldsymbol{\mathcal{G}} - (\boldsymbol{\mathcal{S}} - \frac{1}{\beta}\boldsymbol{\mathcal{H}})||_F^2.
\end{aligned} \tag{6}$$

The closed-form solution of $\boldsymbol{\mathcal{G}}$ can be obtained by the tensor tubal-shrinkage operator $\mathcal{T}$ proposed in [13],

$$\boldsymbol{\mathcal{G}}^* = \mathcal{T}_{m/\beta}(\boldsymbol{\mathcal{S}} - \frac{1}{\beta}\boldsymbol{\mathcal{H}}). \tag{7}$$

Update $\boldsymbol{\mathcal{S}}$: Updating tensor $\boldsymbol{\mathcal{S}}$ is equivalent to updating each frontal slice of $\boldsymbol{\mathcal{S}}$

$$\begin{aligned}
&\min_{\mathbf{S}^{(v)}} \alpha (w^{(v)})^r \|\mathbf{S}^{(v)} - \mathbf{A}\|_F^2 + \langle \mathbf{Y}^{(v)}, \mathbf{Z}^{(v)} - \mathbf{S}^{(v)} - \mathbf{E}^{(v)} \rangle \\
&+ \frac{\mu}{2}\|\mathbf{Z}^{(v)} - \mathbf{S}^{(v)} - \mathbf{E}^{(v)}\|_F^2 + \langle \mathbf{H}^{(v)}, \mathbf{G}^{(v)} - \mathbf{S}^{(v)} \rangle \\
&+ \frac{\beta}{2}\|\mathbf{G}^{(v)} - \mathbf{S}^{(v)}\|_F^2.
\end{aligned} \tag{8}$$

Taking the partial derivative of $\mathbf{S}^{(v)}$ and setting the value to zero, we have

$$\mathbf{S}^{(v)*} = \frac{2\alpha(w^{(v)})^r \mathbf{A} + \mu(\mathbf{Z}^{(v)} - \mathbf{E}^{(v)}) + \mathbf{Y}^{(v)} + \beta\mathbf{G}^{(v)} + \mathbf{H}^{(v)}}{2\alpha(w^{(v)})^r + \mu + \beta}. \tag{9}$$

Update $\boldsymbol{\mathcal{E}}$: Fixing the other variables, updating $\boldsymbol{\mathcal{E}}$ is to solve the following problem

$$\begin{aligned}
&\min_{\boldsymbol{\mathcal{E}}} \lambda\|\boldsymbol{\mathcal{E}}\|_{2,1} + \langle \boldsymbol{\mathcal{Y}}, \boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{E}} \rangle + \frac{\mu}{2}\|\boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{E}}\|_F^2 \\
&= \min_{\mathbf{E}_{(3)}} \lambda\|\mathbf{E}_{(3)}\|_{2,1} + \frac{\mu}{2}\|\mathbf{E}_{(3)} - (\mathbf{Z}_{(3)} - \mathbf{S}_{(3)} + \mathbf{Y}_{(3)}/\mu)\|_F^2.
\end{aligned} \tag{10}$$

Letting $\mathbf{Z}_{(3)} - \mathbf{S}_{(3)} + \mathbf{Y}_{(3)}/\mu = \mathbf{P}$, the closed-form solution of $\mathbf{E}_{(3)}$ according to [12] is presented as follows

$$\mathbf{E}_{(3):,j}^* = \begin{cases} \frac{||\mathbf{P}_{:,j}||_2 - \frac{\lambda}{\mu}}{||\mathbf{P}_{:,j}||_2} \mathbf{P}_{:,j}, & \|\mathbf{P}_{:,j}\|_2 > \frac{\lambda}{\mu}; \\ \mathbf{0}, & \text{otherwise.} \end{cases} \tag{11}$$

Update $\{w^{(v)}\}_{v=1}^{m}$: When the terms irrelevant to $\{w^{(v)}\}_{v=1}^{m}$ are fixed, we have

$$\begin{aligned}
&\min_{w^{(v)}} \sum_{v=1}^{m} (w^{(v)})^r (\|\mathbf{S}^{(v)} - \mathbf{A}\|_F^2). \\
&\text{s.t. } \mathbf{w}^T\mathbf{1} = 1, w^{(v)} \geq 0.
\end{aligned} \tag{12}$$

Assuming $\mathbf{J}^{(v)} = \|\mathbf{S}^{(v)} - \mathbf{A}\|_F^2$, the Lagrangian function of Eq. (12) is formulated as

$$\mathcal{L} = \sum_{v=1}^{m} (w^{(v)})^r \mathbf{J}^{(v)} - \eta(\sum_{v=1}^{m} w^{(v)} - 1). \tag{13}$$

Table 1: Comparison of experimental results on all datasets for multi-view clustering.

Method	ALOI			COIL-20			HW			MSRC-v1		
	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
SPC_{best}	57.98	71.8	53.53	73.54	80.42	65.47	66.48	63.01	52.63	70.29	55.03	46.69
LRR_{best}	60.19	55.42	45.82	76.12	83.12	73.67	59.32	51.73	39.63	60.29	52.74	40.24
LTMSC	61.89	68.41	53.30	76.26	83.44	69.42	85.64	75.65	71.52	82.86	73.86	66.19
CSMSC	75.66	73.32	63.61	74.65	82.57	67.07	72.09	57.75	51.09	81.62	75.77	70.96
t-SVD-MSC	79.78	81.06	72.89	84.52	88.47	79.68	96.65	92.61	92.62	95.33	91.71	89.71
ETLMSC	72.23	77.48	65.06	87.72	94.78	86.29	85.89	86.37	79.79	88.38	86.23	80.61
MCLES	49.69	51.84	35.62	78.21	87.84	74.71	95.50	90.61	90.27	85.95	80.47	73.81
MLLTO	**93.70**	**89.87**	**86.79**	**95.63**	**96.75**	**93.52**	**99.90**	**99.73**	**99.78**	**96.00**	**92.04**	**90.91**

Setting the derivative of Eq. (13) with respect to $w^{(v)}$ to zero, the solution of $w^{(v)}$ is obtained by

$$w^{(v)^*} = \frac{(\mathbf{J}^{(v)})^{1/(1-r)}}{\sum_{v=1}^{m}(\mathbf{J}^{(v)})^{1/(1-r)}}. \tag{14}$$

Update A: All variables except $\mathbf{A}$ are fixed, we obtain the following problem

$$\min_{\mathbf{A}} \sum_{v=1}^{m}(w^{(v)})^r(\|\mathbf{S}^{(v)} - \mathbf{A}\|_F^2). \tag{15}$$

Setting the derivative of $\mathbf{A}$ in Eq. (15) to zero, we have

$$\mathbf{A}^* = \frac{\sum_{v=1}^{m}(w^{(v)})^r\mathbf{S}^{(v)}}{\sum_{v=1}^{m}(w^{(v)})^r}. \tag{16}$$

Update $\mathcal{Y}, \mathcal{H}, \mu$ and β: The Lagrange multipliers and penalty parameters are updated by

$$\begin{aligned} &\mathcal{Y}^* = \mathcal{Y} + \mu(\mathcal{Z} - \mathcal{S} - \mathcal{E}); \mathcal{H}^* = \mathcal{H} + \beta(\mathcal{G} - \mathcal{S}); \\ &\mu^* = min(\omega * \mu, \mu_{max}); \beta^* = min(\omega * \beta, \beta_{max}). \end{aligned} \tag{17}$$

The computational complexity of the proposed method mainly lies in the calculation of $\mathcal{G}$ and $\mathcal{E}$. Due to FFT, inverse FFT and t-SVD operations, updating $\mathcal{G}$ takes $\mathcal{O}(n^2mlog(n) + n^2m^2)$. As for $\mathcal{E}$, its computation process requires $\mathcal{O}(mn^2)$. In general, the computational complexity is $\mathcal{O}(k(n^2mlog(n) + n^2m^2))$ and k is the number of iterations. Algorithm 1 summarizes the main steps of the proposed method MLLTO.

4. EXPERIMENTS

In this section, we perform substantial experiments in terms of multi-view clustering and semi-supervised classification to verify the effectiveness of the proposed MLLTO. Numerous traditional and multi-view approaches are compared with four universally used multi-view datasets.

4.1. Experimental settings

In our experiments, four benchmark datasets for multi-view learning are selected to measure the performance of all compared methods. ALOI[1] contains 1,079 images that are divided into 10 classes, and each image has 4 views. COIL-20[2] is composed of 1,440 images with 20 categories, each sample is represented as 4 features. HW[3] contains 2,000 images of handwritten numerals from 0 to 9, each image with 6 types of representations. MSRC-v1[4] is constituted of a total of 210 images of 7 objects, each image with 5 feature representations. For clustering tasks, the proposed method is compared with following methods: SPC_{best} [14], LRR_{best} [15], LTMSC [8], CSMSC [2], t-SVD-MSC [16], ETLMSC [17], MCLES [3]. It is noted that SPC_{best} and LRR_{best} are single-view methods. We perform these methods on each view and record the best performance. As to semi-supervised classification tasks, MLLTO is compared with KNN, AMGL [18], MVAR [5], MLAN [19], HLR-MVS [20]. To measure the performance of all methods, we evaluate the results of clustering tasks with clustering accuracy (ACC), normalized mutual information (NMI) and adjusted rand index (ARI). Experimental results of semi-supervised classification tasks are evaluated by classification accuracy. All experiments are run ten times and we record the average values of experimental outcomes.

4.2. Performance comparison

MLLTO has two hyperparameters: λ and α. On ALOI, λ is set as 0.01 for clustering and 0.013 for classification. On COIL-20, λ is set as 0.05 and 0.058 for classification. On HW, λ is set as 0.008 for clustering and classification. On MSRC-v1, λ is tuned as 0.099 for clustering and classification. As for α, it is always set 1 regardless of the dataset and the task.

Table 1 exhibits the performance of multi-view clustering tasks. In light of this table, We have the following beneficial

[1]https://elki-project.github.io/datasets/multi view
[2]https://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php
[3]http://yann.lecun.com/exdb/mnist/
[4]http://riemenschneider.hayko.at/vision/dataset/task.php?did=35

Table 2: The performance of compared classification methods when the ratio of labeled samples is fixed as 10%.

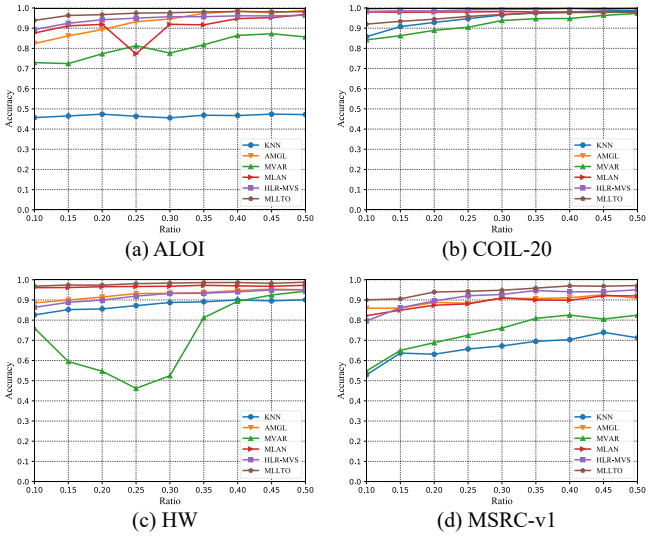
Method	ALOI	COIL-20	HW	MSRC-v1
KNN	45.68	85.79	82.58	52.91
AMGL	82.41	**99.52**	88.57	85.93
MVAR	72.94	84.2	76.06	54.81
MLAN	87.61	98.06	96.09	82.22
HLR-MVS	89.34	98.25	86.37	79.58
MLLTO	**93.85**	92.02	**96.84**	**90.05**

Table 3: Running time (seconds) of various multi-view clustering methods on all datasets.

Method	ALOI	COIL-20	HW	MSRC-v1
LTMSC	131.55	214.20	2035	8.62
CSMSC	90.50	161.44	482.03	16.06
t-SVD-MSC	55.87	123.58	194.56	3.02
ETLMSC	17.12	87.31	411.02	0.87
MCLES	12180	9791	72555	141.49
MLLTO	24.40	58.23	171.93	0.95

observations. Firstly, all of methods with regard to multi-view clustering commonly behave favorably in comparison with single-view methods SPC_{best} and LRR_{best}. It follows that these multi-view methods succeed in learning complementary information among multiple views and promoting the accuracy of clustering. In addition, the proposed MLLTO significantly outperforms these compared methods by all metrics, which suggests the effectiveness of the proposed algorithm. This leading situation is explainable. In brief, MLLTO utilizes the low-rank tensor representation to explore the high-order correlations compared with CSMSC and MCLES. Different from LTMSC, t-SVD-MSC and ETLMSC, MLLTO learns the low-rank tensor representation and the final affinity matrix simultaneously, which is beneficial to utilize the dependence between them. Moreover, MLLTO takes into account the diversity between different data features and learns the adaptive weights for all views. Finally, since the target tensor is constructed by the Gaussian kernel function instead of subspace representation, the efficiency of MLLTO is promising as revealed in Table 3.

Additionally, we conduct experiments on semi-supervised classification to further validate the proposed MLLTO. Figure 1 demonstrates the classification accuracy of all compared methods with different ratios of labeled data ranging in $\{0.10, 0.15, \cdots, 0.50\}$. It can be seen that MLLTO achieves desired experimental results that are comparable or superior to other compared methods, and behaves satisfactorily with limited labeled data. Besides, the performance of MLLTO is relatively stable as the supervision ratio increases. Table 2 presents the classification results when the ratio of labeled samples is fixed as 10%.



(a) ALOI (b) COIL-20 (c) HW (d) MSRC-v1

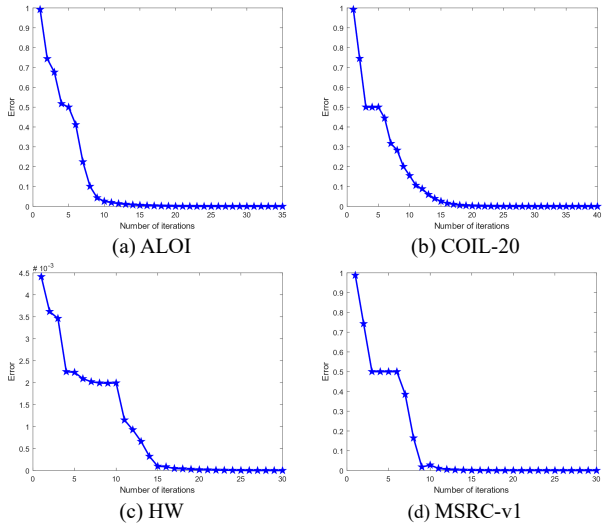
Fig. 1: The performance of the proposed method MLLTO with different ratios of labeled data ranging in $\{0.10, 0.15, \cdots, 0.50\}$.

4.3. Convergence Analyses

For the sake of presentation on the feasibility of the proposed MLLTO, we examine the convergence of the model with all tested datasets in Figure 2. The convergence of multi-view clustering and semi-supervised classification is analogous, thereby we only display the curves of loss values in clustering tasks. The error is defined as $error = max(||\boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{S}}_{k+1} - \boldsymbol{\mathcal{E}}_{k+1}||_\infty \leq \varepsilon, ||\boldsymbol{\mathcal{G}}_{k+1} - \boldsymbol{\mathcal{S}}_{k+1}||_\infty \leq \varepsilon, ||\mathbf{A}_{k+1} - \mathbf{A}_k||_\infty \leq \varepsilon)$. The experimental results indicate that loss values of MLLTO descent rapidly and converge within a certain number of iterations finally, which is suggestive of the effectiveness of the proposed algorithm.

5. CONCLUSION

In this paper, we propose a new framework termed as MLLTO that can be applied to multi-view clustering and semi-supervised classification at the same time. For the sake of algorithm efficiency, we construct the similarity tensor based on the Gaussian kernel function. The t-SVD based tensor nuclear norm is employed to minimize the rank of the target tensor. The difference between views is considered and an adaptive weight is learned for each view. Based on the adaptive weight learning, the final matrix is acquired via weighted multi-view fusion. Furthermore, we integrate the fusion process and the low-rank tensor learning into a joint procedure. Experimental results verify the effectiveness of the proposed MLLTO by clustering and classification accuracy.



(a) ALOI (b) COIL-20

(c) HW (d) MSRC-v1

Fig. 2: The convergent curves of the proposed method MLLTO with respect to clustering tasks on all tested datasets.

6. REFERENCES

[1] Rongkai Xia, Yan Pan, Lei Du, and Jian Yin, "Robust multi-view spectral clustering via low-rank and sparse decomposition," in *AAAI*, 2014, pp. 2149–2155.

[2] Shirui Luo, Changqing Zhang, Wei Zhang, and Xiaochun Cao, "Consistent and specific multi-view subspace clustering," in *IJCAI*, 2018, pp. 3730–3737.

[3] Mansheng Chen, Ling Huang, Changdong Wang, and Dong Huang, "Multi-view clustering in latent embedding space," in *AAAI*, 2020, pp. 3513–3520.

[4] Changqing Zhang, Huazhu Fu, Qinghua Hu, Xiaochun Cao, Yuan Xie, Dacheng Tao, and Dong Xu, "Generalized latent multi-view subspace clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 1, pp. 86–99, 2020.

[5] Hong Tao, Chenping Hou, Feiping Nie, Jubo Zhu, and Dongyun Yi, "Scalable multi-view semi-supervised classification via adaptive regression," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4283–4296, 2017.

[6] Feiping Nie, Guohao Cai, Jing Li, and Xuelong Li, "Auto-weighted multi-view learning for image clustering and semi-supervised classification," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1501–1511, 2018.

[7] Shiping Wang, Zhewen Wang, and Wenzhong Guo, "Accelerated manifold embedding for multi-view semi-supervised classification," *Information Sciences*, 2021.

[8] Changqing Zhang, Huazhu Fu, Si Liu, Guangcan Liu, and Xiaochun Cao, "Low-rank tensor constrained multiview subspace clustering," in *ICCV*, 2015, pp. 1582–1590.

[9] Zhan Wang, Lizhi Wang, and Hua Huang, "Structure preserving multi-view dimensionality reduction," in *ICME*, 2020, pp. 1–6.

[10] Xiaoli Sun, Youjuan Wang, and Xiujun Zhang, "Multi-view subspace clustering via non-convex tensor rank minimization," in *ICME*, 2020, pp. 1–6.

[11] Misha E Kilmer, Karen Braman, Ning Hao, and Randy C Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 1, pp. 148–172, 2013.

[12] 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*, vol. 35, no. 1, pp. 171–184, 2012.

[13] Wenrui Hu, Dacheng Tao, Wensheng Zhang, Yuan Xie, and Yehui Yang, "The twist tensor nuclear norm for video completion," *IEEE Tansactions on Neural Networks and Learning Systems*, vol. 28, no. 12, pp. 2961–2973, 2016.

[14] Andrew Y Ng, Michael I Jordan, and Yair Weiss, "On spectral clustering: Analysis and an algorithm," in *NIPS*, 2001, pp. 849–856.

[15] 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*, vol. 35, no. 1, pp. 171–184, 2013.

[16] Yuan Xie, Dacheng Tao, Wensheng Zhang, Yan Liu, Lei Zhang, and Yanyun Qu, "On unifying multi-view self-representations for clustering by tensor multi-rank minimization," *International Journal of Computer Vision*, vol. 126, no. 11, pp. 1157–1179, 2018.

[17] Jianlong Wu, Zhouchen Lin, and Hongbin Zha, "Essential tensor learning for multi-view spectral clustering," *IEEE Transactions on Image Processing*, vol. 28, no. 12, pp. 5910–5922, 2019.

[18] Feiping Nie, Jing Li, and Xuelong Li, "Parameter-free auto-weighted multiple graph learning: A framework for multiview clustering and semi-supervised classification," in *IJCAI*, 2016, pp. 1881–1887.

[19] Feiping Nie, Guohao Cai, and Xuelong Li, "Multi-view clustering and semi-supervised classification with adaptive neighbours," in *AAAI*, 2017, pp. 2408–2414.

[20] Yuan Xie, Wensheng Zhang, Yanyun Qu, Longquan Dai, and Dacheng Tao, "Hyper-laplacian regularized multilinear multiview self-representations for clustering and semisupervised learning," *IEEE Transactions on Cybernetics*, vol. 50, no. 2, pp. 572–586, 2020.