

Research Article

Incomplete Multiview Clustering via Low-Rank Tensor Ring Completion

Jinshi Yu,¹ Haonan Huang,² Qi Duan,³ Yafei Wang ,¹ and Tao Zou ,^{1,4}

¹School of Mechanical and Electrical Engineering, Guangzhou University, Guangzhou 510006, China

²School of Automation, Guangdong University of Technology, Guangzhou 510006, China

³Guangzhou Panyu Polytechnic, Guangzhou 510006, China

⁴Pazhou Lab, Guangzhou 510330, China

Correspondence should be addressed to Yafei Wang; yw@gzhu.edu.cn and Tao Zou; tzou@gzhu.edu.cn

Received 2 December 2022; Revised 28 March 2023; Accepted 7 April 2023; Published 20 May 2023

Academic Editor: Vasudevan Rajamohan

Copyright © 2023 Jinshi Yu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Since real-world multiview data frequently contains numerous samples that are not observed from some viewpoints, the incomplete multiview clustering (IMC) issue has received a great deal of attention recently. However, most existing IMC methods choose to zero-fill the missing instances, which leads to the failure to exploit information hidden in the missing instances, and high-order interactions between various views. To tackle these problems, we proposed an effective IMC method using low-rank tensor ring completion, which was demonstrated to be powerful in exploiting high-order correlation. Specifically, we first stack the incomplete similarity graphs of all views into a 3rd-order incomplete tensor and then restore it via the tensor ring decomposition. Next, using an adaptive weighting technique, we apply multiview spectral clustering to all entire graphs in order to balance the contributions of different viewpoints and identify the consensus representation for grouping. Finally, we employ the alternating direction method of multipliers (ADMM) to optimize the suggested model. Numerous experimental findings on numerous different datasets show that the suggested approach is superior to other cutting-edge approaches.

1. Introduction

Since technology has advanced, real-world data frequently originates from a variety of sources. In other words, an object can be described from several views. For instance, using the image and text, a product can be explained; the disease can be diagnosed from blood tests, urine tests, and magnetic resonance imaging (MRI). These data are called multiview data [1–5]. Since multiview data typically offers compatible and complementary information, it is more comprehensive for object description in comparison with the single-view data [6–8]. Since this benefit has garnered a lot of attention in recent years, numerous efficient multiview clustering techniques have been put forth to enhance clustering performance by integrating the information presented in different views. For example, the canonical correlation analysis (CCA)-based method [9] is proposed to learn the common low-dimension subspace from multiview

data space, and subsequently implemented spectral clustering on the common representation. In reference [10], an effective multiview clustering method with con-regularization is proposed to exploit the consistent clustering structure across views, where the co-regularization is applied to make the clustering of different views be agreed with each other. Cai et al. [11] proposed a robust multiview k -means clustering (RMKMC) method for handling large-scale multiview clustering problems. Besides, some other multiview data clustering techniques are also developed in many references [12–17].

However, the design of the majority of prior multiview clustering studies frequently relies on the unrealistic assumption that each example could be completely seen across all views. In fact, many samples cannot be viewed in some views due to several unavoidable limitations in the collecting of multiview data. Hence, each view may have different available instances, and the multiview clustering in this

condition is called the IMC problem [18, 19]. It, therefore, brings a challenge to the conventional clustering methods. One intuitive way is to use the mean of the available instances of the corresponding view to fill in the missing instances. However, this strategy has a higher likelihood to damage the underlying structure inside each view, worsening performance, especially when the missing rate is large. In recent years, numerous attempts have been made to effectively tackle the IMC problem. For instance, Rai et al. [18] first developed the kernel-based clustering method to address the incomplete multiview problem. The method then explored the KCCA to recover the kernel matrix of the missing view (kernel canonical correlation analysis). Using nonnegative matrix factorization to take advantage of a latent subspace for effective clustering, Li et al. [20] suggested the partial multiview clustering (PVC) approach to deal with the multiview data with partial views. In [21], Zhao et al. suggested using a novel graph Laplacian term to address the IMC problem while maintaining the compact global structure. Xu et al. [22] presented a new clustering approach called partial multiview subspace clustering (PMSC), where a more comprehensive representation is learned by establishing the underlying structure of the original data. But only two-view data with a single complete view or some fully seen samples can be used with these strategies. To release this limitation, numerous generalized clustering techniques have been suggested to process incomplete multiview data. As well as we know, multi-incomplete-view clustering (MIC) [23] is proposed based on a weighted nonnegative matrix factorization with $L_{2,1}$ regularization, with a goal to learn a consensus representation by minimizing the difference between each view representation and the consensus representation. Subsequently, to reduce the memory requirement, Shao et al. [24] further proposed an online method OMVC (online multiview clustering), where a joint weighted nonnegative matrix factorization is applied to chunk-by-chunk handling of the multiview data. In [25], for the multiview clustering problem with k partial-view scenario, Rai et al. proposed a method called GPMVC (graph regularized partial multiview clustering) by exploring each view's fundamental geometry using the view-specific graph Laplacian regularization. Hu and Chen [26] proposed a doubly aligned IMC algorithm by introducing a regress technique to capture more information among multiple views. Besides, some other IMC techniques could be seen in references [27–31].

It should be noted that the methods mentioned above frequently choose to zero-fill the missing instances when dealing with incomplete multiview data; however, doing so would prevent them from being used to their full potential and result in poorer clustering performance, particularly when dealing with high rates of missing instances. To tackle this drawback, some other valid approaches have been recently proposed, such as efficient and effective IMC (EE-IMVC) [32], unified embedding alignment framework (UEAF) [33], and adaptive graph completion-based IMC (AGC_IMC) [34]. Besides, some researchers seek to combine the deep learning

and conventional IMC approach to improve the performance. For example, Shang et al. [35] proposed a two-view approach, named VIGAN, for view imputation via generative adversarial networks (GANs) by combining the denoising autoencoder and GANs, where the denoising autoencoder is used to reconstruct the missing views according to the outputs of GANs. Wang et al. [36] provided a consistent GANs for partial two-view clustering based on autoencoder (AE) and GANs, which are named by PVC-GAN, where the common representation are used to generate the missing data by GANs. Recently, based on the PVC-GAN model, Wang et al. [37] proposed a novel generative partial multiview clustering model with adaptive fusion and cycle consistency, termed as GPMVC-GAN in this paper.

Their common advantage is to use the sample correlation or view correlation to restore missing information and then further improve the clustering performance. This led us to realize that a well descriptions of sample-level and view-level correlations would help to understand the data and hence help to learn a good clustering indicator for clustering improvement. For fully-observed samples, sample correlation could be found by directly calculating the similarity graph. But for the missing sample, its correlations to other samples are failed to be calculated, resulting into the incomplete similarity graph. Fortunately, samples are generally drawn from several low-rank subspaces, this indicates that the corresponding similarity graph also should have the low-rank structure, and hence the incomplete similarity graph (i.e., the missing sample correlation) could be recovered by exploring its low-rank structure. Since an object is described from several views, the generated data of different views should admit the same underlying clustering, i.e., samples in different views should have the same cluster relationship. This means that there is low-rank correlation among similarity graphs of different views. To obtain a well description of sample-level and view-level correlations, we integrate all similarity graphs into a graph tensor, and then perform the low-rank tensor ring decomposition [38, 39] on it to learn sample-level and view-level correlations, simultaneously, where the tensor ring decomposition has been shown to be powerful for high-order correlation exploration and has achieved remarkable results in the incomplete tensor restoration in recent years [40–42]. The multiview spectral clustering is then applied to all of the entire graphs in order to determine the consensus clustering indicator. We combine spectral clustering and graph completion into a unified model to produce complete graphs that considerably increase clustering performance. The whole process is depicted in Figure 1. Besides, considering that distinct views' contributions are often not equal to each other, an adaptive-weighting strategy is applied to the multiview spectral clustering for IMC improvement. Finally, the ADMM is developed to optimize the suggested model, and we test the proposed approach using a number of real-life multiview datasets by comparing it to other cutting-edge techniques. Following is a summary of our paper's novelty and contributions:

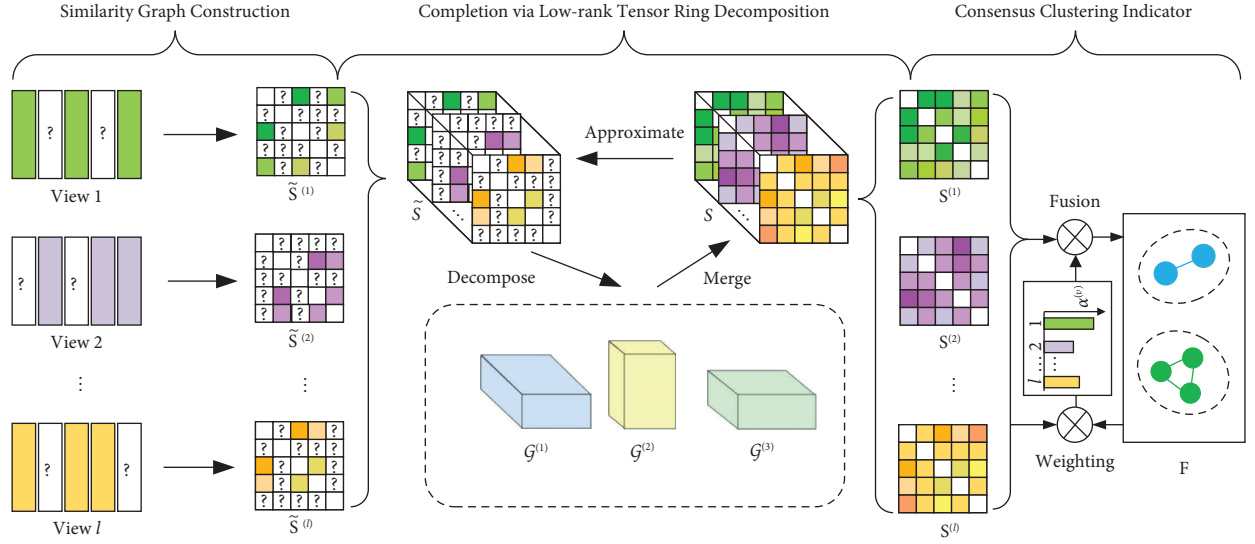


FIGURE 1: The suggested IMC method's framework. For the given multiview data with missing instances, the constructed similarity graphs is incomplete such that fails to provide the complementary information for a good consensus clustering indicator learning \mathbf{F} . To restore the missing information, the tensor ring decomposition is applied to learn a complete tensor representation \mathcal{S} from the incomplete tensor $\tilde{\mathcal{S}}$ generated by stacking incomplete graphs $\tilde{\mathbf{S}}^{(v)}$, $v = 1, 2, \dots, l$ into a 3rd-order tensor. In this way, the complete graphs can be restored and then be fused for learning a consensus clustering indicator \mathbf{F} . Because each viewpoint's contributions vary, the consensus clustering indicator learning process makes use of the adaptive weighting technique.

- (i) We propose a novel IMC algorithm, where the missing information resulting from the missing instances could be restored through tensor ring completion to discover both sample-level and view-level correlations to explore the high-order relationship.
- (ii) The graph tensor completion and consensus clustering indicators are integrated into a unified model and optimized jointly, to fuse different views adaptively and ensure the complete graphs could benefit the clustering task for more accurate learning.
- (iii) The proposed model is optimized using the ADMM technique. The suggested method performs at the cutting edge in the IMC, as demonstrated by experimental results on a number of real-life multiview data.

2. Preliminaries

2.1. Notations. The following simply indicates some of the notations and operations employed by this work. Matrices are represented by bold-face capitals, e.g. \mathbf{D} , \mathbf{F} , and tensors denoted by Calligraphic letters, such as \mathcal{S} , \mathcal{G} . We use $\mathcal{S} = \Phi(\mathbf{S}^{(1)}, \mathbf{S}^{(2)}, \dots, \mathbf{S}^{(l)})$ to stack multiple matrices $\mathbf{S}^{(v)}$, $v = 1, 2, \dots, l$ into a tensor \mathcal{S} with three order, i.e., satisfying $\mathcal{S}(:, :, v) = \mathbf{S}^{(v)}$. Besides, two mode- k unfolding operations of tensor are adopted in our work. For particular, given tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, its standard mode- k unfolding [43] can be represented by matrix $\mathbf{X}_{(k)}$ and calculated by

$$\mathbf{X}_{(k)}(i_k, \overline{i_1 \dots i_{k-1} i_{k+1} \dots i_N}) = \mathcal{X}(i_1, i_2, \dots, i_N), \quad (1)$$

where

$$\begin{aligned} & \overline{i_1 \dots i_{k-1} i_{k+1} \dots i_N} \\ &= i_1 + (i_2 - 1)I_1 + \dots + (i_{k-1} - 1)I_1 \dots I_{k-2} \\ &+ (i_{k+1} - 1)I_1 \dots I_{k-1} + \dots + (i_N - 1) \prod_{j \neq k, N} I_j. \end{aligned} \quad (2)$$

The other one is the modified mode- k unfolding operation [38], which is computed by

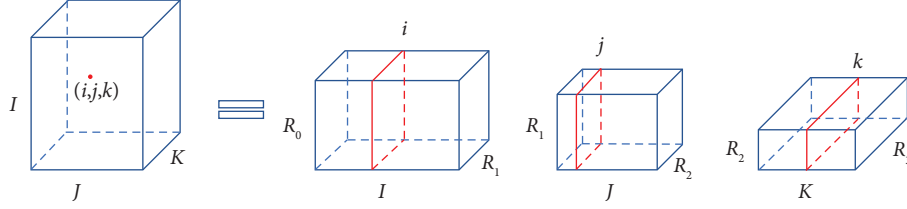
$$\mathbf{X}_{[k]}(i_k, \overline{i_{k+1} \dots i_N i_1 \dots i_{k-1}}) = \mathcal{X}(i_1, i_2, \dots, i_N), \quad (3)$$

where

$$\begin{aligned} & \overline{i_{k+1} i_{k+2} \dots i_N i_1 \dots i_{k-1}} \\ &= i_{k+1} + (i_{k+2} - 1)I_{k+1} + \dots + (i_N - 1)I_{k+1} I_{k+2} \dots I_{N-1} \\ &+ (i_1 - 1)I_{k+1} I_{k+2} \dots I_N + \dots + (i_{k-1} - 1) \prod_{j \neq k, k-1} I_j. \end{aligned} \quad (4)$$

2.2. Tensor Ring (TR) Decomposition [38]. This decomposition has a powerful ability to exploit the high-order correlation and hence has been widely applied for tensor completion, motivated by this advantage; we apply it to graph completion in our work. To make the work self-contained, much relevant knowledge about it is introduced.

2.2.1. Definition. TR decomposition aims to factorize tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ into a sequence of low-order tensors $\{\mathcal{G}^{(1)}, \dots, \mathcal{G}^{(N)}\}$, where $\mathcal{G}^{(k)} \in \mathbb{R}^{R_{k-1} \times I_k \times R_k}$ for $k = 1, 2, \dots, N$, $R_0 = R_N$, are also called the TR-cores and the TR-rank is defined as $[R_1, R_2, \dots, R_N]^T$. The element of \mathcal{X} at index

FIGURE 2: Illustration of tensor ring decomposition on 3rd-order tensors.

(i_1, i_2, \dots, i_N) can be represented by the circular products over N lateral slices $\{\mathbf{G}_{i_1}^{(1)}, \mathbf{G}_{i_2}^{(2)}, \dots, \mathbf{G}_{i_N}^{(N)}\}$, i.e.,

$$\begin{aligned} \mathcal{X}(i_1, \dots, i_N) &= \sum_{r_1=1}^{R_1} \cdots \sum_{r_N=1}^{R_N} \mathbf{G}_{i_1}^{(1)}(r_N, r_1) \cdots \mathbf{G}_{i_N}^{(N)}(r_{N-1}, r_N) \\ &= \text{tr}(\mathbf{G}_{i_1}^{(1)} \cdots \mathbf{G}_{i_N}^{(N)}), \end{aligned} \quad (5)$$

where the trace operation is indicated by $\text{tr}(\cdot)$ and the i_k th lateral slice of $\mathcal{G}^{(k)}$ is represented as $\mathbf{G}_{i_k}^{(k)}$. For simplicity, we denote the tensor ring decomposition of \mathcal{X} using $\mathcal{X} = \mathfrak{R}(\mathcal{G}^{(1)}, \mathcal{G}^{(2)}, \dots, \mathcal{G}^{(N)})$. Besides, to help readers understand, we depict a 3rd-order tensor's TR decomposition in Figure 2.

2.2.2. Merging the Adjacent TR-Cores. This is the commonly used operation in optimizing TR decomposition. Assuming that the tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ with TR-cores $\mathcal{G}^{(k)} \in \mathbb{R}^{R_{k-1} \times I_k \times R_k}$, for $k = 1, \dots, N$, the merging of the two adjacent cores $\{\mathcal{G}^{(k)}, \mathcal{G}^{(k+1)}\}$ can be calculated by

$$\begin{aligned} \mathcal{G}^{(k,k+1)}(r_{k-1}, \overline{i_k i_{k+1}}, r_{k+1}) &= \sum_{r_k=1}^{R_k} \mathbf{G}_{i_k}^{(k)}(r_{k-1}, r_k) \mathbf{G}_{i_{k+1}}^{(k+1)}(r_k, r_{k+1}), \end{aligned} \quad (6)$$

where $\overline{i_k i_{k+1}} = i_k + (i_{k+1} - 1)I_k$. Generally, for d adjacent cores $\{\mathcal{G}^{(k+1)}, \dots, \mathcal{G}^{(k+d)}\}$, their merging can be computed by

$$\begin{aligned} \mathcal{G}^{(k+1,k+2,\dots,k+d)}(r_k, \overline{i_{k+1} i_{k+2} \cdots i_{k+d}}, r_{k+d}) &= \sum_{r_{k+1}=1}^{R_{k+1}} \sum_{r_{k+2}=1}^{R_{k+2}} \cdots \sum_{r_{k+d-1}=1}^{R_{k+d-1}} \mathbf{G}_{i_{k+1}}^{(k+1)}(r_k, r_{k+1}) \\ &\quad \cdot \mathbf{G}_{i_{k+2}}^{(k+2)}(r_{k+1}, r_{k+2}) \cdots \mathbf{G}_{i_{k+d}}^{(k+d)}(r_{k+d-1}, r_{k+d}), \end{aligned} \quad (7)$$

where $\overline{i_{k+1} i_{k+2} \cdots i_{k+d}} = i_{k+1} + (i_{k+2} - 1)I_{k+1} + \dots + (i_{k+d} - 1)I_{k+1} \cdots I_{k+d-1}$.

2.3. Spectral Clustering. Spectral clustering has been shown to be a successful technique for extracting a low-dimensional feature matrix for clustering in past decades. Specifically, assuming that dataset \mathbf{X} is composed of n samples with m features, i.e., $\mathbf{X} = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{m \times n}$, the goal of spectral clustering is to learn a low-dimensional representation, often known as a clustering indicator, \mathbf{F} , from a symmetric nonnegative similarity-graph \mathbf{S} ; where the element $\mathbf{S}_{i,j}$ or $\mathbf{S}_{j,i}$

represents the probability of the i th and j th samples belonging to the same class. The following is how spectral clustering's goal function is typically expressed.

$$\min_{\mathbf{F}} \text{tr}(\mathbf{F}^T \mathbf{L}_S \mathbf{F}) \quad \text{s.t. } \mathbf{F}^T \mathbf{F} = \mathbf{I}, \quad (8)$$

where \mathbf{I} represents the identity matrix, $\mathbf{F} \in \mathbb{R}^{n \times c}$ is the optimized clustering indicator, c denotes the new feature dimension and is usually selected to be the cluster number. $\mathbf{L}_S \in \mathbb{R}^{n \times n}$ denotes the Laplacian matrix of similarity graph \mathbf{S} , computed using the ratio cut [44] $\mathbf{L}_S = \mathbf{D} - \mathbf{S}$ or the normalized cut [45] $\mathbf{L}_S = \mathbf{I} - \mathbf{D}^{-1/2} \mathbf{S} \mathbf{D}^{-1/2}$; where the i th diagonal element of the diagonal matrix \mathbf{D} is calculated by the sum of the i th row or column of \mathbf{S} .

3. The Proposed Method

3.1. Learning Model. Owing to some unavoidable factors in the real-world applications, the collected multiview data often suffers from incompleteness, i.e., only partial samples can be observed from some views. To be specific, for an incomplete multiview data describing n samples from l views, suppose only n_ν ($n_\nu \leq n$) samples with m_ν features can be observed in the ν th view, i.e., $\mathbf{Y}^{(\nu)} \in \mathbb{R}^{m_\nu \times n_\nu}$, then only the connections of n_ν available samples can be revealed, i.e., $\tilde{\mathbf{S}}^{(\nu)} \in \mathbb{R}^{n_\nu \times n_\nu}$. This tends to result in the failure to explore the complementary information for learning a good consensus representation. To overcome this problem, we attempt to propose an IMC approach via tensor completion using low rank decomposition, which mainly consists of three parts: available-connection preservation, low-rank high-order relationship exploration, and consensus clustering indicator learning.

3.1.1. Available-Connection Preservation. Generally, the connections of available instances of the ν th view, i.e., $\tilde{\mathbf{S}}^{(\nu)} \in \mathbb{R}^{n_\nu \times n_\nu}$ should be retained in the recovery complete graph \mathcal{S} . In order to achieve this, we develop the following model:

$$\begin{aligned} \min_{\mathbf{S}^{(\nu)}} \left\| \mathbf{W}^{(\nu)} \odot (\mathbf{S}^{(\nu)} - \tilde{\mathbf{S}}^{(\nu)}) \right\|_F^2 \\ \text{s.t. } 0 \leq \mathbf{S}^{(\nu)} \leq \mathbf{1}, \mathbf{S}^{(\nu)T} \mathbf{1} = \mathbf{1}, \mathbf{S}_{i,i}^{(\nu)} = 0, \end{aligned} \quad (9)$$

where symbol \odot denotes the Hadamard (element-wise) product. $\{\mathbf{W}^{(\nu)}\}_{\nu=1}^l$ is a set of matrix marks, where $\mathbf{W}_{i,j}^{(\nu)} = 1$ denotes that the i th sample and the j th sample both have the instances of the ν th view, otherwise, $\mathbf{W}_{i,j}^{(\nu)} = 0$. graph $\tilde{\mathbf{S}}^{(\nu)}$ is obtained by zero-filling $\tilde{\mathbf{S}}^{(\nu)}$ into the size of $n \times n$ via the following formula:

$$\tilde{\mathbf{S}}^{(v)} = \mathbf{Z}^{(v)} \bar{\mathbf{S}}^{(v)} \mathbf{Z}^{(v)\top}, \quad (10)$$

where $\mathbf{Z}^{(v)} \in \mathbb{R}^{n \times n}$, is defined according to the indexes of available instances:

$$\mathbf{Z}_{i,j}^{(v)} = \begin{cases} 1, & \text{if } y_i^{(v)} \text{ is the } v\text{th view of the } i\text{th sample,} \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

3.1.2. Low-Rank High-Order Relationship Exploration. Since all views admit the same underlying clustering structure, there should exist low-rank relationship among similarity graphs of different views $\mathbf{S}^{(v)}$, $v = 1, 2, \dots, l$. Besides, owing to n samples comes from different groups; there also exist low-rank relationship within view. These demonstrate the stacked graph tensor's low-rank characteristic, where $\mathcal{S} = \Phi(\mathbf{S}^{(1)}, \mathbf{S}^{(2)}, \dots, \mathbf{S}^{(l)})$, $\mathcal{S}(:, :, v) = \mathbf{S}^{(v)}$, and thus provide the reliability for graph completion via exploring the low-rank high-order relationship among samples. In this way, the incomplete graphs, resulting from the missing instances, could be restored. Inspired by this motivation, one powerful tool of exploiting low-rank high-order correlation plays an important role in missing information recovery. To this end, one powerful tensor tool is selected in our work, i.e., tensor ring decomposition, which is shown to be powerful for high-order correlation exploration and has achieved remarkable results in the incomplete tensor restoration in recent years. Thus, the recovery model can be formulated as follow:

$$\begin{aligned} \min_{\mathcal{S}^{(v)}} & \frac{1}{2} \left\| \mathcal{S} - \mathcal{R}(\mathcal{G}^{(1)}, \mathcal{G}^{(2)}, \mathcal{G}^{(3)}) \right\|_F^2 \\ \text{s.t.} & \mathcal{S} = \Phi(\mathbf{S}^{(1)}, \mathbf{S}^{(2)}, \dots, \mathbf{S}^{(l)}), \end{aligned} \quad (12)$$

where $\{\mathcal{G}^{(1)} \in \mathbb{R}^{R_3 \times I_1 \times R_1}, \mathcal{G}^{(2)} \in \mathbb{R}^{R_1 \times I_2 \times R_2}, \mathcal{G}^{(3)} \in \mathbb{R}^{R_2 \times I_3 \times R_3}\}$ denote the TR-cores of the graph tensor \mathcal{S} , and $[R_1, R_2, R_3]^\top$ represents the TR-rank. By exploring the low-rank tensoring structure via optimizing model (12), the complete graph tensor \mathcal{S} , in turn, can be obtained. However, as shown in references [41, 46], the TR decomposition is easily affected by its TR-rank selection during the tensor completion process. To overcome this drawback, following the work in [46], we further add the Frobenius norm of TR-core, i.e. $\mu/2 \sum_{k=1}^3 \|\mathcal{G}^{(k)}\|_F^2$, to reduce the sensitivity of the TR decomposition to its TR-rank selection, which has been verified to can achieve rather good completion results even when the selected TR-rank increases and more details can be found in [46]. In order to do this, the recovery model might be further defined as

$$\begin{aligned} \min_{\mathcal{S}^{(v)}} & \frac{1}{2} \left\| \mathcal{S} - \mathcal{R}(\mathcal{G}^{(1)}, \mathcal{G}^{(2)}, \mathcal{G}^{(3)}) \right\|_F^2 + \frac{\mu}{2} \sum_{k=1}^3 \|\mathcal{G}^{(k)}\|_F^2 \\ \text{s.t.} & \mathcal{S} = \Phi(\mathbf{S}^{(1)}, \mathbf{S}^{(2)}, \dots, \mathbf{S}^{(l)}), \end{aligned} \quad (13)$$

where parameter μ is positive value to balance the importance of the corresponding term.

3.1.3. Consensus Clustering Indicator Learning. After incomplete graphs of all views are restored, the consensus clustering indicator shared by every view is often learned via the following spectral clustering model, i.e.,

$$\min_{\mathbf{F}} \sum_{v=1}^l \text{tr}(\mathbf{F}^\top \mathbf{L}_{\mathcal{S}^{(v)}} \mathbf{F}) \quad \text{s.t. } \mathbf{F}^\top \mathbf{F} = \mathbf{I}, \quad (14)$$

where $\mathbf{F} \in \mathbb{R}^{n \times c}$ denotes the consensus clustering indicator, c is the manually chosen dimension that is frequently specified as the cluster number. Since our work's optimized $\mathbf{S}^{(v)}$ is not a symmetric matrix, we compute $\mathbf{L}_{\mathcal{S}^{(v)}}$ by

$$\mathbf{L}_{\mathcal{S}^{(v)}} = \mathbf{D}^{(v)} - \frac{\mathbf{S}^{(v)} + \mathbf{S}^{(v)\top}}{2}, \quad (15)$$

where $\mathbf{D}^{(v)}$ is a diagonal matrix whose i th diagonal element is computed as

$$\mathbf{D}_{i,i}^{(v)} = \sum_{j=1}^n \frac{(\mathbf{S}_{i,j}^{(v)} + \mathbf{S}_{j,i}^{(v)})}{2}. \quad (16)$$

Note that model (15) treats all views equally to the consensus clustering indicator learning, which may actually reduce the flexibility of the method because the contributions of all views are commonly different. Inspired by this motivation, the following adaptively weighting strategy $\{\alpha^{(v)}\}_{v=1}^l$ is leveraged to improve the model, i.e.,

$$\begin{aligned} \min_{\mathbf{F}, \alpha^{(v)}} & \sum_{v=1}^l \alpha^{(v)2} \text{tr}(\mathbf{F}^\top \mathbf{L}_{\mathcal{S}^{(v)}} \mathbf{F}) \\ \text{s.t.} & \sum_{v=1}^l \alpha^{(v)} = 1, \alpha^{(v)} \geq 0, \mathbf{F}^\top \mathbf{F} = \mathbf{I}, \end{aligned} \quad (17)$$

where the non-negative $\alpha^{(v)}$ denotes the normalized weighting parameter of the v -th view.

3.1.4. Overall Objective Function. As analyzed above, the above-mentioned three parts play different key roles in IMC. To take full advantage of these three parts, we integrate them into a unified framework and thus getting the overall objective model:

$$\begin{aligned} \min_{\mathcal{S}^{(v)}, \mathbf{F}, \alpha^{(v)}} & \frac{1}{2} \left\| \mathbf{W}^{(v)} \odot (\mathbf{S}^{(v)} - \tilde{\mathbf{S}}^{(v)}) \right\|_F^2 \\ & + \frac{1}{2} \left\| \mathcal{S} - \mathcal{R}(\mathcal{G}^{(1)}, \mathcal{G}^{(2)}, \mathcal{G}^{(3)}) \right\|_F^2 + \frac{\mu}{2} \sum_{k=1}^3 \|\mathcal{G}^{(k)}\|_F^2 \\ & + \beta \sum_{v=1}^l \alpha^{(v)2} \text{tr}(\mathbf{F}^\top \mathbf{L}_{\mathcal{S}^{(v)}} \mathbf{F}) \\ \text{s.t.} & 0 \leq \mathbf{S}^{(v)} \leq 1, \mathbf{S}^{(v)\top} \mathbf{1} = \mathbf{1}, \mathbf{S}_{i,i}^{(v)} = 0, \sum_{v=1}^l \alpha^{(v)} = 1, \\ & \alpha^{(v)} \geq 0, \mathbf{F}^\top \mathbf{F} = \mathbf{I}, \mathcal{S} = \Phi(\mathbf{S}^{(1)}, \mathbf{S}^{(2)}, \dots, \mathbf{S}^{(l)}), \end{aligned} \quad (18)$$

where parameter β is positive value to balance the importance of the corresponding term.

As seen from equation (18), the first term plays a role to preserve the available connection of observed samples during the optimization of graphs $\{\mathbf{S}^{(v)}, v = 1, 2, \dots, l\}$. The middle two terms attempt to discover the low-rank high-order corrections of samples across all views by implementing the tensor ring decomposition on a stacked graph tensor $\mathcal{S} = \Phi(\mathbf{S}^{(1)}, \mathbf{S}^{(2)}, \dots, \mathbf{S}^{(l)})$, satisfying $\mathcal{S}(:, :, v) = \mathbf{S}^{(v)}$, and then to obtain a series of complete graphs $\{\mathbf{S}^{(1)}, \mathbf{S}^{(2)}, \dots, \mathbf{S}^{(l)}\}$. Finally, the last term aims to learn a consensus clustering indicator \mathbf{F} via implementing spectral clustering on the optimal complete graphs, where adaptive weights $\{\alpha^{(v)}\}_{v=1}^l$ are leveraged to weigh the contributions of different views. Furthermore, to ensure the

complete graphs would bring a significant improvement in the clustering performance, these terms are integrated into a unified framework and optimized jointly. To summarize, we highlight that the proposed method has two advantages: (1) In order to examine the high-order correlations, our model concurrently takes into account sample-level and view-level correlations, which is important for incomplete multiview data. (2) The suggested approach incorporates consensus multiview information that might guide the learners for more precise clustering.

3.2. Optimization. The augmented Lagrangian function of model (18) can be defined as

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \|\mathcal{W} \odot (\mathcal{S} - \tilde{\mathcal{S}})\|_F^2 + \frac{1}{2} \|\mathcal{S} - \mathcal{R}(\mathcal{G}^{(1)}, \mathcal{G}^{(2)}, \mathcal{G}^{(3)})\|_F^2 \\ & + \frac{\mu}{2} \sum_{k=1}^3 \|\mathcal{G}^{(k)}\|_F^2 + \beta \sum_{v=1}^l \alpha^{(v)2} \text{tr}(\mathbf{F}^\top \mathbf{L}_{\mathbf{S}^{(v)}} \mathbf{F}) \end{aligned} \quad (19)$$

$$\text{s.t.} \quad 0 \leq \mathbf{S}^{(v)} \leq \mathbf{1}, \mathbf{S}^{(v)\top} \mathbf{1} = \mathbf{1}, \mathbf{S}_{i,i}^{(v)} = 0, \sum_{v=1}^l \alpha^{(v)} = 1,$$

$$\alpha^{(v)} \geq 0, \mathbf{F}^\top \mathbf{F} = \mathbf{I}, \mathcal{S} = \Phi(\mathbf{S}^{(1)}, \mathbf{S}^{(2)}, \dots, \mathbf{S}^{(l)}).$$

Consequently, the model can be optimized as follows:

3.2.1. Update Variable $\mathcal{G}^{(k)}$. Note that according to reference [38], we can get

$$\|\mathcal{R}(\mathcal{G}^{(1)}, \mathcal{G}^{(2)}, \mathcal{G}^{(3)})\|_F^2 = \|\mathbf{G}_{(2)}^{(k)} (\mathbf{G}_{[2]}^{(\#k)})^\top\|_F^2, \quad (20)$$

where $\mathcal{G}^{(\#k)}$ is a subchain tensor generated by merging all but k th core tensor. Hence, fixing other variables, the augmented Lagrangian function with respect to $\mathcal{G}^{(k)}$ can be simplified as

$$\mathcal{L}(\mathbf{G}_{(2)}^{(k)}) = \frac{1}{2} \|\mathbf{S}_{[k]} - \mathbf{G}_{(2)}^{(k)} (\mathbf{G}_{[2]}^{(\#k)})^\top\|_F^2 + \frac{\mu}{2} \|\mathbf{G}_{(2)}^{(k)}\|_F^2, \quad (21)$$

The derivative of $\mathcal{L}(\mathbf{G}_{[k]}^{(k)})$ with respect to $\mathbf{G}_{[k]}^{(k)}$ is

$$\frac{\partial \mathcal{L}}{\partial \mathbf{G}_{(2)}^{(k)}} = \left(\mathbf{G}_{(2)}^{(k)} (\mathbf{G}_{[2]}^{(\#k)})^\top - \mathbf{S}_{[k]} \right) \mathbf{G}_{(2)}^{(k)} + \mu \mathbf{G}_{(2)}^{(k)}. \quad (22)$$

Let $\partial \mathcal{L} / \partial \mathbf{G}_{(2)}^{(k)} = 0$, we can obtain

$$\mathbf{G}_{(2)}^{(k)} = \mathbf{S}_{[k]} \mathbf{G}_{[2]}^{(\#k)} \left(\mu \mathbf{I} + (\mathbf{G}_{[2]}^{(\#k)})^\top \mathbf{G}_{[2]}^{(\#k)} \right)^{-1}. \quad (23)$$

3.2.2. Update Variable $\mathbf{S}^{(v)}$. Fixing other variables, the problem with respect to $\mathbf{S}^{(v)}$ is reduced to solve the following problem:

$$\begin{aligned} \min_{\mathbf{S}^{(v)}} & \frac{1}{2} \left\| \left(\mathbf{S}^{(v)} - \tilde{\mathbf{S}}^{(v)} \right) \odot \mathbf{W}^{(v)} \right\|_F^2 + \beta \alpha^{(v)2} \text{tr}(\mathbf{F}^\top \mathbf{L}_{\mathbf{S}^{(v)}} \mathbf{F}) \\ & + \frac{1}{2} \|\mathbf{S}^{(v)} - \mathbf{Z}^{(v)}\|_F^2 \end{aligned}$$

$$\text{s.t.} \quad 0 \leq \mathbf{S}^{(v)} \leq \mathbf{1}, \mathbf{S}^{(v)\top} \mathbf{1} = \mathbf{1}, \mathbf{S}_{i,i}^{(v)} = 0, \quad (24)$$

where we define $\mathbf{Z}^{(v)} = \mathcal{R}(\mathcal{G}^{(1)}, \mathcal{G}^{(2)}, \mathbf{G}_v^{(3)})$. Note that $\text{tr}(\mathbf{F}^\top \mathbf{L}_{\mathbf{S}^{(v)}} \mathbf{F}) = 1/2 \sum_{i=1}^n \sum_{j=1}^n \mathbf{S}_{i,j}^{(v)} \|\mathbf{F}_{i,:} - \mathbf{F}_{j,:}\|_2^2$, by defining $\mathbf{H}_{i,j} = \|\mathbf{F}_{i,:} - \mathbf{F}_{j,:}\|_2^2$, the problem (24) can be rewritten into

$$\begin{aligned} \min_{\mathbf{S}^{(v)}} & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \mathbf{W}_{i,j}^{(v)} \left(\mathbf{S}_{i,j}^{(v)} - \tilde{\mathbf{S}}_{i,j}^{(v)} \right)^2 \\ & + \frac{\beta \alpha^{(v)2}}{2} \sum_{i=1}^n \sum_{j=1}^n \mathbf{H}_{i,j} \mathbf{S}_{i,j}^{(v)} + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \left(\mathbf{S}_{i,j}^{(v)} - \mathbf{Z}_{i,j}^{(v)} \right)^2 \end{aligned} \quad (25)$$

$$\text{s.t.} \quad 0 \leq \mathbf{S}^{(v)} \leq \mathbf{1}, \mathbf{S}^{(v)\top} \mathbf{1} = \mathbf{1}, \mathbf{S}_{i,i}^{(v)} = 0.$$

Through mathematical transformation, the above problem is equivalent to solve the following problem:

$$\min_{\mathbf{S}^{(v)}} \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\mathbf{S}_{i,j}^{(v)} - \mathbf{T}_{i,j}^{(v)})^2 \quad (26)$$

$$\text{s.t. } 0 \leq \mathbf{S}^{(v)} \leq \mathbf{1}, \mathbf{S}^{(v)\top} \mathbf{1} = \mathbf{1}, \mathbf{S}_{i,i}^{(v)} = 0,$$

where

$$\mathbf{T}_{i,j}^{(v)} = \frac{W_{i,j}^{(v)} \tilde{\mathbf{S}}_{i,j}^{(v)} + Z_{i,j}^{(v)} - (\beta \alpha^{(v)2} / 2) \mathbf{H}_{i,j}}{W_{i,j}^{(v)}}. \quad (27)$$

Since problem (26) is independent with respect to each column, we can solve it column by column using the optimization method in reference [47], i.e., the solution of (26) can be given by

$$\mathbf{S}_{i,j}^{(v)} = \begin{cases} \max(\mathbf{T}_{i,j}^{(v)} + \eta_j, 0), & i \neq j, \\ 0, & i = j. \end{cases} \quad (28)$$

Because of the constraints $\mathbf{S}^{(v)\top} \mathbf{1} = \mathbf{1}$ and $\mathbf{S}_{i,i}^{(v)} = 0$, we can get

$$\eta_j = \frac{1 - \sum_{i=1, i \neq j}^n \mathbf{T}_{i,j}^{(v)}}{n-1}. \quad (29)$$

3.2.3. Update Variable F. By fixing other variables, we can obtain the new representation \mathbf{F} by solving the following subproblem, i.e.,

$$\min_{\mathbf{F}} \sum_{v=1}^l \alpha^{(v)2} \text{tr}(\mathbf{F}^\top \mathbf{L}_{\mathbf{S}^{(v)}} \mathbf{F}) \quad \text{s.t. } \mathbf{F}^\top \mathbf{F} = \mathbf{I}. \quad (30)$$

Through mathematical transformation, the above problem can be further rewritten as

$$\min_{\mathbf{F}} \text{tr} \left(\mathbf{F}^\top \sum_{v=1}^l \alpha^{(v)2} \mathbf{L}_{\mathbf{S}^{(v)}} \mathbf{F} \right) \quad \text{s.t. } \mathbf{F}^\top \mathbf{F} = \mathbf{I}. \quad (31)$$

The optimal solution \mathbf{F} can be given by the eigenvectors set corresponding to the first c smallest eigenvalues of matrix $\sum_{v=1}^l \alpha^{(v)2} \mathbf{L}_{\mathbf{S}^{(v)}}$.

3.2.4. Update variable $\alpha^{(v)}$, $v = 1, 2, \dots, l$. To properly balance the role of each view, we adaptively update parameters $\alpha^{(v)}$, $v = 1, 2, \dots, l$ using an effective mechanism, i.e.,

$$\begin{aligned} \min_{\alpha^{(v)}} \sum_{v=1}^l \alpha^{(v)2} \text{tr}(\mathbf{F}^\top \mathbf{L}_{\mathbf{S}^{(v)}} \mathbf{F}) \\ \text{s.t. } \sum_{v=1}^l \alpha^{(v)} = 1, \alpha^{(v)} \geq 0. \end{aligned} \quad (32)$$

Let $\phi^{(v)} = \text{tr}(\mathbf{F}^\top \mathbf{L}_{\mathbf{S}^{(v)}} \mathbf{F})$, then (32) can be rewritten into

$$\min_{\alpha^{(v)}} \sum_{v=1}^l \alpha^{(v)2} \phi^{(v)} \quad \text{s.t. } \sum_{v=1}^l \alpha^{(v)} = 1, \alpha^{(v)} \geq 0. \quad (33)$$

The Lagrange function of equation (33) is

$$L(\alpha^{(v)}) = \sum_{v=1}^l \alpha^{(v)2} \phi^{(v)} - \lambda \left(\sum_{v=1}^l \alpha^{(v)} - 1 \right). \quad (34)$$

Let the derivative of (34) wrt. $\alpha^{(v)}$ to be zero, i.e.,

$$\frac{\partial L(\alpha^{(v)})}{\partial \alpha^{(v)}} = 2\alpha^{(v)} \phi^{(v)} - \lambda = 0, \quad (35)$$

we have

$$\alpha^{(v)} = \frac{\lambda}{2\alpha^{(v)}}. \quad (36)$$

According to the condition $\sum_{v=1}^l \alpha^{(v)} = 1$, we can get the optimal $\alpha^{(v)}$ is

$$\alpha^{(v)} = \frac{\phi^{(v)-1}}{\sum_{v=1}^l \phi^{(v)-1}}. \quad (37)$$

The above computation procedures are summarized in Algorithm 1. The k -means technique is applied to the consensus representation F to produce the final clustering results.

3.3. Computational Complexity Analysis. As seen from Algorithm 1, the computation cost of the suggested approach is caused by updating the variables $\mathcal{G}^{(k)}$, $\mathbf{S}^{(v)}$, \mathbf{F} , $\alpha^{(v)}$. The suggested approach's computational cost is evaluated as follows, assuming the multiview data has l views, n samples of m_v features, and c class number. In the update of $\mathcal{G}^{(k)}$, the main computational burden comes from the inversion part $(\mu \mathbf{I} + (\mathbf{G}_{[2]}^{(\#k)})^\top \mathbf{G}_{[2]}^{(\#k)})^{-1}$ with complexity $\mathcal{O}(R^4 n^2 + R^6)$, where the TR-rank is selected by $R_1 = R_2 = R_3 = R$ in default. Due to the fact that the update of $\mathbf{S}^{(v)}$ just uses some-wise based operations, its computational cost can be disregarded. The update of \mathbf{F} needs to calculate the first c smallest eigenvalues of $\sum_{v=1}^l \alpha^{(v)2} \mathbf{L}_{\mathbf{S}^{(v)}}$, which can be speeded up using an efficient function "eigs" [48] with computational complexity $\mathcal{O}(cn^2)$. As seen from (37), since the numerical division operation may be used to calculate the variable $\alpha^{(v)}$, the complexity of their computations can be disregarded. According to the study above, the proposed method's overall computational complexity is around $\mathcal{O}(R^4 n^2 + R^6 + cn^2)$ for each iteration.

4. Experiments

The proposed method, IMC-LTR, would be compared with other cutting-edge methods through studies on various actual multiview datasets with different missing rate samples. Additionally, we do experiments to test the proposed method's convergence characteristics as well as the impact of the adaptive weighting strategy. The suggested approach has two hyperparameters: μ and β , which are both tuned in our experiments from the range of $[10^{-5}, 10^{-4}, \dots, 10^1]$. Furthermore, we simply set $R_1 = R_2 = R_3 = R$ for the TR-rank of the proposed method, $[R_1, R_2, R_3]$, and then fine-tune the value of R to get the optimum outcomes. Note that, other compared methods are adjusted for the best results, per the relevant articles.

Require: Multiview data $\{\mathbf{X}^{(v)}\}_{v=1}^l$, parameters μ, β .

Initialization: Construct the similarity graphs $\bar{\mathbf{S}}^{(v)}, v = 1, \dots, l$ from observable instances of each view $\mathbf{X}^{(v)}$, and then fill it into $\bar{\mathbf{S}}^{(v)}$ via formular (10). Initialize \mathbf{F} by solving (31).

- (1) **while** not converged **do**
- (2) Update $\mathcal{G}^{(k)}, k = 1, 2, 3$ by equation (23).
- (3) Update variable $\mathbf{S}^{(v)}, v = 1, \dots, l$ by equation (28).
- (4) Update \mathbf{F} by solving (31).
- (5) Update variable $\alpha^{(v)}, v = 1, \dots, l$ by equation (37).
- (6) **end while**
- (7) Return \mathbf{F} .

ALGORITHM 1: Algorithm to solve equation (18).

4.1. Dataset Description and Incomplete Multiview Data Construction

- (1) MSRC-v1 (<https://github.com/youweiliang/ConsistentGraphLearning/>): This dataset contains 8 categories of images. Following [49], 7 categories are chosen in our experiment, i.e., cow, airplane, building, face, bicycle, and car, where each has 30 images. Here, CENTRIST, GIST, color moment, HOG, and LBP are selected as five views, with feature dimensions 254, 512, 24, 576, and 256, respectively.
- (2) NewsGroups Dataset (<https://lig-membres.imag.fr/grimal/data.html>): This contains subsets of the NG20 dataset. In the experiment, we select a subset composed of 500 documents over 5 topics with 3 different preprocessing: supervised mutual information, partitioning around medoids, and unsupervised mutual information; where each preprocessing has a feature dimension of 2000.
- (3) ORL Database (<https://www.cl.cam.ac.uk/research/dtg/attarchive/>): This image dataset contains 400 face images from 40 distinct subjects, where some of them were taken at different times, varying the lighting, facial expressions, and facial details. Here, 3 types of features are selected as 3 views, i.e., 4096 dimension intensity feature, 3304 dimension LBP feature, and 6750 dimension Gabor feature.
- (4) Yale (<https://www.cl.cam.ac.uk/research/dtg/attarchive/>): This image dataset contains 165 gray-scale images of 15 individuals. We select 3 types of features as 3 views, i.e., 4096 dimension intensity feature, 3304 dimension LBP feature, and 6750 dimension Gabor feature.

Incomplete Multiview Data Construction: Assuming the dataset with n_v views, we randomly select p of samples to be fully observed from all views, and then make the remaining samples only can be observed from one view, i.e., each view is constructed by p fully-observed samples and $1 - p/n_v$ single-view samples. Here, p can be seen as the fully-observed rate, and it is chosen as 30%, 50%, and 70% in our experiment.

4.2. Compared Methods and Evaluation Metric. Other approaches, designed to deal with the incomplete multiview data, are chosen to compare the suggested approach, i.e.,

4.2.1. *Best Single View (BSV)* [21]. The best single-view clustering result was determined by performing k-means on every view, separately, where the missing instances are filled by using the mean of other observable instances of the same view.

4.2.2. *Concat* [21]. Concat filled the missing instances by using the mean of other observable instances of the same view, and then stacked all instances into a single-view dataset of a large feature dimension. Next, the clustering result is obtained by implementing k-means on the single-view dataset.

4.2.3. *Multi-Incomplete-View Clustering (MIC)* [23]. MIC is proposed based on a weighted nonnegative matrix factorization with $L_{2,1}$ regularization, with a goal to learn a consensus representation via the minimization of the difference between each view representation and the consensus representation.

4.2.4. *Online Multiview Clustering (OMVC)* [24]. OMVC created a joint weighted nonnegative matrix factorization to handle the multiview data chunk by chunk in order to save memory.

4.2.5. *Graph Regularized Partial Multiview Clustering (GPMVC)* [25]. By using the view-specific graph Laplacian regularization to explore each view's intrinsic geometry, GPMVC seeks to tackle the incomplete multiview issue.

4.2.6. *Adaptive Graph Completion-Based IMC (AGC_IMC)* [34]. To improve the multiview clustering performance, AGC_IMC is proposed to learn an effective common representation via a multiview spectral learning model that includes the graph completion.

4.2.7. *Incomplete Multiview Tensor Spectral Clustering with Missing View Inferring (IMVTSC-MVI)* [50]. This approach introduces a tensor low-rank constraint and manifold space based incomplete multiview graph learning method.

4.2.8. *Localized Sparse IMC (LSIMVC)* [51]. Recently, the LSIMVC method was proposed as a way to extract incomplete multiview data and learn a sparse and structured shared low-dimensional representation.

4.2.9. *Generative Partial Multiview Clustering Based on GAN (GPMVC-GAN)* [37]. To solve the incomplete multiview problem by explicitly generating the data of missing views, this generative partial multiview clustering model was proposed by design and build with adaptive fusion and cycle consistency, which is the newly proposed GAN-based IMC method.

We compare the accuracy (ACC), normalized mutual information (NMI), purity, adjusted rand index (AR), and F-score [52–54] of the suggested method to those of previous incomplete multiview approaches in attempt to demonstrate the advantage of the suggested approach. Higher values indicate greater clustering performance for the aforementioned well-known metrics. In our tests, each approach is applied multiple times with various views absent groups, and the method's final findings are given based on the average results.

4.3. *Experiment Results and Analyses.* The suggested approach is evaluated against other approaches using a number of real-world multiview datasets with varying observed rate samples. Experimental findings are reported in Table 1 and Figure 3. The following conclusions can be drawn from these experimental findings:

- (1) The suggested approach typically outperforms other IMC algorithms on multiview datasets with various missing rate samples. For instance, the proposed method improves ACC by around 12 percent, 8 percent, and 8 percent, respectively, on the MSRC-v1 dataset, in contrast towards the second-best results, at observed rates of 0.3, 0.5, and 0.7, respectively. On the NewsGroups dataset with observed rates of 0.3, 0.5, and 0.7, the proposed method, respectively, obtains about 14%, 9%, and 7% improvement of NMI over the second-best AGC_IMC. These outcomes demonstrate the effectiveness of the suggested technique in IMC, which may be due to the tensor ring decomposition's potent capability in exploiting the higher relationship among different samples cross all the views.
- (2) Observe that, in most situations, multiview approaches outperform single-view approaches, i.e., BSV and Concat. This may benefit from the compatible and supplementary data offered by a variety of perspectives. Also due to the same reason, Concat outperforms BSV in most cases. However,

sometimes, multiview methods are inferior to single-view methods under some missing cases, which demonstrate that data missing would affect the capacity of multiview methods in learning the complementary information from multiple views to some extent.

- (3) Compared methods BSV, Concat, MIC, OMVC, GPMVC, LSIMVC, and GPMVC-GAN are inferior to completion based methods, i.e. AGC_IMC, IMVTSC-MVI, and the proposed method, in most missing cases. The reason may come from the difference of their strategies in handling the missing instance. Note that BSV, Concat, MIC, OMVC, GPMVC, and LSIMVC, chose to replace the empty instances using the mean of the existing instances or zero, which is more likely to damage the intrinsic structure of each view and hence result in poorer performance. In contrast, AGC_IMC, IMVTSC-MVI, and the proposed method are completion-based methods, which could adaptively fill in the missing elements during the optimization, hence obtaining an optimal solution for clustering. These show that a good filling strategy is crucial for achieving good clustering performance. Besides, we find that the GPMVC-GAN method obtains poor results and which may be caused by its poor robustness to high-dimensional feature data. This reflects the advantages of the proposed model in processing high-dimensional data compared with GAN-based the IMC method.
- (4) When compared to other completion-based IMC approaches, such as AGC_IMC and IMVTSC-MVI, the suggested approach greatly improves the clustering performance in the majority of cases. It reveals that the suggested approach fills in the missing data more effectively than the other two approaches. This result may benefit from the powerful ability of the tensor ring decomposition in exploring the low-rank structure.

4.4. *Sensitivity Analysis of the Penalty Parameters.* The proposed method has two penalty parameters μ and β . In this section, we would analyze the sensitivities of these two parameters in terms of the clustering accuracy.

We conduct experiments on the above mentioned datasets (MSRC-v1, NewsGroups, ORL, and Yale) with 30% observable paired-sample under different combinations of parameters μ and β , selected from a set $\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10^1, 10^2, 10^3, 10^4, 10^5\}$. Experimental results of the proposed method on the above-mentioned four datasets are shown in Figure 4. As for the MSRC-v1 dataset, the proposed method can obtain a relative good clustering performance when $\mu \in [10^{-5}, 10^5]$ and $\beta \in [10^{-5}, 1]$. As for the NewsGroups dataset, the proposed dataset can achieve better clustering results when $\mu \in [10^{-5}, 10^5]$ and $\beta \in [10^{-5}, 10^{-1}]$. As for the ORL dataset, a relative good clustering result could be obtained when $\mu \in [10^{-5}, 10^5]$ and $\beta \in [10^{-5}, 1]$. As for the Yale dataset, the

TABLE 1: Average values of ACC (%), NMI (%), and purity (%) of different methods on MSRC-v1, NewsGroups, ORL, and Yale datasets with different observed rates of paired-samples.

Dataset	Method/rate	ACC (%)			NMI (%)			Purity (%)		
		0.3	0.5	0.7	0.3	0.5	0.7	0.3	0.5	0.7
MSRC-v1	BSV [21]	38.67	45.90	57.29	31.02	37.56	47.95	39.19	46.67	58.19
	Concat [21]	32.05	36.19	39.24	24.45	27.05	30.98	33.67	37.62	41.10
	MIC [23]	43.29	61.52	68.10	37.19	54.09	61.45	45.95	63.48	70.00
	OMVC [24]	44.24	37.81	41.29	38.06	29.33	31.57	47.29	39.43	42.48
	GPMVC [25]	52.57	54.95	66.24	41.52	45.48	55.74	53.90	56.76	67.48
	AGC_IMC [34]	55.62	71.33	71.57	42.99	57.36	61.66	57.48	72.38	74.86
	IMVTSC-MVI [50]	48.10	60.00	73.81	36.52	50.17	65.79	48.73	60.32	73.81
	LSIMVC [51]	34.00	35.14	36.24	27.04	27.37	27.68	37.10	37.19	38.29
	GPMVC-GAN [37]	22.19	22.38	24.73	5.77	5.88	9.23	23.24	23.57	26.10
	IMC-LTR	67.43	79.14	81.71	51.11	64.69	68.53	67.71	79.14	81.71
NewsGroups	BSV [21]	20.60	21.68	23.68	4.48	7.58	10.13	20.88	22.14	23.84
	Concat [21]	20.70	21.00	22.06	4.88	5.53	7.61	21.02	21.28	22.26
	MIC [23]	26.44	28.94	29.54	11.26	15.85	17.15	26.88	29.96	31.08
	OMVC [24]	23.70	23.32	24.46	7.26	7.54	8.00	24.04	23.78	24.76
	GPMVC [25]	32.32	34.06	37.18	9.93	11.30	15.19	34.44	36.32	40.30
	AGC_IMC [34]	67.26	84.12	88.86	45.88	61.55	71.01	68.06	84.12	88.86
	IMVTSC-MVI [50]	46.00	59.20	80.07	22.08	42.60	65.68	46.20	59.47	80.20
	LSIMVC [51]	28.66	32.62	32.72	8.48	12.96	16.46	30.52	36.38	36.56
	GPMVC-GAN [37]	25.08	26.72	27.59	5.90	7.80	8.27	25.76	27.36	28.15
	IMC-LTR	83.06	88.74	92.58	59.01	70.14	78.87	83.06	88.74	92.58
ORL	BSV [21]	26.98	35.18	43.33	46.54	54.24	62.44	31.03	38.85	47.28
	Concat [21]	32.20	37.23	43.75	50.52	56.60	63.19	34.78	40.93	47.73
	MIC [23]	42.83	60.68	63.88	62.44	78.03	81.46	45.88	64.63	68.68
	OMVC [24]	29.88	30.13	33.85	54.39	55.01	57.41	32.40	32.58	36.48
	GPMVC [25]	42.78	49.28	53.78	62.72	67.53	71.34	45.53	52.53	56.68
	AGC_IMC [34]	40.88	53.18	64.38	62.85	70.91	79.03	43.93	55.98	67.30
	IMVTSC-MVI [50]	29.50	45.00	59.25	49.67	62.58	74.94	31.33	47.58	63.25
	LSIMVC [51]	52.53	57.65	64.30	70.98	75.32	78.33	55.68	60.83	66.43
	GPMVC-GAN [37]	14.70	15.08	15.92	34.36	35.93	37.68	16.25	16.32	16.88
	IMC-LTR	50.75	62.58	69.83	69.05	76.27	81.34	52.90	64.20	71.43
Yale	BSV [21]	29.94	37.33	42.48	37.72	44.31	49.03	31.88	38.73	44.36
	Concat [21]	33.76	39.09	41.39	40.46	44.75	47.01	35.15	40.61	43.64
	MIC [23]	35.70	47.88	52.73	42.33	52.32	56.54	37.21	48.91	53.33
	OMVC [24]	27.09	25.82	27.70	34.57	32.61	34.97	30.00	27.82	29.09
	GPMVC [25]	42.18	49.21	56.24	50.13	56.15	61.96	44.00	50.06	56.79
	AGC_IMC [34]	39.21	46.85	60.48	45.09	53.19	64.12	40.61	48.48	61.27
	IMVTSC-MVI [50]	37.78	48.28	63.64	42.35	51.90	65.72	38.38	48.69	64.04
	LSIMVC [51]	43.21	50.36	54.18	50.60	55.32	61.00	44.18	51.82	55.94
	GPMVC-GAN [37]	18.79	19.03	19.84	23.26	22.73	23.70	20.61	20.79	21.66
	IMC-LTR	49.15	57.70	65.52	51.91	59.51	65.71	50.12	58.61	65.58

Bold numbers denote the best result.

proposed method obtains stable and better performance when $\mu \in [1, 10^5]$ and $\beta \in [10^{-5}, 1]$. The experimental results show that it is easy to select the penalty parameters of the proposed method. Based on the analysis, we simply select the values of μ and β by $\mu = 1$ and $\beta = 10^{-3}$ in the experiments of the previous section.

4.5. Effect of Adaptive Weighting Strategy. To demonstrate the efficacy of the adaptive weighting strategy in multiview clustering, we undertake trials on the MRSC-v1 and Yale datasets with paired-sample observed rates of 0.3, 0.5, and 0.7. Experimental findings in terms of ACC, NMI, and purity, are displayed in Figure 5, where IMC-LTR-W and IMC-LTR-WO denote the proposed methods with and

without adaptive weighting, respectively. Note that, since the adaptive weighting strategy is removed in IMC-LTR-WO, the weights $\alpha^{(v)}$, $v = 1, \dots, l$ are all equal to $1/l$. Observe from the experimental findings in Figure 5 that the adaptive weighting strategy contributes to the enhancement of multiview clustering.

4.6. Convergence Analysis. Note that, since the proposed model is nonconvex, it would be difficult to guarantee the convergence theoretically. Fortunately, experimental findings on the MRSC-v1, NewsGroups, ORL, and Yale datasets with 50% observed paired-samples clearly demonstrate the convergence of the suggested approach. Specifically, we define the goal function as $1/2\|\mathcal{S}^{(t)} - \mathcal{R}(\mathcal{Z}^{(1)}, \mathcal{Z}^{(2)}, \mathcal{Z}^{(3)})$

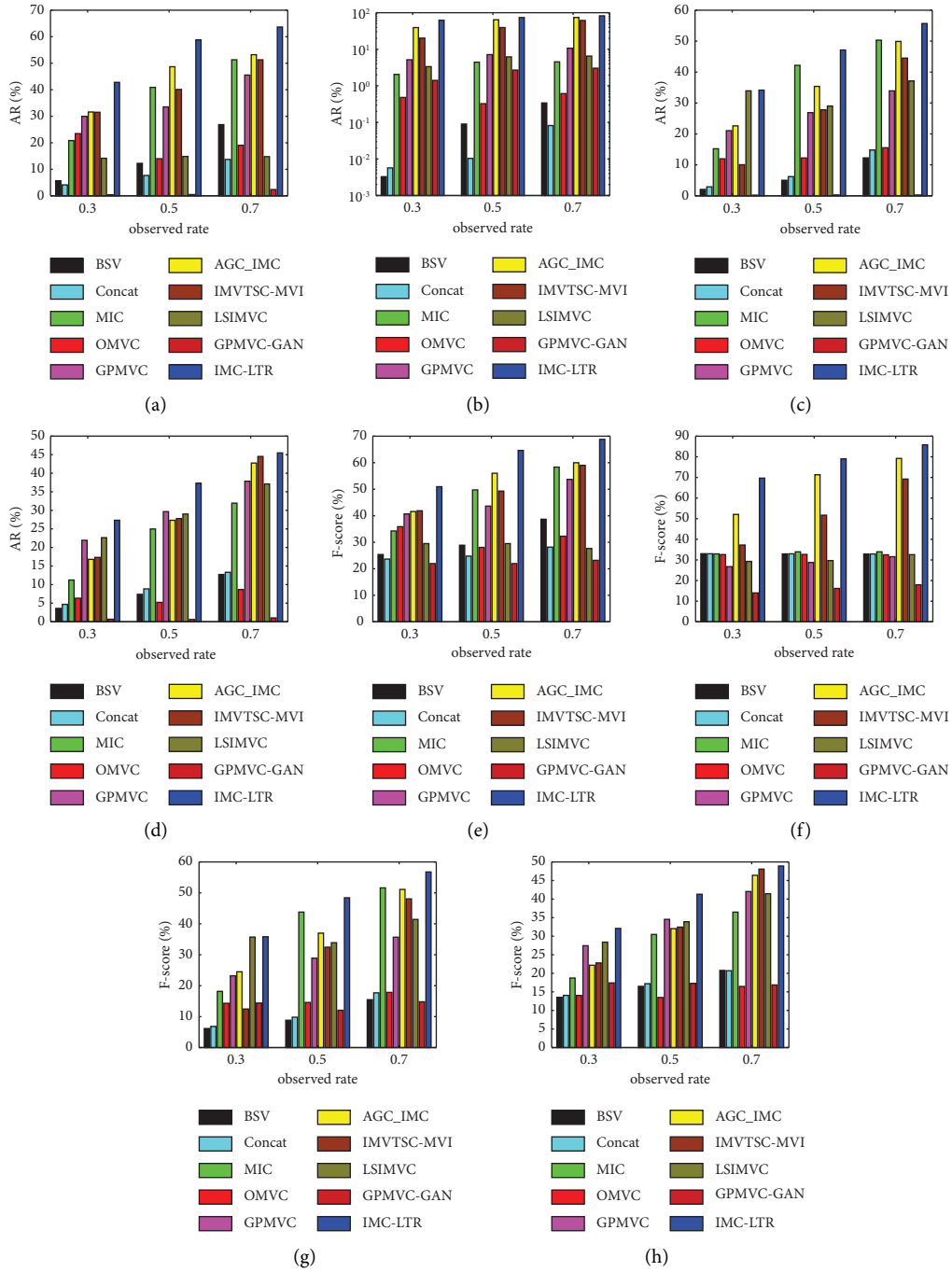


FIGURE 3: ARs (%) and F -score (%) of different methods on the MSRC-v1 dataset, NewsGroups dataset, ORL dataset, and Yale dataset with different observed rates of paired-samples. (a) MSRC-v1. (b) NewsGroups. (c) ORL. (d) Yale. (e) MSRC-v1. (f) NewsGroups. (g) ORL. (h) Yale.

$\|F + 1/2\|\mathcal{S}^{(t)} - \mathcal{S}^{(t-1)}\|_F$, and then record the objective value versus each iteration in Figure 6. According to the convergence curves in Figure 6, which exhibits the strong convergent

quality of the proposed method in multiview clustering, all the curves rapidly decline and then stabilize with an increase in the number of iterations.

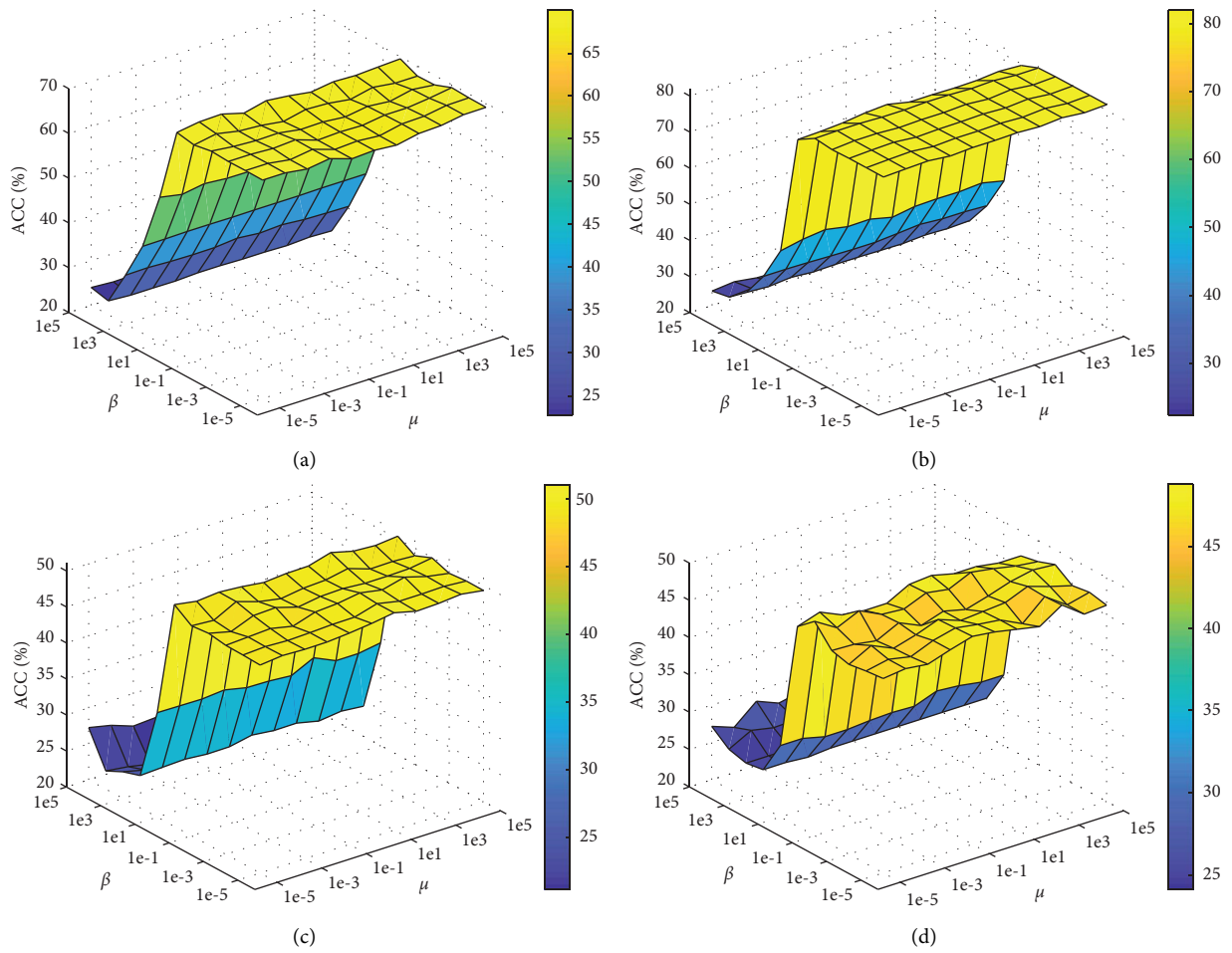


FIGURE 4: ACC (%) versus parameters μ and β of the proposed method on the (a) MSRC-v1 dataset, (b) NewsGroups dataset, (c) ORL dataset, and (d) Yale dataset, with 30% observable paired-sample.

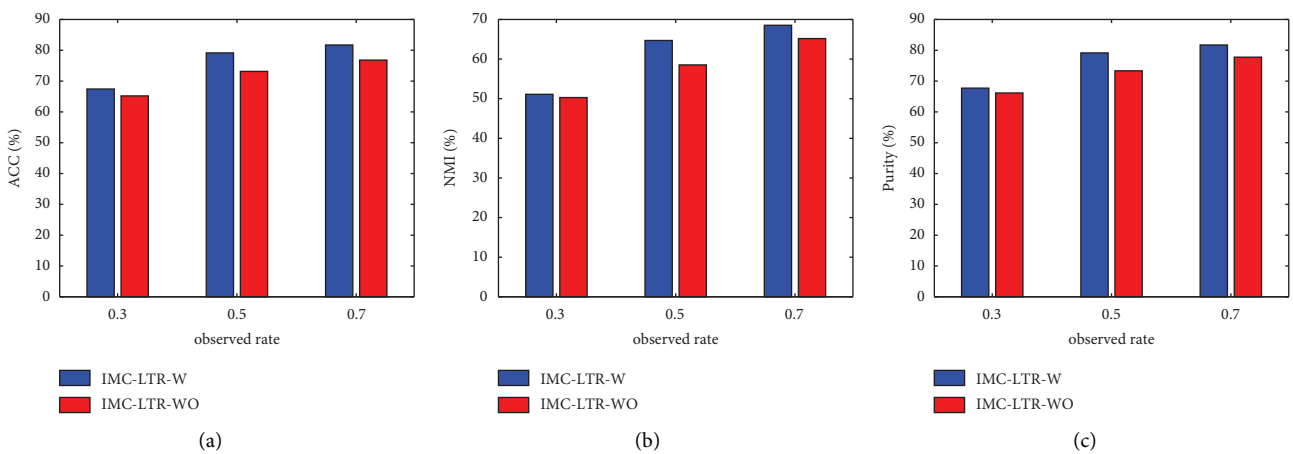


FIGURE 5: Continued.

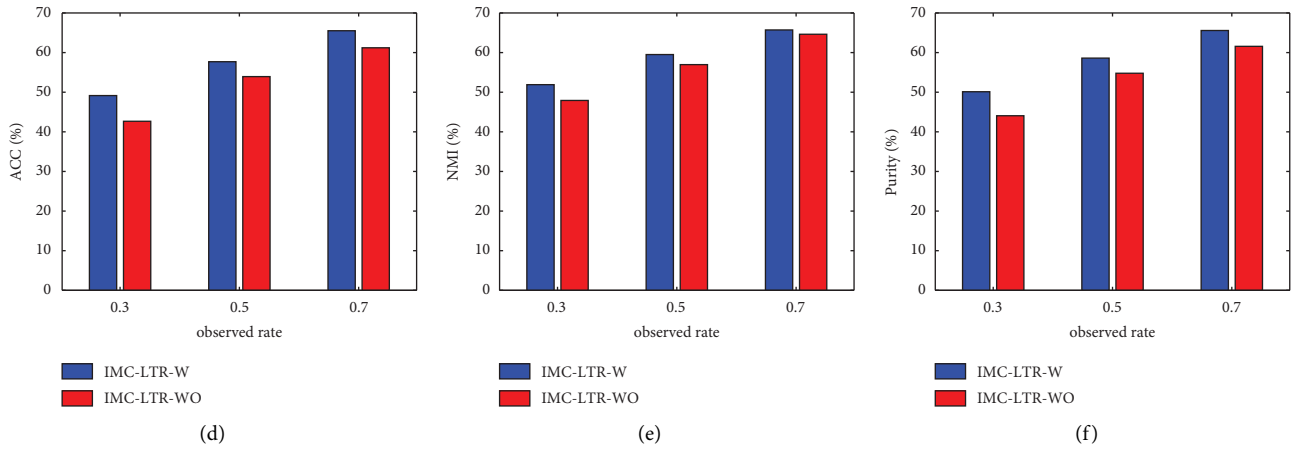


FIGURE 5: ACC (%), NMI (%), and purity (%) with or without adaptive weighting strategy on the MSRC-v1 dataset and Yale dataset with different observed rates of paired-samples, where IMC-LTR-WO denotes the proposed method without adaptive weighting. (a-c) MSRC-v1 and (d-f) Yale.

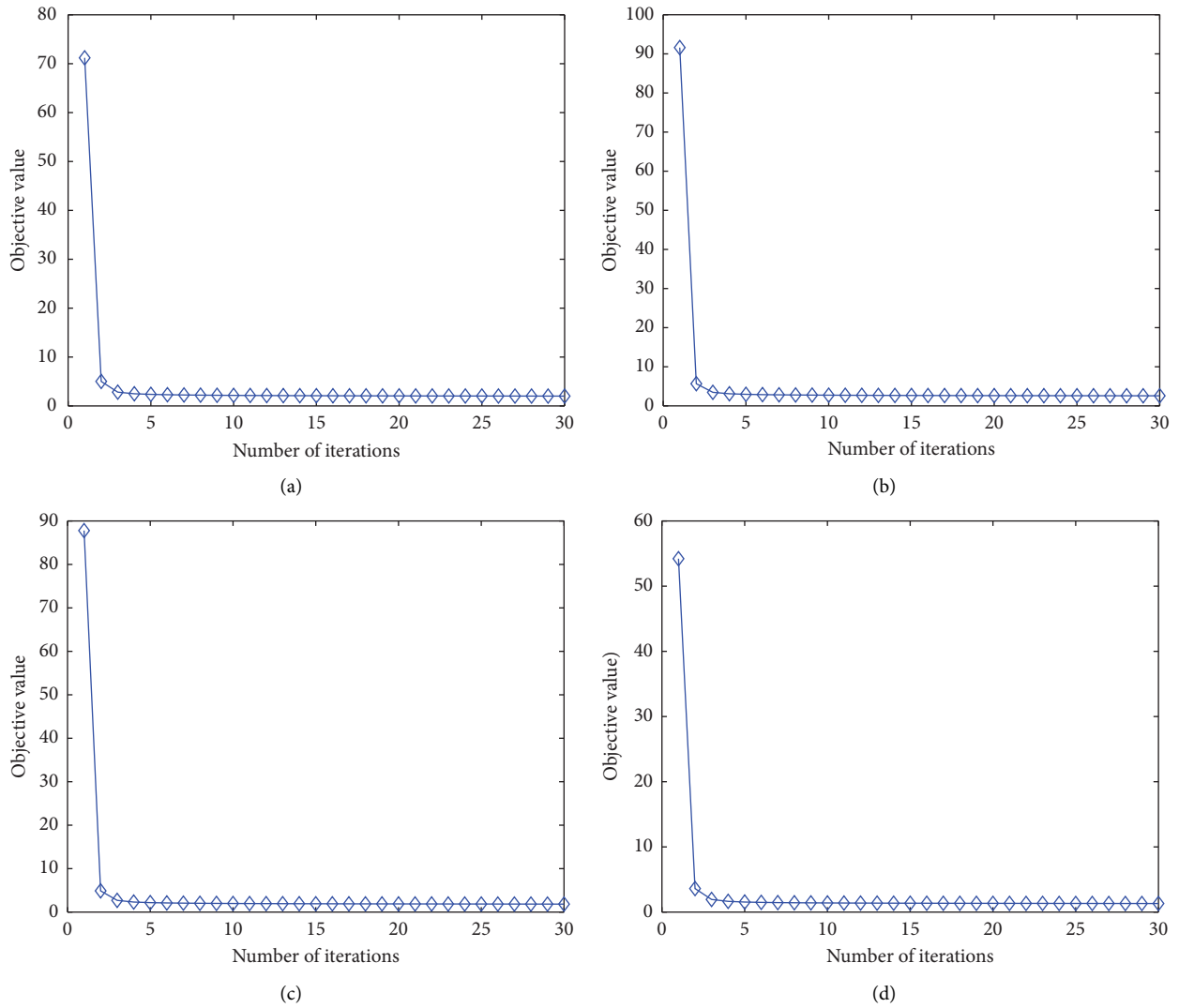


FIGURE 6: The convergence curves of the proposed method on four datasets: (a) MRSC-v1 dataset, (b) NewsGroups dataset, (c) ORL dataset, and (d) Yale dataset with 50% observed rates of paired-samples.

5. Conclusion

This paper established a low-rank tensor ring decomposition-based tensor completion method for IMC. The suggested approach is to first stack the incomplete similarity graphs of all views into a 3rd-order incomplete tensor, and then perform tensor ring decomposition on it for graphs completion. In this way, the incomplete graphs could be completed via the low-rank high-order correlation cross-views. The suggested model explores the high-order correlations by concurrently taking into account sample-level and view-level correlations, and it is shown that the optimum full graphs outperform previous cutting-edge approaches for improving the clustering performance. For further research, we are looking to expand the suggested paradigm into a semisupervised environment.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Nos. 62203128 and 52171331), in part by the Science and Technology Planning Project of Guangzhou City under Grants 202102010411, in part by the Basic and Applied Basic Research of Guangzhou Science and Technology Planning Project (2023A04J1726), and in part by the 2023 Municipal University-Enterprise Joint Funding of Guangzhou Science and Technology Plan Project (2023A03J0124).

References

- [1] X. Yu, H. Liu, Y. Wu, and H. Ruan, "Kernel-based low-rank tensorized multiview spectral clustering," *International Journal of Intelligent Systems*, vol. 36, no. 2, pp. 757–777, 2021.
- [2] H. Lian, H. Xu, S. Wang, M. Li, X. Zhu, and X. Liu, "Partial multiview clustering with locality graph regularization," *International Journal of Intelligent Systems*, vol. 36, no. 6, pp. 2991–3010, 2021.
- [3] A. Sharma, A. Kumar, H. Daume, and D. W. Jacobs, "Generalized multiview analysis: a discriminative latent space," in *Proceedings of the 2012 IEEE conference on computer vision and pattern recognition*, pp. 2160–2167, IEEE, Providence, RI, USA, June 2012.
- [4] J. Fan, Z. Ma, J. Wen, Z. Lai, W. Xie, and W. K. Wong, "Self-inferring incomplete multi-view clustering," *IET Computer Vision*, 2022.
- [5] F. Nie, J. Li, and X. Li, "Self-weighted multiview clustering with multiple graphs," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)*, pp. 2564–2570, Xi'an, China, December 2017.
- [6] J. C. Rojas-Thomas and M. Santos, "New internal clustering validation measure for contiguous arbitrary-shape clusters," *International Journal of Intelligent Systems*, vol. 36, no. 10, pp. 5506–5529, 2021.
- [7] Z. Wu, B. Wang, and C. Li, "A new robust fuzzy clustering framework considering different data weights in different clusters," *Expert Systems with Applications*, vol. 206, Article ID 117728, 2022.
- [8] Z. Liu, W. Jin, and Y. Mu, "Learning robust graph for clustering," *International Journal of Intelligent Systems*, vol. 37, no. 10, pp. 7736–7766, 2022.
- [9] K. Chaudhuri, S. M. Kakade, K. Livescu, and K. Sridharan, "Multi-view clustering via canonical correlation analysis," in *Proceedings of the 26th annual international conference on machine learning*, pp. 129–136, Xi'an, China, December 2009.
- [10] A. Kumar, P. Rai, and H. Daume, "Co-regularized multi-view spectral clustering," *Advances in Neural Information Processing Systems*, vol. 24, pp. 1413–1421, 2011.
- [11] X. Cai, F. Nie, and H. Huang, "Multi-view k-means clustering on big data," in *Proceedings of the Twenty-Third International Joint Conference on artificial intelligence*, August 2013.
- [12] W. Guo, Z. Wang, and W. Du, "Pseudolabel-guided multiview consensus graph learning for semisupervised classification," *International Journal of Intelligent Systems*, vol. 37, no. 11, pp. 8611–8634, 2022.
- [13] M. M. Kalayeh, H. Idrees, and M. Shah, "Nmf-knn: image annotation using weighted multi-view non-negative matrix factorization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 184–191, Columbus, OH, USA, June 2014.
- [14] H. Zhao, Z. Ding, and Y. Fu, "Multi-view clustering via deep matrix factorization," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, Washington, DC, USA, January 2017.
- [15] S. Huang, Z. Kang, and Z. Xu, "Auto-weighted multi-view clustering via deep matrix decomposition," *Pattern Recognition*, vol. 97, Article ID 107015, 2020.
- [16] H. Huang, G. Zhou, N. Liang, Q. Zhao, and S. Xie, "Diverse deep matrix factorization with hypergraph regularization for multiview data representation," *IEEE/CAA Journal of Automatica Sinica*, vol. 34, pp. 1–14, 2022.
- [17] H. Huang, Z. Yang, Z. Li, and W. Sun, "A converged deep graph semi-nmf algorithm for learning data representation," *Circuits, Systems, and Signal Processing*, vol. 41, no. 2, pp. 1146–1165, 2022.
- [18] P. Rai, A. Trivedi, H. Daumé III, and S. L. DuVall, "Multiview clustering with incomplete views," in *Proceedings of the NIPS Workshop on Machine Learning for Social Computing, Cite-seer*, Washington, DC, USA, December 2010.
- [19] J. Wen, Z. Zhang, Z. Zhang, L. Fei, and M. Wang, "Generalized incomplete multiview clustering with flexible locality structure diffusion," *IEEE Transactions on Cybernetics*, vol. 51, no. 1, pp. 101–114, 2021.
- [20] S.-Y. Li, Y. Jiang, and Z.-H. Zhou, "Partial multi-view clustering," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 28, no. 1, 2014.
- [21] H. Zhao, H. Liu, and Y. Fu, "Incomplete multi-modal visual data grouping," in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)*, pp. 2392–2398, Boston, CA, USA, June 2016.
- [22] N. Xu, Y. Guo, X. Zheng, Q. Wang, and X. Luo, "Partial multi-view subspace clustering," in *Proceedings of the 26th ACM International conference on multimedia*, pp. 1794–1801, New York, NY, USA, October 2018.
- [23] W. Shao, L. He, and S. Y. Philip, "Multiple incomplete views clustering via weighted nonnegative matrix factorization with

- $l_{2,1}$ regularization,” in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 318–334, Springer, Heidelberg, Germany, 2015.
- [24] W. Shao, L. He, C.-T. Lu, and S. Y. Philip, “Online multi-view clustering with incomplete views,” in *Proceedings of the 2016 IEEE International Conference on Big Data (Big Data)*, pp. 1012–1017, IEEE, Washington, DC, USA, December 2016.
- [25] N. Rai, S. Negi, S. Chaudhury, and O. Deshmukh, “Partial multi-view clustering using graph regularized nmf,” in *Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR)*, pp. 2192–2197, IEEE, Cancun, Mexico, April 2016.
- [26] M. Hu and S. Chen, “Doubly aligned incomplete multi-view clustering,” 2019, <https://arxiv.org/abs/1903.02785>.
- [27] X. Zheng, X. Liu, J. Chen, and E. Zhu, “Adaptive partial graph learning and fusion for incomplete multi-view clustering,” *International Journal of Intelligent Systems*, vol. 37, no. 1, pp. 991–1009, 2022.
- [28] H. Wang, L. Zong, B. Liu, Y. Yang, and W. Zhou, “Spectral perturbation meets incomplete multi-view data,” 2019, <https://arxiv.org/abs/1906.00098>.
- [29] Z. Yu, M. Ye, S. Xiao, and L. Tian, “Learning missing instances in latent space for incomplete multi-view clustering,” *Knowledge-Based Systems*, vol. 250, Article ID 109122, 2022.
- [30] L. Zhao, Z. Chen, Y. Yang, Z. Jane Wang, and V. C. Leung, “Incomplete multi-view clustering via deep semantic mapping,” *Neurocomputing*, vol. 275, pp. 1053–1062, 2018.
- [31] J. Yu, Q. Duan, H. Huang, S. He, and T. Zou, “Effective incomplete multi-view clustering via low-rank graph tensor completion,” *Mathematics*, vol. 11, no. 3, p. 652, 2023.
- [32] X. Liu, X. Zhu, M. Li et al., “Efficient and effective incomplete multi-view clustering,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 1, pp. 4392–4399, 2019.
- [33] J. Wen, Z. Zhang, Y. Xu, B. Zhang, L. Fei, and H. Liu, “Unified embedding alignment with missing views inferring for incomplete multi-view clustering,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 1, pp. 5393–5400, 2019.
- [34] J. Wen, K. Yan, Z. Zhang et al., “Adaptive graph completion based incomplete multi-view clustering,” *IEEE Transactions on Multimedia*, vol. 23, pp. 2493–2504, 2021.
- [35] C. Shang, A. Palmer, J. Sun, K.-S. Chen, J. Lu, and J. Bi, “Vigan: missing view imputation with generative adversarial networks,” in *Proceedings of the 2017 IEEE International Conference on Big Data (Big Data)*, pp. 766–775, IEEE, Osaka, Japan, August 2017.
- [36] Q. Wang, Z. Ding, Z. Tao, Q. Gao, and Y. Fu, “Partial multi-view clustering via consistent gan,” in *Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM)*, pp. 1290–1295, IEEE, Singapore, November 2018.
- [37] Q. Wang, Z. Ding, Z. Tao, Q. Gao, and Y. Fu, “Generative partial multi-view clustering with adaptive fusion and cycle consistency,” *IEEE Transactions on Image Processing*, vol. 30, pp. 1771–1783, 2021.
- [38] Q. Zhao, G. Zhou, S. Xie, L. Zhang, and A. Cichocki, “Tensor ring decomposition,” 2016, <http://arxiv.org/abs/1606.05535>.
- [39] Q. Zhao, M. Sugiyama, L. Yuan, and A. Cichocki, “Learning efficient tensor representations with ring-structured networks,” in *Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8608–8612, IEEE, Brighton, UK, May 2019.
- [40] W. Wang, V. Aggarwal, and S. Aeron, “Efficient low rank tensor ring completion,” in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 5698–5706, Venice, Italy, May 2017.
- [41] L. Yuan, C. Li, D. Mandic, J. Cao, and Q. Zhao, “Tensor ring decomposition with rank minimization on latent space: an efficient approach for tensor completion,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 1, pp. 9151–9158, 2019.
- [42] J. Yu, G. Zhou, C. Li, Q. Zhao, and S. Xie, “Low tensor-ring rank completion by parallel matrix factorization,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 7, pp. 3020–3033, 2021.
- [43] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.
- [44] L. Hagen and A. B. Kahng, “New spectral methods for ratio cut partitioning and clustering,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 11, no. 9, pp. 1074–1085, 1992.
- [45] J. Shi and J. Malik, “Normalized cuts and image segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.
- [46] J. Yu, G. Zhou, W. Sun, and S. Xie, “Robust to rank selection: low-rank sparse tensor-ring completion,” *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2021.
- [47] F. Nie, X. Wang, M. Jordan, and H. Huang, “The constrained laplacian rank algorithm for graph-based clustering,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, 2016.
- [48] T. G. Wright and L. N. Trefethen, “Large-scale computation of pseudospectra using arpack and eigs,” *SIAM Journal on Scientific Computing*, vol. 23, no. 2, pp. 591–605, 2001.
- [49] F. Nie, G. Cai, J. Li, and X. Li, “Auto-weighted multiview learning for image clustering and semi-supervised classification,” *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1501–1511, 2018.
- [50] J. Wen, Z. Zhang, Z. Zhang et al., “Unified tensor framework for incomplete multiview clustering and missing-view inferring,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 11, pp. 10273–10281, 2021.
- [51] C. Liu, Z. Wu, J. Wen, Y. Xu, and C. Huang, “Localized sparse incomplete multiview clustering,” *IEEE Transactions on Multimedia*, pp. 1–13, 2022.
- [52] C. Zhang, H. Fu, S. Liu, G. Liu, and X. Cao, “Low-rank tensor constrained multiview subspace clustering,” in *Proceedings of the IEEE international conference on computer vision*, pp. 1582–1590, Santiago, Chile, December 2015.
- [53] H. Schütze, C. D. Manning, and P. Raghavan, *Introduction to Information Retrieval*, vol. 39, Cambridge University Press, Cambridge, UK, 2008.
- [54] L. Hubert and P. Arabie, “Comparing partitions,” *Journal of Classification*, vol. 2, no. 1, pp. 193–218, 1985.