Enhanced Tensor Low-Rank and Sparse Representation Recovery for Incomplete Multi-View Clustering

Chao Zhang[1], Huaxiong Li[1]*, Wei Lv[1], Zizheng Huang[1], Yang Gao[2], Chunlin Chen[1]
[1]Department of Control Science and Intelligence Engineering, Nanjing University
[2]State Key Laboratory for Novel Software Technology, Nanjing University
{chaozhang, weilv, zizhenghuang}@smail.nju.edu.cn, {huaxiongli, gaoy, clchen}@nju.edu.cn

Abstract

Incomplete multi-view clustering (IMVC) has attracted remarkable attention due to the emergence of multi-view data with missing views in real applications. Recent methods attempt to recover the missing information to address the IMVC problem. However, they generally cannot fully explore the underlying properties and correlations of data similarities across views. This paper proposes a novel Enhanced Tensor Low-rank and Sparse Representation Recovery (ETLSRR) method, which reformulates the IMVC problem as a joint incomplete similarity graph learning and complete tensor representation recovery problem. Specifically, ETLSRR learns the intra-view similarity graphs and constructs a 3-way tensor by stacking the graphs to explore the inter-view correlations. To alleviate the negative influence of missing views and data noise, ETLSRR decomposes the tensor into two parts: a sparse tensor and an intrinsic tensor, which models the noise and underlying true data similarities, respectively. Both global low-rank and local structured sparse characteristics of the intrinsic tensor are considered, which enhances the discrimination of similarity matrix. Moreover, instead of using the convex tensor nuclear norm, ETLSRR introduces a generalized nonconvex tensor low-rank regularization to alleviate the biased approximation. Experiments on several datasets demonstrate the effectiveness and superiority of our method compared with the state-of-the-art methods.

Introduction

In the big data era, multi-view data become very common in real-world applications, and multi-view learning has attracted much attention in the machine learning and data mining communities (Han et al. 2022; Li et al. 2023; Xia et al. 2022b; Hassani and Khasahmadi 2020; Zhang et al. 2022d). As one of the most important tasks of multi-view learning, multi-view clustering (MVC) aims to partition the multi-view data into groups by exploiting the potential consistent and complementary information across views. Recently, various successful MVC methods have been developed (Zhang et al. 2017; Peng et al. 2019; Wang et al. 2019; Chen et al. 2020; Tang et al. 2020; Pan and Kang 2021; Xu et al. 2022b). These methods have achieved promising clustering performance in the case that the information of all views is available. However, in real scenarios, the collected multi-view data are usually incomplete with missing views. For example, some webpages contain texts, images, and videos, while some others only have texts. In industrial monitor, the multi-source data are often collected independently by a set of detectors, which can be incomplete due to detector failure. Traditional MVC methods cannot handle this problem, and increasing attention has been paid to incomplete multi-view clustering (IMVC) (Zhao, Liu, and Fu 2016; Liu et al. 2021; Wen et al. 2020; Lin et al. 2021; Zhang et al. 2022a).

Most existing IMVC methods can be categorized into non-imputation and imputation based approaches. Non-imputation based methods usually learn a common representation or subspace from existing data. Partial multi-view clustering (PVC) learns a common representation for aligned samples and a private representation for unaligned ones by non-negative matrix factorization (Li, Jiang, and Zhou 2014). Anchor based partial multi-view clustering (APMC) selects anchors from paired samples and captures the pairwise similarities based on anchors (Guo and Ye 2019). Consensus bipartite graph based IMVC (IMVC-CBG) adaptively learns latent anchors and bipartite graphs in a unified framework (Wang et al. 2022). Different from the above methods that neglect the underlying information of missing views, imputation based methods try to fill the missing information. Early works directly fill the missing views by zero values or average features, which will introduce large errors under high data missing rates. Some recent methods attempt to explore the multi-view correlations for filling the missing views (Wen et al. 2019; Zhang et al. 2022b). Zhang et al. recently proposed a Low-rank tensor regularized views recovery (LATER) method to fill the missing views for IMVC (Zhang et al. 2022b). It assumes a shared latent representation for all views to recovery the missing views, and the inter-view correlations are regularized by a low-rank tensor (Xie et al. 2018; Wu, Lin, and Zha 2019; Zhang et al. 2020) to enhance the recovery quality. Incomplete multi-view tensor spectral clustering with missing view inferring (IMVTSC-MVI) (Wen et al. 2021b) fills the missing views by error matrices, and learns view-specific self-expression graphs with low-rank tensor regularization.

More recently, instead of directly filling the original views, some researchers perform imputation on incomplete similarity graphs which essentially transfers the missing data

*Corresponding author.

problem from original feature space to graph space (Wen et al. 2021a; Xia et al. 2022a; Wen, Xu, and Liu 2020). For example, adaptive graph completion based method (AGC-IMC) performs low-rank matrix completion on all kNN similarity graphs of incomplete views, and learns a consensus representation for clustering (Wen et al. 2021a). By stacking the similarity graphs with missing values into a 3-order tensor, Xia et al. proposed a tensor completion based method (TCIMC), which recovers a complete low-rank tensor and learns a common representation under the guidance of recovered tensor (Xia et al. 2022a). Recovering the similarity graphs is interesting and natural to address the IMVC problem. Although the above graph completion methods have achieved promising performance, there are still some limitations. On the one hand, these methods generally conduct completion on pre-defined graphs (e.g., kNN graphs), which separates the graph learning and graph recovery into two steps, making the clustering performance sensitive to the initial graphs quality. On the other hand, these methods mainly focus on the global low-rank property, while the local structures in graphs are not fully exploited. The local structures characterize the intra-class and inter-class relationships among data, which is significant for effective clustering. Moreover, most methods use the convex nuclear norm for low-rank approximation, which is a biased estimation of rank function.

To this end, we propose a novel enhanced tensor low-rank and sparse representation recovery method, called ETLSRR. ETLSRR formulates the IMVC problem as a joint incomplete graph learning and tensor based complete graph recovery problem. The graph learning and tensor recovery are integrated into a unified framework and coordinates with each other to boost graph quality. The noisy tensor stacked by incomplete graphs is decomposed into a sparse noise tensor and an intrinsic similarity tensor with low-rank and sparse property. Moreover, a generalized tensor low-rank regularization is introduced to enhance the low-rank property. We summarize the main contributions as follows:

- ETLSRR provides a novel joint graph learning and tensor based graph recovery optimization framework to address the IMVC problem.
- To alleviate the negative influence of noise, ETLSRR adopts tensor decomposition to explicitly model the noisy values, and recover an intrinsic tensor with both global low-rank and local structure sparse regularization.
- ETLSRR introduces a generalized nonconvex approximation for tensor low-rank property, and considers the local structure with row fiber sparsity regularization, which enhances the discrimination of learned affinity graphs.

Notations and Preliminaries

We use bold lower case letters (e.g. $\mathbf{a}$) for vectors, upper case letters (e.g., $\mathbf{A}$) for matrices, and calligraphy letters (e.g., $\mathcal{A}$) for tensors. For a matrix $\mathbf{A} \in \mathbb{R}^{n_1 \times n_2}$, its Frobenius norm, nuclear norm, and $\ell_{2,1}$-norm are defined as $\|\mathbf{A}\|_F = \sqrt{\sum_{ij} a_{ij}^2}$, $\|\mathbf{A}\|_* = \sum_i \delta_i(\mathbf{A})$, and $\|\mathbf{A}\|_{2,1} = \sum_j \sqrt{\sum_i a_{ij}^2}$, respectively, where a_{ij} is the element of $\mathbf{A}$ at position (i,j), and $\delta_i(\mathbf{A})$ is the i-th singular value of $\mathbf{A}$. For a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, its ℓ_1 norm, Frobenius norm, and infinity norm are defined as $\|\mathcal{A}\|_1 = \sum_{ijk} |a_{ijk}|$, $\|\mathcal{A}\|_F = \sqrt{\sum_{ijk} a_{ijk}^2}$, and $\|\mathcal{A}\|_\infty = \max_{ijk} |a_{ijk}|$, respectively. $\mathcal{A}^{(:,:,i)}$, $\mathcal{A}^{(:,i,:)}$, $\mathcal{A}^{(i,:,:)}$ denote the i-th frontal, lateral, and horizantal slice, respectively. For convenience, we use $\mathbf{A}^{(i)}$ to represent $\mathcal{A}^{(:,:,i)}$. The inner product between $\mathcal{A}$ and $\mathcal{B}$ is $\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i=1}^{n_3} \langle \mathbf{A}^{(i)}, \mathbf{B}^{(i)} \rangle$. $\hat{\mathcal{A}}$ denotes the fast Fourier transformation (FFT) of $\mathcal{A}$ along the third dimension, i.e., $\hat{\mathcal{A}} = \text{fft}(\mathcal{A}, [], 3)$, and $\mathcal{A}$ can be recovered from $\hat{\mathcal{A}}$ by the inverse FFT operation, i.e., $\mathcal{A} = \text{ifft}(\hat{\mathcal{A}}, [], 3)$(Lu et al. 2020).

Definition 1 (t-SVD (Kilmer et al. 2013)**).** For a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, its t-SVD is defined as

$$\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T,$$

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 tensors, $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is an f-diagonal tensor, whose frontal slices are all diagonal matrices, and "$*$" denotes the t-product.

Based on t-SVD, the tensor multi-rank and tensor nuclear norm are defined as follows.

Definition 2 (tensor multi-rank (Zhou et al. 2018)**).** Given a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, its multi-rank $rank_m(\mathcal{A})$ is a vector defined as

$$rank_m(\mathcal{A}) = [rank(\hat{\mathbf{A}}^{(1)}); rank(\hat{\mathbf{A}}^{(2)}); ...; rank(\hat{\mathbf{A}}^{(n_3)})].$$

Definition 3 (tensor nuclear norm (Zhou et al. 2018)**).** Tensor nuclear norm is the tightest convex relaxation to the ℓ_1 norm of tensor multi-rank, which is defined as

$$\|\mathcal{A}\|_{\circledast} = \sum_{k=1}^{n_3} \|\hat{\mathbf{A}}^{(k)}\|_* = \sum_{i=1}^{\min(n_1,n_2)} \sum_{k=1}^{n_3} \delta_i(\hat{\mathbf{A}}^{(k)}),$$

where $\delta_i(\hat{\mathbf{A}}^{(k)})$ is the i-th singular value of $\hat{\mathbf{A}}^{(k)}$.

Let $\tilde{\mathbf{X}} = \{\tilde{\mathbf{X}}^v\}_{v=1}^m$ be an incomplete multi-view dataset with m views, and $\tilde{\mathbf{X}}^v \in \mathbf{R}^{d_v \times n_v}$ denotes the existing data matrix of the i-th view, where d_v is the feature dimension and n_v is the number of samples in view v. $\mathbf{X} = \{\mathbf{X}^v\}_{v=1}^m$ with $\mathbf{X}^v \in \mathbb{R}^{d_v \times n}$ represents the complete dataset by filling the missing values in $\tilde{\mathbf{X}}$, and n is the number of all samples ($n \geq n_v$). Note that we do not fill $\tilde{\mathbf{X}}$ in our method and just use $\mathbf{X}$ for ease of expression.

Method

Recovering the missing information is significant for IMVC. Considering that directly recovering the original missing features may be difficult due to the complex intra- and inter-class variations, we attempt to recover the latent similarity graphs for clustering, and reformulate the IMVC problem as a joint graph learning and tensor recovery problem. The overall framework of our ETLSRR is shown in Figure 1.

Formulation

Graph based methods are popular and have been developed in recent years, which seek the pairwise similarity among

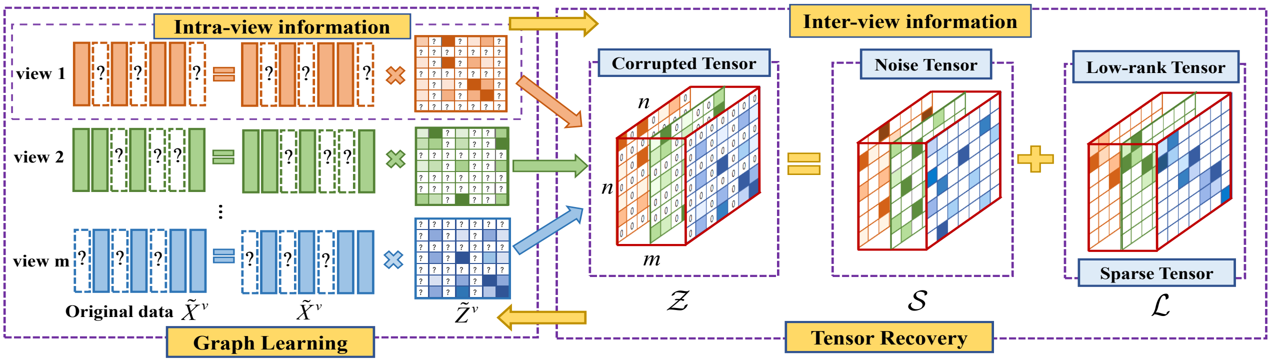


Figure 1: The overall framework of ETLSRR.

all data points for clustering. Low-rank representation based graph learning is widely adopted, in which a common graph with nuclear norm regularization is usually learned for all views (Zhang et al. 2017; Chen et al. 2020). However, for incomplete multi-view data, the common graph cannot be directly obtained due to the missing views. A straightforward strategy to deal with the incomplete data can be formulated as

$$\min_{\tilde{\mathbf{Z}}^v,\mathbf{E}^v} \sum_{v=1}^{m} \|\tilde{\mathbf{Z}}^v\|_* + \lambda\varphi(\mathbf{E}^v) \; s.t. \; \tilde{\mathbf{X}}^v = \tilde{\mathbf{X}}^v\tilde{\mathbf{Z}}^v + \mathbf{E}^v, \quad (1)$$

where $\tilde{\mathbf{Z}}^v \in \mathbb{R}^{n_v \times n_v}$ is the similarity graph of the v-th view, $\mathbf{E}^v \in \mathbb{R}^{d_v \times n_v}$ is the reconstruction error, φ is a function to estimate the data noise, and λ is a tradeoff parameter.

Due to the missing views of samples, each $\tilde{\mathbf{Z}}^v$ is incomplete in which the similarities related to missing samples are not encoded. For effective clustering, it is necessary to obtain a complete graph by leveraging and exploring all incomplete graphs. For its purpose, we first define an index matrix $\mathbf{H}^v \in \mathbb{R}^{n \times n_v}$ for the v-th view as follows:

$$h_{ij}^v = \begin{cases} 1, & \text{if the } i\text{-th sample in } \mathbf{X}^v \text{ is the} \\ & j\text{-th existing sample in } \tilde{\mathbf{X}}^v, \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

$\mathbf{H}^v$ indicates the positions of missing samples in view v. Then, a complete graph $\mathbf{Z}^v \in \mathbb{R}^{n \times n}$ can be constructed by

$$\mathbf{Z}^v = \mathbf{H}^v\tilde{\mathbf{Z}}^v\mathbf{H}^{vT}. \quad (3)$$

$\mathbf{Z}^v$ is equivalent to filling the missing values in $\tilde{\mathbf{Z}}^v$ by zero, which is corrupted and introduces noisy information. Directly using these corrupted graphs for clustering will lead to degraded performance. Thus, we expect to recover an intrinsic similarity matrix from these corrupted ones.

By stacking the view-specific graphs into a 3-order tensor, low-rank tensor based graph learning has been proved effective to explore the underlying correlations across views (Xie et al. 2018; Wu et al. 2020; Gao et al. 2020). Inspired by this, we first stack $\{\mathbf{Z}^v\}_{v=1}^m$ into a tensor $\mathcal{Z} \in \mathbb{R}^{n \times m \times n}$ (see Figure 1), and then reformulate the intrinsic graph learning problem as a tensor recovery problem. Specifically, the corrupted tensor $\mathcal{Z}$ can be decomposed into a low-rank tensor $\mathcal{L}$ and a sparse noise tensor $\mathcal{S}$, i.e., $\mathcal{Z} = \mathcal{L} + \mathcal{S}$. The overall model can be described as follows:

$$\begin{aligned} &\min_{\Omega} \sum_{v=1}^{m} \|\mathbf{E}^v\|_{2,1} + \lambda\|\mathcal{L}\|_{\circledast} + \theta\|\mathcal{S}\|_1 \\ &s.t. \; \tilde{\mathbf{X}}^v = \tilde{\mathbf{X}}^v\tilde{\mathbf{Z}}^v + \mathbf{E}^v, \mathcal{Z} = \mathcal{L} + \mathcal{S}, \\ &\quad \mathcal{Z} = \Phi(\mathbf{H}^1\tilde{\mathbf{Z}}^1\mathbf{H}^{1T}, ..., \mathbf{H}^m\tilde{\mathbf{Z}}^m\mathbf{H}^{mT}), \end{aligned} \quad (4)$$

where Φ is an operator to construct a 3-order tensor by collecting all graphs, and $\Omega = \{\mathbf{E}^v, \tilde{\mathbf{Z}}^v, \mathcal{Z}, \mathcal{L}, \mathcal{S}\}$ is the target variables set. $\ell_{2,1}$ norm is used to characterize the reconstruction errors, the intrisic tensor $\mathcal{L}$ is regularized by the popular tensor nuclear norm which encodes the true relationships among all samples including existing and missing ones in each view, and the sparse noise tensor $\mathcal{S}$ characterizes the noisy values caused by missing values imputation and original graph learning. According to definition 2, the convex nuclear norm in Eq. (4) directly sums all nuclear norms, which treats each singular value equally and ignores their different roles, leading to a biased approximation (Gao et al. 2020). Besides, Eq. (4) mainly focuses on the global low-rank property, but neglects the local structures in graphs. As shown in Figure 1, due to the local intra-class compactness, the frontal slices of recovered tensor $\mathcal{L}$ describes the sample-level similarities across views, and they should be structured sparse, i.e., row-wise sparse.

To enhance the low-rank property of recovered tensor, we extend the tensor nuclear norm by introducing a general surrogate to approximate the tensor rank as follows:

$$\|\mathcal{L}\|_{\gamma,\circledast} = \sum_{i=1}^{\min(n,m)} \sum_{k=1}^{n} \psi_\gamma(\delta_i(\hat{\mathbf{L}}^{(k)})), \quad (5)$$

where $\psi_\gamma(\cdot)$ is a nonconvex function with adjustment parameter $\gamma \in \mathbb{R}^+$. Obviously, when $\psi_\gamma(\cdot)$ is convex function $\psi_\gamma(x) = x$, Eq. (5) becomes the tensor nuclear norm. Inspired by (Lu et al. 2016), we define

$$\psi_\gamma(\delta_i(\hat{\mathbf{L}}^{(k)})) = 1 - exp(\frac{-\delta_i(\hat{\mathbf{L}}^{(k)})}{\gamma}), \quad (6)$$

which is the Laplace function, and it has been widely used for matrix low-rank nonconvex surrogates.

To capture the structured sparse property, we define $\|\mathcal{L}\|_{\mathrm{F1F}} = \sum_{ik} \|\mathcal{L}^{(i,:,k)}\|_2$, where $\mathcal{L}^{(i,:,k)}$ is a mode-2 (row) fiber. Minimizing $\|\mathcal{L}\|_{\mathrm{F1F}}$ encourages the sparsity at row fibers level. By intergrating the nonconvex low-rank and sparse regularization terms into Eq. (4), the objective function of ETLSRR is described as

$$\begin{aligned}
&\min_{\Omega} \sum_{v=1}^{m} \|\mathbf{E}^v\|_{2,1} + \lambda\|\mathcal{L}\|_{\gamma,\circledast} + \mu\|\mathcal{L}\|_{\mathrm{F1F}} + \theta\|\mathcal{S}\|_1 \\
&s.t.\ \tilde{\mathbf{X}}^v = \tilde{\mathbf{X}}^v\tilde{\mathbf{Z}}^v + \mathbf{E}^v, \mathcal{Z} = \mathcal{L} + \mathcal{S}, \\
&\qquad \mathcal{Z} = \Phi(\mathbf{H}^1\tilde{\mathbf{Z}}^1\mathbf{H}^{1^T}, ..., \mathbf{H}^m\tilde{\mathbf{Z}}^m\mathbf{H}^{m^T}).
\end{aligned} \tag{7}$$

As can be observed, ETLSRR jointly learns graphs and recovers an intrinsic tensor from a corrupted one with low-rank and sparse regularization. Once the optimal tensor $\mathcal{L}$ is acquired, we can compute the final complete similarity matrix $\mathbf{Q} \in \mathbb{R}^{n\times n}$ by

$$\mathbf{Q} = \frac{1}{m}\sum_{v=1}^{m} \frac{|\mathbf{L}^v| + |\mathbf{L}^v|^T}{2}, \tag{8}$$

where $\mathbf{L}^v$ is the v-th graph in $\mathcal{L}$ (i.e., the v-th lateral slice). Then, we can perform spectral clustering on $\mathbf{Q}$ to obtain clusters (Zhang et al. 2017).

Optimization

The ETLSRR model can be solved by the alternation direction method of multipliers algorithm (Boyd et al. 2011; Zhang et al. 2022c). We first introduce two tensor variables $\mathcal{B}$ and $\mathcal{C}$, and rewrite the original problem as follows:

$$\begin{aligned}
&\min_{\Omega} \sum_{v=1}^{m} \|\mathbf{E}^v\|_{2,1} + \lambda\|\mathcal{B}\|_{\gamma,\circledast} + \mu\|\mathcal{C}\|_{\mathrm{F1F}} + \theta\|\mathcal{S}\|_1 \\
&s.t.\ \mathbf{X}^v = \tilde{\mathbf{X}}^v\tilde{\mathbf{Z}}^v + \mathbf{E}^v, \mathcal{Z} = \mathcal{L} + \mathcal{S}, \mathcal{L} = \mathcal{B}, \mathcal{L} = \mathcal{C}, \\
&\qquad \mathcal{Z} = \Phi(\mathbf{H}^1\tilde{\mathbf{Z}}^1\mathbf{H}^{1^T}, ..., \mathbf{H}^m\tilde{\mathbf{Z}}^m\mathbf{H}^{m^T}).
\end{aligned} \tag{9}$$

The augmented Lagrangian function of problem (9) is:

$$\begin{aligned}
L_\rho &= \sum_{v=1}^{m} \|\mathbf{E}^v\|_{2,1} + \lambda\|\mathcal{B}\|_{\gamma,\circledast} + \theta\|\mathcal{S}\|_1 + \mu\|\mathcal{C}\|_{\mathrm{F1F}} \\
&+ \langle \tilde{\mathbf{X}}^v - \tilde{\mathbf{X}}^v\tilde{\mathbf{Z}}^v - \mathbf{E}^v, \mathbf{Y}^v\rangle + \langle \mathcal{Z} - \mathcal{L} - \mathcal{S}, \mathcal{F}_1\rangle \\
&+ \langle \mathcal{L} - \mathcal{B}, \mathcal{F}_2\rangle + \langle \mathcal{L} - \mathcal{C}, \mathcal{F}_3\rangle + \frac{\rho}{2}\|\mathcal{L} - \mathcal{B}\|_F^2 + \frac{\rho}{2}\|\mathcal{L} - \mathcal{C}\|_F^2 \\
&+ \frac{\rho}{2}\|\tilde{\mathbf{X}}^v - \tilde{\mathbf{X}}^v\tilde{\mathbf{Z}}^v - \mathbf{E}^v\|_F^2 + \frac{\rho}{2}\|\mathcal{Z} - \mathcal{L} - \mathcal{S}\|_F^2, \\
&s.t.\ \mathcal{Z} = \Phi(\mathbf{H}^1\tilde{\mathbf{Z}}^1\mathbf{H}^{1^T}, ..., \mathbf{H}^m\tilde{\mathbf{Z}}^m\mathbf{H}^{m^T}).
\end{aligned}$$

where $\rho > 0$ is a penalty factor, $\mathbf{Y}^v$, $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$ are Lagrange multipliers. Then, all variables can be updated by iteratively solving the corresponding sub-problems. Specifically, in the $(t+1)$-th iteration, it solves the following problems in sequence.

Step 1: By fixing other variables, $\tilde{\mathbf{Z}}^v$ can be updated by

$$\tilde{\mathbf{Z}}_{t+1}^v = \arg\min_{\tilde{\mathbf{Z}}^v} \|\mathbf{Q}_t^v - \tilde{\mathbf{X}}^v\tilde{\mathbf{Z}}^v\|_F^2 + \|\mathbf{H}^v\tilde{\mathbf{Z}}^v\mathbf{H}^{v^T} - \mathbf{P}_t^v\|_F^2, \tag{10}$$

where $\mathbf{Q}_t^v = \tilde{\mathbf{X}}^v - \mathbf{E}_t^v + \mathbf{Y}_t^v/\rho_t$ and $\mathbf{P}^v = \mathbf{L}_t^v + \mathbf{S}_t^v - \mathbf{F}_1^v/\rho_t$. By setting the derivative over $\tilde{\mathbf{Z}}^v$ to zero, we can obtain

$$(\mathbf{H}^{v^T}\mathbf{H}^v)^{-1}\tilde{\mathbf{X}}^{v^T}\tilde{\mathbf{X}}^v\tilde{\mathbf{Z}}^v + \tilde{\mathbf{Z}}^v\mathbf{H}^{v^T}\mathbf{H}^v = \mathbf{M}^v, \tag{11}$$

where $\mathbf{M}^v = (\mathbf{H}^{v^T}\mathbf{H}^v)^{-1}(\tilde{\mathbf{X}}^{v^T}\mathbf{Q}_t^v + \mathbf{H}^{v^T}\mathbf{P}_t^v\mathbf{H}^v)$. It is a typical Sylvester equation and can be solved by the Bartels-Stewart algorithm (Zhang et al. 2017).

Step 2: $\mathbf{E}^v$ can be updated by

$$\mathbf{E}_{t+1}^v = \arg\min_{\mathbf{E}^v} \|\mathbf{E}^v\|_{2,1} + \frac{\rho_t}{2}\|\mathbf{E}^v - \mathbf{J}_{t+1}^v\|_F^2, \tag{12}$$

where $\mathbf{J}_{t+1}^v = \tilde{\mathbf{X}}^v - \tilde{\mathbf{X}}^v\tilde{\mathbf{Z}}_{t+1}^v + \mathbf{Y}_t^v/\rho_t$. The optimal solution can be obtained by Lemma 4.1 in (Liu et al. 2013).

Step 3: $\mathcal{L}$ can be updated by

$$\begin{aligned}
\mathcal{L}_{t+1} = \arg\min_{\mathcal{L}} &\|\mathcal{Z}_{t+1} - \mathcal{L} - \mathcal{S}_t + \frac{\mathcal{F}_{1,t}}{\rho_t}\|_F^2 + \\
&\|\mathcal{L} - \mathcal{B}_t + \frac{\mathcal{F}_{2,t}}{\rho_t}\|_F^2 + \|\mathcal{L} - \mathcal{C}_t + \frac{\mathcal{F}_{3,t}}{\rho_t}\|_F^2.
\end{aligned}$$

The optimal solution is

$$\mathcal{L}_{t+1} = (\mathcal{K} + \frac{\mathcal{F}_{1,t} - \mathcal{F}_{2,t} - \mathcal{F}_{3,t}}{\rho_t})/3, \tag{13}$$

where $\mathcal{K} = \mathcal{Z}_{t+1} - \mathcal{S}_t + \mathcal{B}_t + \mathcal{C}_t$.

Step 4: $\mathcal{B}$ can be updated by

$$\mathcal{B}_{t+1} = \arg\min_{\mathcal{B}} \frac{\lambda}{\rho_t}\|\mathcal{B}\|_{\gamma,\circledast} + \frac{1}{2}\|\mathcal{B}_t - \mathcal{D}_{t+1}\|_F^2, \tag{14}$$

where $\mathcal{D}_{t+1} = \mathcal{L}_{t+1} + \mathcal{F}_{2,t}/\rho_t$. It is difficult to directly yield the closed-form solution due to the nonconvexity of the generalized tensor rank approximation. Based on Eq. (5), we transform problem (14) into n separated sub-problems in the frequency domain, and the k-th problem is

$$\mathbf{B}_{t+1}^{(k)} = \arg\min_{\hat{\mathbf{B}}^{(k)}} \frac{\lambda}{\rho_t}\sum_{i=1}^{m} \psi_\gamma(\delta_i(\hat{\mathbf{B}}^{(k)})) + \frac{1}{2}\|\hat{\mathbf{B}}^{(k)} - \hat{\mathbf{D}}_{t+1}^{(k)}\|_F^2. \tag{15}$$

Due to the antimonotone property of gradient of the nonconvex function in Eq. (5) and the descending order of singular values, it holds that $0 \le \nabla\psi_\gamma(\delta_1^t) \le \nabla\psi_\gamma(\delta_2^t) \le ... \le \nabla\psi_\gamma(\delta_m^t)$. Then, we can obtain

$$\psi_\gamma(\delta_i(\hat{\mathbf{B}}^{(k)})) \le \psi_\gamma(\delta_i^t) + \nabla\psi_\gamma(\delta_i^t)(\delta_i(\hat{\mathbf{B}}^{(k)}) - \delta_i^t),$$

where δ_i^t denotes the i-th singular values of $\hat{\mathbf{B}}_t^{(k)}$, and $\nabla\psi_\gamma(\delta_i^t)$ is the gradient of $\psi_\gamma(\hat{\mathbf{B}}^{(k)})$ at δ_i^t. By using the inequality, we relax problem (15) as follows:

$$\begin{aligned}
\mathbf{B}_{t+1}^{(k)} &= \arg\min_{\hat{\mathbf{L}}^{(k)}} \frac{\lambda}{\rho_t}\sum_{i=1}^{m} \psi_\gamma(\delta_i^t) + \nabla\psi_\gamma(\delta_i^t)(\delta_i(\hat{\mathbf{B}}^{(k)}) - \delta_i^t) \\
&\quad + \frac{1}{2}\|\hat{\mathbf{B}}^{(k)} - \hat{\mathbf{D}}_{t+1}^{(k)}\|_F^2 \\
&= \arg\min_{\hat{\mathbf{B}}^{(k)}} \frac{\lambda}{\rho_t}\sum_{i=1}^{V} \nabla\psi_\gamma(\delta_i^t)\delta_i(\hat{\mathbf{B}}^{(k)}) + \frac{1}{2}\|\hat{\mathbf{B}}^{(k)} - \hat{\mathbf{D}}_{t+1}^{(k)}\|_F^2,
\end{aligned}$$

which can be viewed as a weighted nuclear norm regularized problem. We solve it by the generalized weighted singular value thresholding operator (Lu et al. 2015), i.e.,

$$\mathbf{B}_{t+1}^{(k)} = \mathbf{U}\mathcal{H}_{\frac{\lambda\nabla\psi_\gamma}{\rho_t}}(\mathbf{\Sigma})\mathbf{V}^T, \tag{16}$$

where $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ is the SVD of $\hat{\mathbf{D}}_{t+1}^{(k)}$, and $\mathcal{H}_{\frac{\lambda\nabla\psi_\gamma}{\rho_t}}(\mathbf{\Sigma}) = \mathrm{diag}\{\max(0, \mathbf{\Sigma}_{i,i} - \lambda\nabla\psi_\gamma(\delta_i^t)/\rho_t)\}$.

Algorithm 1: ETLSRR algorithm

Input: Data matrices $\{\tilde{\mathbf{X}}^v\}_{v=1}^m$, parameters $\lambda, \mu, \theta, \gamma$.

Initialize $\mathbf{E}^{(v)} = \mathbf{0}$, $\mathbf{Z}^{(v)} = \mathbf{0}$, $\mathbf{Y}^{(v)} = \mathbf{0}$, $\mathcal{L} = \mathcal{S} = 0$, $\mathcal{F}_1 = \mathcal{F}_2 = \mathcal{F}_3 = 0$, $\rho = 10^{-3}$, $\delta = 1.3$, $\mu_{\max} = 10^6$, $\epsilon = 10^{-5}$.
while not converged **do**
    Update $\tilde{\mathbf{Z}}^v$ by solving (11);
    Update $\mathbf{E}^v$ by solving (12);
    Construct $\mathcal{Z} = \Phi(\mathbf{H}^1\tilde{\mathbf{Z}}^1\mathbf{H}^{1^T}, ..., \mathbf{H}^m\tilde{\mathbf{Z}}^m\mathbf{H}^{m^T})$;
    Update $\mathcal{L}$ by solving (13);
    Update $\mathcal{B}$ by solving (16);
    Update $\mathcal{C}$ by Eq. (18);
    Update $\mathcal{S}$ by solving (19);
    Update the Lagrange multipliers and penalty factors by Eqs. (20);
end while

Output: Compute the final similarity graph $\mathbf{Q}$ by Eq. (8).

Step 5: $\mathcal{C}$ can be updated by

$$\mathcal{C}_{t+1} = \arg\min_{\mathcal{C}} \frac{\mu}{\rho_t}\|\mathcal{C}\|_{\text{F1F}} + \frac{1}{2}\|\mathcal{C} - \mathcal{G}_{t+1}\|_F^2, \tag{17}$$

where $\mathcal{G}_{t+1} = \mathcal{L}_{t+1} + \mathcal{F}_{3,t}/\rho_t$. The optimal solution $\mathcal{C}_{t+1}$ can be obtained by operating on each row fiber, i.e.,

$$\mathcal{C}_{t+1}^{(i,:,j)} = \begin{cases} \frac{\left\|\mathcal{G}_{t+1}^{(i,:,j)}\right\|_2 - \frac{\mu}{\rho_t}}{\left\|\mathcal{G}_{t+1}^{(i,:,j)}\right\|_2}\mathcal{G}_{t+1}^{(i,:,j)}, & \text{if } \left\|\mathcal{G}_{t+1}^{(i,:,j)}\right\|_2 > \frac{\mu}{\rho_t}, \\ \mathbf{0}, & \text{otherwise.} \end{cases} \tag{18}$$

Step 6: $\mathcal{S}$ can be updated by

$$\mathcal{S}_{t+1} = \arg\min_{\mathcal{S}} \frac{\theta}{\rho_t}\|\mathcal{S}\|_1 + \frac{1}{2}\|\mathcal{S} - \mathcal{Q}_{t+1}\|_F^2, \tag{19}$$

where $\mathcal{Q}_{t+1} = \mathcal{Z}_{t+1} - \mathcal{L}_{t+1} + \frac{1}{\rho_t}\mathcal{F}_{1,t}$. It can be solved by the well-known soft-threshold operator (Yang and Zhang 2011).

Step 7: The Lagrange multipliers and penalty factor can be updated by

$$\begin{aligned} &\mathbf{Y}_{t+1}^{(v)} = \mathbf{Y}_t^{(v)} + \rho_t(\tilde{\mathbf{X}}^{(v)} - \tilde{\mathbf{X}}^{(v)}\tilde{\mathbf{Z}}_{t+1}^{(v)} - \mathbf{E}^{(v)}), \\ &\mathcal{F}_{1,t+1} = \mathcal{F}_{1,t} + \rho_t\left(\mathcal{Z}_{t+1} - \mathcal{L}_{t+1} - \mathcal{S}_{t+1}\right), \\ &\mathcal{F}_{2,t+1} = \mathcal{F}_{2,t} + \rho_t\left(\mathcal{L}_{t+1} - \mathcal{B}_{t+1}\right), \\ &\mathcal{F}_{3,t+1} = \mathcal{F}_{3,t} + \rho_t\left(\mathcal{L}_{t+1} - \mathcal{C}_{t+1}\right), \\ &\rho_{t+1} = \min\left(\delta\rho_t, \rho_{\max}\right), \end{aligned} \tag{20}$$

where δ and $\rho_{\max}$ are positive scalars. The convergence conditions are defined as

$$\max\left\{\begin{array}{l} \left\|\tilde{\mathbf{X}}^v - \tilde{\mathbf{X}}^v\tilde{\mathbf{Z}}_{t+1}^v - \mathbf{E}_{t+1}^v\right\|_\infty \\ \left\|\mathcal{Z}_{t+1} - \mathcal{L}_{t+1} - \mathcal{S}_{t+1}\right\|_\infty \\ \left\|\mathcal{L}_{t+1} - \mathcal{B}_{t+1}\right\|_\infty \\ \left\|\mathcal{L}_{t+1} - \mathcal{C}_{t+1}\right\|_\infty \end{array}\right\} \le \epsilon, \tag{21}$$

where ϵ is a small tolerance. Algorithm 1 summarizes the optimization procedure of ETLSRR.

Computational Complexity Analysis

In $\tilde{\mathbf{Z}}^v$ step, the major time cost is the matrix inverse operation and Sylvester equation. Since index matrix $\mathbf{H}^v$ is fixed and the inverse of $\mathbf{H}^{v^T}\mathbf{H}^v$ can be computed outside the loop. The complexity for solving Sylvester equation is $O(n^3)$. In $\mathcal{L}$ step, the major computational costs are the FFT and inverse FFT and SVD operations. The computational complexity of FFT and inverse FFT is $O(mn^2log(n) + m^2n^2)$ (Wen et al. 2021b), and that of SVD is $O(m^2n^2)$. Since the other steps only have some basic matrix operations which have lower complexity than the above calculations, we ignore their computational costs. Thus, the total computational complexity is about $O(\tau(n^3 + mn^2log(n) + m^2n^2))$ if there are total τ iterations.

Experiment

Experimental Settings

Datasets Five popular datasets are adopted, including: (1) **ORL** contains 400 face images from 40 individuals. Three types of features (intensity, LBP and Gabor) in 4096, 3304, 6750 dimensions, respectively, are extracted as three views. (2) **BBCSport** consists of 544 documents collected from the BBC news website. These documents belong to five topic labels and contain two views with 3183 and 3203 dimensions, respectively. (3) **RGB-D** dataset contains 1449 indoor scenes images and sentential descriptions for each image. The visual and textual views are in 2048 and 300 dimensions, respectively. (4) **UCI** contains 2000 handwritten digit images. Each image has two views with 240 and 76 dimensions, respectively. (5) **Scene** (Xie et al. 2018) has 15 natural scene categories with both indoor and outdoor environments. There are total 4485 images and three views in 1800, 1180, and 1240 dimensions, respectively.

Baselines We adopt the following methods as baselines for comparison: (1) UEAF (Wen et al. 2019), (2) AGC-IMC (Wen et al. 2021a), (3) IMVTSC-MVI (Wen et al. 2021b), (4) TMBSD (Li et al. 2021), (5) TCIMC (Xia et al. 2022a), (6) IMVC-CBG (Wang et al. 2022), and (7) DIMVC (Xu et al. 2022a), in which DIMVC is a deep IMVC method. For our method, the parameters (λ, μ, θ) are set to $(10, 1, 1)$ on ORL, BBCSport, and RGB-D; (10, 0.1, 1) on UCI and Scene. The bandwidth factor γ in Eq. (6) is 20.

Incomplete Data Construction Following (Wang et al. 2022; Xu et al. 2022b), we construct the incomplete data with different missing rates p (0.1, 0.3, 0.5, 0.7). For example, when the missing rate $p = 0.7$, we randomly select 30% samples as complete data and randomly drop partial views of the rest 70% samples.

Evaluation Metrics The clustering accuracy (ACC), normalized mutual information (NMI), and adjusted rand index (ARI) are used as evaluation metrics. For all metrics, the higher values indicate the better performance. We run all experiments 10 times to report the mean values.

	p	0.1			0.3			0.5			0.7		
	Metric	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
ORL	UEAF	63.52	78.45	47.03	50.32	70.34	31.91	45.85	62.65	12.09	42.50	58.16	8.23
	AGC-IMC	74.35	86.59	64.22	74.10	84.87	62.08	62.05	76.86	47.02	56.90	73.06	40.61
	IMVTSC-MVI	84.13	94.40	80.26	81.67	90.86	74.26	80.83	90.95	73.64	78.67	89.44	70.54
	TMBSD	96.00	98.02	95.07	94.60	97.97	92.97	92.54	96.83	89.98	91.30	96.06	87.93
	TCIMC	78.70	89.20	71.09	77.05	87.43	65.99	75.05	85.12	63.00	73.80	84.85	62.28
	IMVC-CBG	74.02	88.42	64.43	73.00	88.50	63.88	72.68	88.42	64.01	72.15	88.04	63.58
	DIMVC	63.66	82.80	46.15	55.28	76.25	37.20	42.16	63.63	27.73	36.31	55.84	17.15
	Ours	**96.55**	**98.95**	**96.31**	**96.40**	**98.90**	**96.26**	**95.62**	**98.40**	**94.99**	**95.19**	**98.24**	**94.50**
BBCSport	UEAF	84.38	68.67	66.15	84.45	65.03	65.92	81.49	63.43	64.69	77.76	58.06	54.17
	AGC-IMC	86.76	78.56	77.13	84.93	73.22	74.94	85.66	72.61	74.26	83.46	68.28	69.53
	IMVTSC-MVI	96.27	94.46	94.45	96.08	94.40	94.23	94.92	92.37	93.32	91.17	87.29	89.54
	TMBSD	94.04	85.77	87.23	92.66	84.55	82.52	91.54	84.00	80.12	88.05	83.64	73.98
	TCIMC	85.08	75.63	72.03	82.50	71.00	69.55	80.15	70.18	64.00	78.13	66.22	64.29
	IMVC-CBG	88.60	73.14	74.57	88.10	72.35	74.17	88.11	72.23	72.21	86.03	69.18	68.18
	DIMVC	87.89	79.02	75.94	78.64	70.37	66.69	74.83	53.75	51.61	68.11	42.45	38.87
	Ours	**99.60**	**98.56**	**99.16**	**99.14**	**97.66**	**98.47**	**99.12**	**96.99**	**97.92**	**97.90**	**93.88**	**94.91**
RGB-D	UEAF	44.29	37.47	25.44	38.16	32.21	18.23	35.97	29.86	13.38	33.18	26.80	9.09
	AGC-IMC	43.93	33.08	23.82	40.86	28.57	19.37	35.40	25.93	15.45	31.82	23.53	13.22
	IMVTSC-MVI	44.28	51.06	28.12	40.60	32.16	19.37	34.74	27.43	13.58	32.62	23.14	12.67
	TCIMC	49.36	42.30	32.68	41.66	35.84	20.64	38.62	33.20	16.07	34.64	30.55	10.57
	TMBSD	42.84	34.61	21.81	37.68	27.57	17.56	32.65	23.76	14.67	32.74	21.43	13.12
	IMVC-CBG	47.73	34.84	26.38	46.72	34.15	25.66	45.69	33.96	24.56	44.81	34.05	24.33
	DIMVC	41.66	37.56	23.77	38.18	30.46	19.17	31.96	22.64	14.51	28.98	17.86	10.98
	Ours	**60.80**	**60.71**	**47.36**	**60.56**	**58.95**	**46.72**	**60.32**	**57.04**	**45.48**	**58.18**	**56.47**	**43.99**
UCI	UEAF	87.40	78.20	74.49	68.62	62.35	50.55	56.85	51.17	32.18	43.92	41.98	17.19
	AGC-IMC	82.70	85.81	78.44	82.85	84.94	78.10	81.85	83.57	76.74	80.20	81.23	74.23
	IMVTSC-MVI	99.60	99.01	99.11	99.35	98.32	98.56	98.90	97.05	97.56	97.80	94.63	95.18
	TMBSD	99.15	98.00	98.13	98.98	97.60	97.77	98.05	95.85	95.72	96.11	94.41	92.24
	TCIMC	81.25	82.62	75.30	73.37	78.42	69.06	68.90	74.99	60.44	48.88	45.85	22.65
	IMVC-CBG	79.54	72.62	65.57	79.19	72.38	65.43	76.20	71.29	63.21	71.49	66.68	56.90
	DIMVC	84.65	78.67	69.82	78.87	73.84	62.59	62.95	62.94	62.49	57.43	56.23	44.00
	Ours	**99.80**	**99.45**	**99.56**	**99.68**	**99.11**	**99.38**	**99.65**	**99.06**	**99.22**	**99.42**	**98.46**	**98.73**
Scene	UEAF	40.15	45.07	26.06	36.07	37.49	19.98	29.96	31.21	15.12	22.99	23.89	7.97
	AGC-IMC	51.12	55.85	36.37	53.71	55.08	38.17	45.86	46.85	29.46	46.13	44.64	28.70
	IMVTSC-MVI	76.43	77.09	67.87	66.64	64.43	52.55	59.67	56.38	41.80	52.71	49.47	31.96
	TMBSD	59.90	56.93	59.59	52.35	50.79	51.17	43.06	40.88	42.59	31.57	29.61	30.82
	TCIMC	54.83	58.88	40.63	52.12	52.41	36.85	49.53	52.69	34.06	46.37	48.14	33.25
	IMVC-CBG	51.22	48.85	33.26	51.39	48.37	32.94	50.40	48.72	32.81	50.38	49.47	33.54
	DIMVC	57.10	52.92	31.10	46.17	41.75	32.98	34.95	32.16	23.62	25.14	21.04	13.39
	Ours	**94.39**	**92.78**	**92.69**	**94.28**	**92.61**	**92.45**	**93.97**	**92.08**	**91.83**	**93.81**	**91.88**	**91.60**

Table 1: Clustering results (%) of different methods on five datasets.

Experimental Results

Clustering Performance Table 1 lists the clustering results of different methods on five datasets with varying missing rates w.r.t. three metrics. It can be observed that: (1) Our ETLSRR achieves the best performance in all cases. Especially on RGB-D and Scene datasets, the improvements of our method are significant. This demonstrates that our method is superior to these methods on underlying information excavation and intrinsic similarity learning of incomplete multi-view data. (2) With the increase of missing rate, the clustering performance of all methods generally decreases. However, ETLSRR shows some robustness against the missing views and the performance drop is small when the missing rate increases from 0.3 to 0.7. (3) ETLSRR outperforms two graph completion based methods, AGC-IMC and TCIMC, demonstrating the effectiveness and superiority of joint graph learning and recovery with tensor low-rank and structured sparse regularization for IMVC.

Parameter Analysis In ETLSRR model, there are four parameters λ, μ, γ and θ that influence the clustering performance. To analyze the parameter sensitivity, Figure 2 shows the change of ACC value of ETLSRR on ORL with different combinations of parameters. We can observe that the proposed method is more sensitive to λ and γ to some extent compared with μ and θ. However, our method can obtain satisfactory results with relatively wide ranges for these parameters. For example, when μ and λ are selected from $[1, 10]$, θ is selected from $[0.5, 2]$, and γ is selected from $[5, 50]$, ETLSRR can obtain good performance.

Ablation Study To evaluate the effectiveness of nonconvex tensor low-rank approximation and row fiber sparsity in ETLSRR, we derive two variants, i.e., ETLSRR-E and

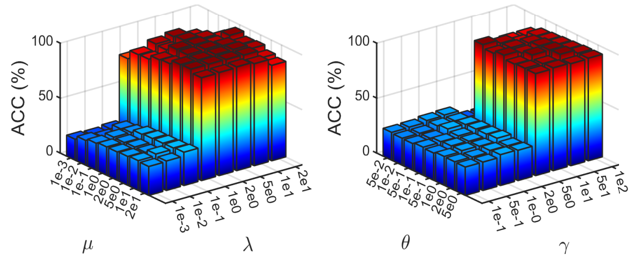
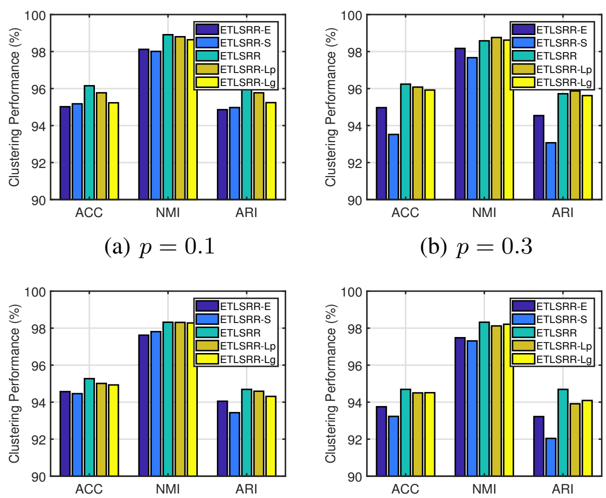


Figure 2: ACC values of ETLSRR under different parameter settings on ORL with missing rate $p = 0.7$.



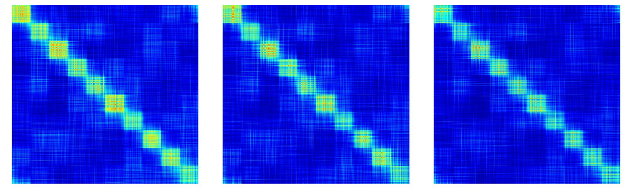
(a) $p = 0.1$ (b) $p = 0.3$

(c) $p = 0.5$ (d) $p = 0.7$

Figure 3: Ablation study results on ORL with varying missing rates.

STLSRR-S. ETLSRR-E replaces the nonconvex tensor low-rank regularization by tensor nuclear norm, and ETLSRR-S drops the tensor sparse regularization term (i.e., the third term in Eq. (7)). Besides, we adopt two another nonconvex functions, ℓ_p and logarithm function (Lu et al. 2016), to replace the laplace function in Eq. (6). The ℓ_p and logarithm functions are defined as $\psi_{\ell_p}(x) = x^p$ $(0 < p < 1)$ and $\psi_{lg}(x) = \log(1 + \gamma x)/\log(1 + \gamma)$. We denote the ETLSRR with the two nonconvex functions as ETLSRR-Lp and ETLSRR-Lg, respectively. Figure 3 shows the clustering results of ETLSRR and the four variants on ORL with varying missing rates. It can be observed that: (1) ETLSRR improves the performance over ETLSRR-E and ETLSRR-S, indicating the effectiveness of nonconvex approximation and structured sparse regularization. (2) Three nonconvex low-rank based methods (ETLSRR, ETLSRR-Lp, ETLSRR-Lg) outperform ETLSRR-E and achieve very competitive results, which implies that ETLSRR may be further improved by adopting appropriate nonconvex relaxations.

Visualization ETLSRR recovers an intrinsic tensor $\mathcal{L}$ from the corrupted tensor $\mathcal{Z}$ for clustering. To show the effectiveness of tensor based graph recovery of ETLSRR,



(a) $p = 0.3$ (b) $p = 0.5$ (c) $p = 0.7$

Figure 4: Visualization of recovered graphs of ETLSRR on UCI with varying missing rates.

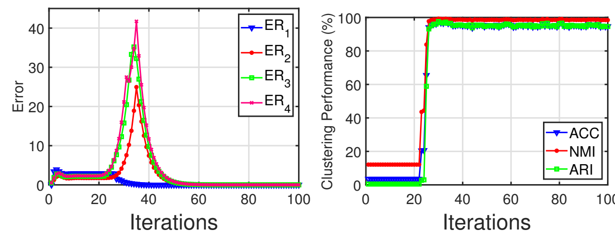


Figure 5: Convergence curves and clustering results w.r.t iterations on ORL with missing rate $p = 0.7$.

Figure 4 shows the learned affinity graph of ETLSRR on UCI dataset, where the number of diagonal blocks is equal to the ground-truth cluster number. It can be observed that all graphs have a clear block diagonal structure, even in the case of large missing rate. These results demonstrate the effectiveness of our method for intrinsic similarity discovery in incomplete multi-view data.

Convergence Figure 5 shows the error values of variables and clustering results in each iteration on ORL with missing rate $p = 0.7$, where the errors of variables are $\mathrm{ER}_1 = \max\{\left\|\tilde{\mathbf{X}}^v - \tilde{\mathbf{X}}^v\tilde{\mathbf{Z}}^v_{t+1} - \mathbf{E}^v\right\|_\infty\}_{v=1}^m$, $\mathrm{ER}_2 = \|\mathcal{Z} - \mathcal{L} - \mathcal{S}\|_\infty$, $\mathrm{ER}_3 = \|\mathcal{L} - \mathcal{B}\|_\infty$, $\mathrm{ER}_4 = \|\mathcal{L} - \mathcal{C}\|_\infty$. We can observe that the errors eventually converge to 0 and our method obtains stable clustering results within a few iterations, demonstrating the convergence property of the optimization algorithm.

Conclusion

This paper proposes a new incomplete multi-view clustering method, called Enhanced Tensor Low-rank and Sparse Representation Recovery (ETLSRR). ETLSRR jointly learns the incomplete similarity graphs and recovers an intrinsic tensor that characterizes the underlying data similarities. To eliminate the impacts of missing views and data noise, ETLSRR decomposes the initial tensor stacked by incomplete graphs into a sparse noise tensor and an intrinsic tensor. The global tensor low-rank property ensured by a generalized nonconvex function and local sparsity of the recovered tensor are captured. Experimental results demonstrate the superiority of our method compared with the state-of-the-art methods. The main limitation of our method is the relatively high computational complexity, and learning bipartite graph is expected to improve the algorithm efficiency.

Acknowledgments

This work is partially supported by the National Natural Science Foundation of China under Grants 62176116, 62073160, 62276136, and the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under Grant 20KJA520006.

References

Boyd, S. P.; Parikh, N.; Chu, E.; Peleato, B.; and Eckstein, J. 2011. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Found. Trends Mach. Learn.*, 3(1): 1–122.

Chen, M.; Huang, L.; Wang, C.; and Huang, D. 2020. Multi-view clustering in latent embedding space. In *AAAI*, 3513–3520.

Gao, Q.; Xia, W.; Wan, Z.; Xie, D.; and Zhang, P. 2020. Tensor-SVD Based Graph Learning for Multi-View Subspace Clustering. In *AAAI*, 3930–3937.

Guo, J.; and Ye, J. 2019. Anchors Bring Ease: An Embarrassingly Simple Approach to Partial Multi-View Clustering. In *AAAI*, 118–125.

Han, Z.; Zhang, C.; Fu, H.; and Zhou, J. T. 2022. Trusted Multi-View Classification with Dynamic Evidential Fusion. *IEEE TPAMI*.

Hassani, K.; and Khasahmadi, A. H. 2020. Contrastive multi-view representation learning on graphs. In *ICML*, 4116–4126.

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 Anal. Appl.*, 34(1): 148–172.

Li, H.; Zhang, C.; Jia, X.; Gao, Y.; and Chen, C. 2023. Adaptive Label Correlation Based Asymmetric Discrete Hashing for Cross-modal Retrieval. *IEEE TKDE*, 35(2): 1185–1199.

Li, S.; Jiang, Y.; and Zhou, Z. 2014. Partial Multi-View Clustering. In *AAAI*, 1968–1974.

Li, Z.; Tang, C.; Liu, X.; Zheng, X.; Zhang, W.; and Zhu, E. 2021. Tensor-Based Multi-View Block-Diagonal Structure Diffusion for Clustering Incomplete Multi-View Data. In *ICME*, 1–6.

Lin, Y.; Gou, Y.; Liu, Z.; Li, B.; Lv, J.; and Peng, X. 2021. COMPLETER: Incomplete Multi-View Clustering via Contrastive Prediction. In *CVPR*, 11174–11183.

Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; and Ma, Y. 2013. Robust Recovery of Subspace Structures by Low-Rank Representation. *IEEE TPAMI*, 35(1): 171–184.

Liu, X.; Li, M.; Tang, C.; Xia, J.; Xiong, J.; Liu, L.; Kloft, M.; and Zhu, E. 2021. Efficient and Effective Regularized Incomplete Multi-View Clustering. *IEEE TPAMI*, 43(8): 2634–2646.

Lu, C.; Feng, J.; Chen, Y.; Liu, W.; Lin, Z.; and Yan, S. 2020. Tensor Robust Principal Component Analysis with a New Tensor Nuclear Norm. *IEEE TPAMI*, 42(4): 925–938.

Lu, C.; Tang, J.; Yan, S.; and Lin, Z. 2016. Nonconvex Nonsmooth Low Rank Minimization via Iteratively Reweighted Nuclear Norm. *IEEE TIP*, 25(2): 829–839.

Lu, C.; Zhu, C.; Xu, C.; Yan, S.; and Lin, Z. 2015. Generalized Singular Value Thresholding. In *AAAI*, 1805–1811.

Pan, E.; and Kang, Z. 2021. Multi-view Contrastive Graph Clustering. In *NeurIPS*, 2148–2159.

Peng, X.; Huang, Z.; Lv, J.; Zhu, H.; and Zhou, J. T. 2019. COMIC: Multi-view clustering without parameter selection. In *ICML*, volume 97, 5092–5101.

Tang, C.; Liu, X.; Zhu, X.; Zhu, E.; Luo, Z.; Wang, L.; and Gao, W. 2020. CGD: Multi-View Clustering via Cross-View Graph Diffusion. In *AAAI*, 5924–5931.

Wang, S.; Liu, X.; Liu, L.; Tu, W.; Zhu, X.; Liu, J.; Zhou, S.; and Zhu, E. 2022. Highly-efficient incomplete large-scale multi-view clustering with consensus bipartite graph. In *CVPR*, 9776–9785.

Wang, S.; Liu, X.; Zhu, E.; Tang, C.; Liu, J.; Hu, J.; Xia, J.; and Yin, J. 2019. Multi-view clustering via late fusion alignment maximization. In *IJCAI*, 3778–3784.

Wen, J.; Xu, Y.; and Liu, H. 2020. Incomplete Multiview Spectral Clustering With Adaptive Graph Learning. *IEEE TCYB*, 50(4): 1418–1429.

Wen, J.; Yan, K.; Zhang, Z.; Xu, Y.; Wang, J.; Fei, L.; and Zhang, B. 2021a. Adaptive Graph Completion Based Incomplete Multi-View Clustering. *IEEE TMM*, 23: 2493–2504.

Wen, J.; Zhang, Z.; Xu, Y.; Zhang, B.; Fei, L.; and Liu, H. 2019. Unified Embedding Alignment with Missing Views Inferring for Incomplete Multi-View Clustering. In *AAAI*, 5393–5400.

Wen, J.; Zhang, Z.; Zhang, Z.; Wu, Z.; Fei, L.; Xu, Y.; and Zhang, B. 2020. DIMC-net: Deep Incomplete Multi-view Clustering Network. In *ACM MM*, 3753–3761.

Wen, J.; Zhang, Z.; Zhang, Z.; Zhu, L.; Fei, L.; Zhang, B.; and Xu, Y. 2021b. Unified Tensor Framework for Incomplete Multi-view Clustering and Missing-view Inferring. In *AAAI*, 10273–10281.

Wu, J.; Lin, Z.; and Zha, H. 2019. Essential Tensor Learning for Multi-View Spectral Clustering. *IEEE TIP*, 28(12): 5910–5922.

Wu, J.; Xie, X.; Nie, L.; Lin, Z.; and Zha, H. 2020. Unified Graph and Low-Rank Tensor Learning for Multi-View Clustering. In *AAAI*, 6388–6395.

Xia, W.; Gao, Q.; Wang, Q.; and Gao, X. 2022a. Tensor Completion-Based Incomplete Multiview Clustering. *IEEE TCYB*.

Xia, W.; Gao, Q.; Wang, Q.; Gao, X.; Ding, C.; and Tao, D. 2022b. Tensorized Bipartite Graph Learning for Multi-View Clustering. *IEEE TPAMI*.

Xie, Y.; Tao, D.; Zhang, W.; Liu, Y.; Zhang, L.; and Qu, Y. 2018. On Unifying Multi-view Self-Representations for Clustering by Tensor Multi-rank Minimization. *IJCV*, 126(11): 1157–1179.

Xu, J.; Li, C.; Ren, Y.; Peng, L.; Mo, Y.; Shi, X.; and Zhu, X. 2022a. Deep Incomplete Multi-view Clustering via Mining Cluster Complementarity. In *AAAI*, 8761–8769.

Xu, J.; Tang, H.; Ren, Y.; Peng, L.; Zhu, X.; and He, L. 2022b. Multi-Level Feature Learning for Contrastive Multi-View Clustering. In *CVPR*, 16051–16060.

Yang, J.; and Zhang, Y. 2011. Alternating Direction Algorithms for $_1$-Problems in Compressive Sensing. *SIAM J. Sci. Comput.*, 33(1): 250–278.

Zhang, C.; Cui, Y.; Han, Z.; Zhou, J. T.; Fu, H.; and Hu, Q. 2022a. Deep Partial Multi-View Learning. *IEEE TPAMI*, 44(5): 2402–2415.

Zhang, C.; Fu, H.; Wang, J.; Li, W.; Cao, X.; and Hu, Q. 2020. Tensorized Multi-view Subspace Representation Learning. *IJCV*, 128(8): 2344–2361.

Zhang, C.; Hu, Q.; Fu, H.; Zhu, P.; and Cao, X. 2017. Latent multi-view subspace clustering. In *CVPR*, 4333–4341.

Zhang, C.; Li, H.; Chen, C.; Jia, X.; and Chen, C. 2022b. Low-Rank Tensor Regularized Views Recovery for Incomplete Multiview Clustering. *IEEE TNNLS*.

Zhang, C.; Li, H.; Chen, C.; Qian, Y.; and Zhou, X. 2022c. Enhanced Group Sparse Regularized Nonconvex Regression for Face Recognition. *IEEE TPAMI*, 44(5): 2438–2452.

Zhang, C.; Li, H.; Gao, Y.; and Chen, C. 2022d. Weakly-Supervised Enhanced Semantic-Aware Hashing for Cross-Modal Retrieval. *IEEE TKDE*.

Zhao, H.; Liu, H.; and Fu, Y. 2016. Incomplete Multi-Modal Visual Data Grouping. In *IJCAI*, 2392–2398.

Zhou, P.; Lu, C.; Lin, Z.; and Zhang, C. 2018. Tensor Factorization for Low-Rank Tensor Completion. *IEEE TIP*, 27(3): 1152–1163.