Tensor Completion-Based Incomplete Multiview Clustering

Wei Xia, *Graduate Student Member, IEEE*, Quanxue Gao, Qianqian Wang, and Xinbo Gao, *Senior Member, IEEE*

Abstract—Incomplete multiview clustering is a challenging problem in the domain of unsupervised learning. However, the existing incomplete multiview clustering methods only consider the similarity structure of intraview while neglecting the similarity structure of interview. Thus, they cannot take advantage of both the complementary information and spatial structure embedded in similarity matrices of different views. To this end, we complete the incomplete graph with missing data referring to tensor complete and present a novel and effective model to handel the incomplete multiview clustering task. To be specific, we consider the similarity of the interview graphs via the tensor Schatten p-norm-based completion technique to make use of both the complementary information and spatial structure. Meanwhile, we employ the connectivity constraint for similarity matrices of different views such that the connected components approximately represent clusters. Thus, the learned entire graph not only has the low-rank structure but also well characterizes the relationship between unmissing data. Extensive experiments show the promising performance of the proposed method comparing with several incomplete multiview approaches in the clustering tasks.

Index Terms—Incomplete data, multiview clustering, tensor completion.

I. INTRODUCTION

MULTIVIEW data are ubiquitous in real scenarios, and provide more useful complementary and discriminative information embedded in multiple views, which help improve the robustness of algorithms [1]–[7]. To date, numerous effective clustering methods have been developed and achieve promising clustering performance. However, all clustering

methods have an assumption that the entries of the given data are fully observed from multiple views. This assumption may not satisfy since some partial views are missing in real applications, for example, disease diagnosing and Webpage clustering. This makes the performance of the existing multiview clustering approaches degenerate remarkably. Thus, there has been a growing interest for incomplete multiview learning [8]–[11].

The graph-based incomplete multiview clustering (GIMC) has attracted much attention due to the fact that it can reveal the relationships between data and complex distribution of data [12]. One of the most representative methods is [8]. It employs the incomplete graphs from partial complete views, to learn the Laplacian matrix of the entire data including missing views, and then learns low-dimensional embedding by kernel CCA. However, it cannot well exploit the complementary information and the low-rank structure embedded in graphs of different views, which are very important for multiview clustering [13], [14]. Another limitation is that it requires at least one complete view. To this end, combined with adaptive graph learning, Wen *et al.* [15] presented a new incomplete multiview spectral clustering (SC) method (IMSC-AGL). It makes use of the low-rank representation technique to explore the low-rank structure embedded in each graph, that is, constructed from the corresponding incomplete view.

Although IMSC-AGL obtains satisfactory results, it still has the following shortcomings.

1) The learned graph cannot well exploit the cluster structure of entire data including missing views. For each view, IMSC-AGL only considers the relationship between the un-missing data. Thus, the learned graph cannot characterize the cluster structure.

2) It ignores the similarity structure of interview. IMSC-AGL only considers the similarity structure of unmissing data of intraview. Thus, the learned graph cannot well explore the complementary information embedded in multiple views.

3) It implicitly assumes that each view contributes equally to the clustering task, which makes no sense in real applications. Each view has some contents of the objects that other views do not contain; thus, there has a significance difference between different views for clustering. However, IMSC-AGL ignores this fact, resulting in degrading the robustness and flexibility of the algorithm.

In this article, inspired from the low-rank structure of the graph, we propose a novel and effective multiview clustering

method for incomplete data, in which we complete the incomplete graph with missing data referring to the tensor complete technique. Our method is the first work that employs the tensor completion technique to learn the view-consensus graph. Specifically, we employ the tensor Schatten p-norm [6], [16] based complete technique to learn the consensus graph, which helps to take advantage of both the spatial structure and complementary information embedded in similarity matrices of different views. This can guarantee that the learned entire graph not only characterizes the similarity structure of interview but also well characterizes the similarity structure of interview. After that, the connectivity constraint is employed on the learned graph to ensure that the connected components approximately indicate clusters, which also helps guide the tensor completion. By alternate optimization, the two processes are seamlessly connected to achieve better clustering performance. To sum up, our contributions are three-fold.

1) We leverage a tensor Schatten p-norm-based complete technique to construct the graph of entire data including missing views. Thus, the learned graph not only characterizes the similarity structure of interview but also preserves the relationship between the un-missing data.

2) We employ the connectivity constraint for similarity matrices of different views. Therefore, the connected components approximately indicate clusters, which helps to characterize the cluster structure of multiview data.

3) The proposed method explicitly considers the difference between different views by adaptive weighting scheme. This helps encode the discriminant information embedded in graph. Extensive experiments show the promising performance of the proposed method comparing with some state-of-the-art incomplete multiview approaches in the clustering tasks.

Notations: The matrices and vectors are represented by bold uppercase letters and bold lowercase letters, respectively. For example, $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\mathbf{Z} \in \mathbb{R}^{n_1 \times n_2}$, and $\mathbf{z} \in \mathbb{R}^n$ are the corresponding 3-order tensor, matrix, and n-dimensional vector, respectively. Moreover, we use $\overline{\mathcal{Z}} = fft(\mathcal{Z}, [], 3)$ to represent the discrete Fast Fourier transform (FFT) of $\mathcal{Z}$, and use $\mathcal{Z} = ifft(\overline{\mathcal{Z}}, [], 3)$ to represent the inverse FFT of $\overline{\mathcal{Z}}$ along the 3rd dimension. $tr(\mathbf{Z})$ is the trace of matrix $\mathbf{Z}$. $\mathbf{I}$ is an identity matrix. $\mathbf{1}$ is a vector whose elements are 1.

II. RELATED WORK

Multiview clustering targets at dividing multiview data into several different clusters, such that the samples in the same cluster have high correlation to each other [14], [17]–[19]. In real word applications, only incomplete fractions of the multiview data can be obtained [20], [21]. This makes the performance of most methods degrade remarkably. Taking this into consideration, there has been a growing interest for incomplete multiview clustering, which can be roughly divided three categories, that is: 1) matrix factorization-based methods (MFIMC); 2) kernel learning-based methods (KIMC); and 3) graph-based methods (GIMC). MFIMC aims to learn a low-dimension common representation for different views by

the matrix factorization approach. Partial multiview clustering (PMVC) [9] is one of the most representative MFIMC methods. For the different views of a sample, PMVC enforces them to have the same representation and finally learns a consensus latent subspace. Different from PMVC, Shao *et al.* [10] first employed the data preprocessing technique to complete the missing views, and then introduced the weighted non-negative matrix factorization to simultaneously learn the view-specific latent representations and the view-consensus representation. However, the aforementioned methods neglect the local geometric structure embedded in data when learning the consensus representation of data. Thus, they cannot obtain the compact representation, which well encodes discriminative information. To tackle this problem, Wen *et al.* [22] presented a generalized incomplete multiview learning method, which well exploits the local geometrical structure.

KIMC integrates the kernel matrix imputation task and clustering task into a unified optimization framework and alternatively solves each of them. For example, Liu *et al.* [23] proposed late fusion incomplete multiview clustering (LF-IMVC). LF-IMVC first learns a view-consensus clustering matrix, and then imputes the incomplete information of them with the learned consensus matrix. Although achieved good clustering results, KIMC methods are sensitive to the quality of the preconstructed kernels.

Different from MFIMC and KIMC methods, GIMC, which can reveal the relationships between data and complex distribution of data, has become an active topic for incomplete multiview learning. It aims to learn the latent feature from different graphs that exploit the relationships between samples. The key challenge is how to effectively learn a reasonable entire similarity graph by exploiting the information implicit in incomplete multiview data. To solve this problem, Trivedi *et al.* [8] proposed to use the Laplace matrix of the complete view to fill in the incomplete the graph of the view with missing samples. Zhao *et al.* [24] leveraged the idea of PMVC to learn the latent consensus representation and then learned the entire graph from the latent representation for incomplete multiview clustering. However, all of them require that few samples contain all views. This reduces the flexibility of algorithms. For each view of multiview data, Zhou *et al.* [25] proposed to construct a complete graph and learned a view-consensus graph automatically by using each constructed view-specific graph. Wang *et al.* [26] explored a spectral perturbation theory and transferred the problem of missing views from data level to a similar graph level, and presented a matrix completion method for incomplete graphs. However, they cannot exploit the low-rank structure embedded in the graph, which is very important for clustering. To this end, Wen *et al.* [15] proposed IMSC-AGL, which simultaneously carries out graph construction and consensus representation learning via low-rank constraint. However, it ignores the similarity structure of interview, resulting in inferior results.

Recently, deep learning-based IMC approaches [21], [27]–[31] keep emerging. The basic idea of these approaches is to leverage the remained information in the complete views to predict the missing samples. For example,

Wen *et al.* [27] developed a weighted fusion layer with the view-indicator matrix to cope with the incomplete issue on multiview clustering, where the filled missing views plays no role in framework optimization. Lin *et al.* [29] integrated view-consistency learning and view completing into a unified framework and proposed incomplete multiview clustering via contrastive prediction (COMPLETER). COMPLETER aims to recover one view from another one by minimizing the conditional entropy between view-specific representation. Although these methods achieve impressive clustering results, they fail to take both the spatial structure and complementary information embedded in multiple views into consideration.

To sum up, our proposed method is significantly different from existing incomplete multiview clustering methods in the following aspects.

1) To the best of our knowledge, this could be the first tensor completion-based incomplete multiview clustering method. Our proposed method takes the similarity of the interview graphs into account via the tensor Schatten p-norm-based complete technique to learn the consensus graph. Thus, the learned entire graph not only well characterizes the complementary information and spatial structure embedded in multiview data, but also well characterizes the relationship between the unmissing data, whereas these existing incomplete multiview clustering methods do not consider the similarity structure of interview graphs.

2) Meanwhile, to well characterize cluster structure, we employ the connectivity constraint for similarity matrices. In contrast, existing methods fail to take this into consideration.

3) Moreover, our method takes the differences among multiple views into account, which helps encode the discriminant information embedded in multiple graphs, whereas most existing methods treat all views equally.

III. METHODOLOGY

A. Problem Formulation

Denote the multiview data by $\{\mathbf{X}^{(v)}\}_{v=1}^{V}$, $\{\mathbf{Y}^{(v)}\}_{v=1}^{V}$ is the set of the corresponding unmissing instances, where $\mathbf{X}^{(v)} \in \mathbb{R}^{d_v \times N}$; $\mathbf{Y}^{(v)} \in \mathbb{R}^{m_v \times N_u}$; d_v and N are the number of dimensions and data points in the vth view, respectively; V is the view number, N_u and m_v are the number of unmissing samples and the feature dimensions of the vth view, respectively. The objective function of incomplete multiview SC with adaptive graph learning [15] is

$$
\begin{aligned}
\min_{\mathbf{D}^{(v)}, \mathbf{E}^{(v)}, \mathbf{F}} \quad & \sum_{v=1}^{V} \left(\left\| \mathbf{D}^{(v)} \right\|_{*} + \lambda_2 \left\| \mathbf{E}^{(v)} \right\|_{1} \right) \\
& \sum_{v=1}^{V} \left(\lambda_1 \mathrm{tr}\left(\mathbf{F}^{(v)\mathrm{T}} \mathbf{G}^{(v)\mathrm{T}} \mathbf{L}^{(v)} \mathbf{G}^{(v)} \mathbf{F}^{(v)} \right) \right) + \frac{\lambda_3}{2} \Gamma\left(\mathbf{F}^{(v)}, \mathbf{U} \right) \\
\text{s.t.} \quad & \mathbf{Y}^{(v)} = \mathbf{Y}^{(v)} \mathbf{D}^{(v)} + \mathbf{E}^{(v)}, \mathbf{D}^{(v)} \mathbf{1} = \mathbf{1} \\
& 0 \leq \mathbf{D}^{(v)}, \mathbf{D}^{(v)}_{i,i} = 0, \mathbf{F}^{(v)} \left(\mathbf{F}^{(v)} \right)^{\mathrm{T}} = \mathbf{I}, \mathbf{U}^{\mathrm{T}} \mathbf{U} = \mathbf{I} \quad (1)
\end{aligned}
$$

where $\mathbf{D}^{(v)} \in \mathbb{R}^{N_u \times N_u}$ is the self-expression coefficient matrix learned from the unmissing instances; $\bar{\mathbf{D}}^{(v)} = \mathbf{G}^{(v)\mathrm{T}} \mathbf{D}^{(v)} \mathbf{G}^{(v)}$ is the completed graph; the Laplacian matrix of $\mathbf{D}^{(v)}$ is $\mathbf{L}^{(v)}$, and we have $\bar{\mathbf{L}}^{(v)} = \mathbf{G}^{(v)\mathrm{T}} \mathbf{L}^{(v)} \mathbf{G}^{(v)}$; $\mathbf{F}^{(v)}$ and $\mathbf{U}$ denote the cluster indicator matrix of the vth view and a consensus cluster indicator matrix, respectively. $\lambda_i (i = 1, 2, 3)$ are penalty parameters. $\Gamma(\cdot)$ is used to measure the disagreement of $\mathbf{F}^{(v)}$ and the consensus cluster indicator matrix $\mathbf{U}$. $\mathbf{G}^{(v)} \in \mathbb{R}^{N_u \times N}$ is an index matrix used to complete the graph, which is defined as

$$
\mathbf{G}_{ij}^{(v)} = \begin{cases} 1, & \text{if } \mathbf{y}_i^{(v)} = \mathbf{x}_j^{(v)} \\ 0, & \text{otherwise} \end{cases} \quad (2)
$$

where $\mathbf{y}_i^{(v)} \in \mathbb{R}^{m_v}$ is the ith unmissing instance in the vth view and $\mathbf{x}_j^{(v)} \in \mathbb{R}^{d_v}$ is the jth instance in the vth view.

Although IMSC-AGL achieves good clustering performance, especially for the case in which all samples have missing views, it still has the following drawbacks. *First*, in model (1), for each view, IMSC-AGL only considers the relationship between the unmissing data by nuclear norm minimization. Thus, the learned entire graph $\bar{\mathbf{D}}^{(v)}$ does not characterize the similarity structure of interview graphs. For multiview clustering, similarity graphs of different views should have not only high similarity but also a similar spatial structure. Unfortunately, IMSC-AGL does not take into account this. In other words, it ignores the similarity structure of interview graphs, which are very important for improving clustering performance. Thus, the learned graph cannot make full use of the complementary information embedded in multiple views. *Second*, the connected components in the learned graph does not approximately indicate the cluster. Thus, the learned graph cannot exploit the cluster structure of data. *Third*, from both the first term and last term in model (1), it is easy to see that IMSC-AGL implicitly treats each view equally, which makes no sense in real application. For multiview data, the features of different view include some content of the objects that other views do not contain; thus, there has a big difference in clustering performance between different views. However, IMSC-AGL ignores this fact, resulting in degrading the robustness and flexibility. To learn the entire graph, which well exploits the complementary information embedded in different views and characterizes the cluster structure, we employ a tensor complete technique with the connectivity constraint to address the aforementioned issues.

B. Objective Function

The multiview clustering aims to divide the data into K clusters, an ideal similarity graph should have low-rank structure and K-connected components. Moreover, to obtain the best clustering performance, similarity graphs between different views not only have high similarity, but also the spatial geometric structure between them is similar. For incomplete clustering, we also hope that the learned graph well characterizes the relationship between the unmissing data. Drawing the inspiration from the tensor complete technique, the proposed
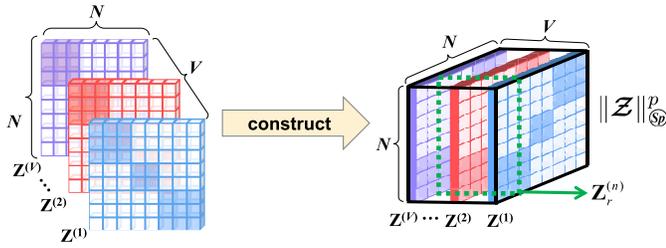
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

4   IEEE TRANSACTIONS ON CYBERNETICS

Fig. 1. Construction of tensor $\mathcal{Z} \in \mathbb{R}^{N \times V \times N}$. $\mathbf{Z}_r^{(n)} \in \mathbb{R}^{N \times V}$ denotes the nth frontal slice of $\mathcal{Z}$ ($n \in \{1, 2, \ldots, N\}$).

objective function is

$$\min_{\mathbf{Z}^{(v)}} \ \|\mathcal{Z}\|_{\textcircled{Sp}}^{p}$$

$$\text{s.t.} \ \ C_{\Omega}(\mathcal{Z}) = C_{\Omega}(\bar{\mathcal{S}}), \ 0 \leq \mathbf{Z}^{(v)}, \ \mathbf{Z}^{(v)} \cdot \mathbf{1} = \mathbf{1} \tag{3}$$

where $\mathcal{Z} \in \mathbb{R}^{N \times V \times N}$ and $\mathcal{Z}(:, v, :) = \mathbf{Z}^{(v)}$, $\bar{\mathbf{S}}^{(v)} = \mathbf{G}^{(v)\mathbf{T}}\mathbf{S}^{(v)}\mathbf{G}^{(v)}$, $\mathbf{S}^{(v)} \in \mathbb{R}^{N_u \times N_u}$ is the similarity matrix of the vth view, $\| \bullet \|_{\textcircled{Sp}}$ is the tensor Schatten p-norm (see Definition 1), Ω is the index of unmissing data in original data, and C_{Ω} denotes a completion operator which can be defined as

$$C_{\Omega}(\mathcal{Z})_{i,j,k} = \begin{cases} \mathcal{Z}_{i,j,k}, & (i, j, k) \in \Omega \\ 0, & (i, j, k) \notin \Omega. \end{cases} \tag{4}$$

Remark 1: The tensor Schatten p-norm in our objective (3) is used to explore the complementary information embedded in interview graphs $\mathbf{Z}^{(v)}$. For tensor $\mathcal{Z}$, as shown in Fig. 1, the nth frontal slice $\mathbf{Z}_r^{(n)} \in \mathbb{R}^{N \times V}$ describes the similarity between N sample points in different views. The good graph $\mathbf{Z}^{(v)}$ should ensure that the relationship between N data points is consistent in different views. Considering the fact that different views usually show different cluster structures, we impose a tensor Schatten p-norm minimization [16] based completion technique, that is, the tensor multirank minimization constraint on $\mathcal{Z}$, which can make sure each $\mathbf{Z}_r^{(n)}$ has spatial low-rank structure. After it happened, $\mathbf{Z}_r^{(n)}$ can well characterize the complementary information embedded in interview.

Definition 1 [16]: Given $\mathcal{Z} \in \mathbb{R}^{N \times V \times N}$, $h = \min(N, V)$, tensor Schatten p-norm of $\mathcal{Z}$ is defined as

$$\|\mathcal{Z}\|_{\textcircled{Sp}} = \left(\sum_{i=1}^{N} \left\| \overline{\mathcal{Z}}^{(i)} \right\|_{\textcircled{Sp}}^{p} \right)^{\frac{1}{p}} = \left(\sum_{i=1}^{N} \sum_{j=1}^{h} \sigma_j \left(\overline{\mathcal{Z}}^{(i)} \right)^{p} \right)^{\frac{1}{p}} \tag{5}$$

where $p \in (0, 1]$ is a parameter of power and $\sigma_j(\overline{\mathcal{Z}}^{(i)})$ is the jth singular value of $\overline{\mathcal{Z}}^{(i)}$.

Remark 2 Note that when $p = 1$, the tensor Schatten p-norm of $\mathcal{Z}$ is the tensor nuclear norm [32]: $\|\mathcal{Z}\|_{\circledast} = \sum_{i=1}^{N} \sum_{j=1}^{h} \sigma_j(\overline{\mathcal{Z}}^{(i)})$. Let us use matrices to illustrate the Schatten p-norm. Consider $\mathbf{Z}_r \in \mathbb{R}^{N \times V}$, $\sigma_1, \ldots, \sigma_h$ is the singular values of $\mathbf{Z}_r$ in the descending order. Then, for $p > 0$, we may consider $\|\mathbf{Z}_r\|_{Sp}^{p} = \sigma_1^p + \cdots + \sigma_h^p$. If we let $p \to 0$, one can see $\lim_{p \to 0} \|\mathbf{Z}_r\|_{Sp}^{p} = \#\{i : \sigma_i \neq 0\} = \text{rank}(\mathbf{Z}_r)$. Hence, for $0 \leq p \leq 1$, the Schatten p-norm (which is a quasinorm) is introduced for the rank approximation.

From model (3), the learned $\mathbf{Z}^{(v)}$ well characterize both spatial low-rank structure and the relationship between the

unmissing data of the vth view. However, the learned $\mathbf{Z}^{(v)}$ does not satisfy the K-connected components. To achieve the ideal similarity graph, the elements Z_{ij} in model (3) should be constrained such that the completed graph has approximately K-connected components which characterize the cluster structure. Inspired by Lemma 1, we use constraint $\text{rank}(\mathbf{L}^{(v)}) = n - K$ instead of the connected constraint and obtain

$$\min_{0 \leq \mathbf{Z}^{(v)}, \ \mathbf{Z}^{(v)} \cdot \mathbf{1} = \mathbf{1}} \ \|\mathcal{Z}\|_{\textcircled{Sp}}^{p}$$

$$\text{s.t.} \ \ C_{\Omega}(\mathcal{Z}) = C_{\Omega}(\bar{\mathcal{S}}), \text{rank}(\mathbf{L}^{(v)}) = n - K. \tag{6}$$

Lemma 1 [33]: The multiplicity K of the eigenvalue zero of the Laplacian matrix $\mathbf{L_Z}$ (non-negative) is equal to the number of connected components in the graph with the similarity matrix $\mathbf{Z}$.

Denote $\sigma_j(\mathbf{L}^{(v)})$ as the jth smallest eigenvalue of $(\mathbf{L}^{(v)})$. Note that $\sigma_j(\mathbf{L}^{(v)}) \geq 0$ because $(\mathbf{L}^{(v)})$ is positive semidefinite. According to Ky Fan's theorem [34], the optimal solution of $\mathbf{Z}$ with the rank constraint can be achieved by solving

$$\min_{\mathbf{F}^{\mathbf{T}}\mathbf{F} = \mathbf{I}} \ \text{tr}(\mathbf{F}^{\mathbf{T}}\mathbf{L}^{(v)}\mathbf{F}). \tag{7}$$

Thus, the proposed model (6) can be rewritten as

$$\min_{\mathbf{Z}^{(v)}, \mathbf{F}} \ \|\mathcal{Z}\|_{\textcircled{Sp}}^{p} + \beta \sum_{v=1}^{V} \text{tr}\left(\mathbf{F}^{\mathbf{T}}\mathbf{L}_{\mathbf{Z}^{(v)}}\mathbf{F} \right)$$

$$\text{s.t.} \ \ C_{\Omega}\left(\mathbf{Z}^{(v)} \right) = C_{\Omega}\left(\bar{\mathbf{S}}^{(v)} \right), \mathbf{F}^T\mathbf{F} = \mathbf{I}$$

$$0 \leq \mathbf{Z}^{(v)}, \mathbf{Z}^{(v)}\mathbf{1} = \mathbf{1}. \tag{8}$$

It can be seen that the model (8) implicitly assumes that different views contribute equally to the clustering, resulting in reducing the flexibility of the method. To further improve the robustness, we leverage the following adaptively weighting strategy to encode difference and propose the final objective function as:

$$\min_{\mathbf{Z}^{(v)}, \alpha^{(v)}, \mathbf{F}} \ \|\mathcal{Z}\|_{\textcircled{Sp}}^{p} + \beta \sum_{v=1}^{V} \text{tr}\left(\frac{1}{\alpha^{(v)}} \mathbf{F}^{\mathbf{T}}\mathbf{L}_{\mathbf{Z}^{(v)}}\mathbf{F} \right)$$

$$\text{s.t.} \ \ C_{\Omega}\left(\mathbf{Z}^{(v)} \right) = C_{\Omega}\left(\bar{\mathbf{S}}^{(v)} \right), \mathbf{F}^{\mathbf{T}}\mathbf{F} = \mathbf{I}$$

$$0 \leq \mathbf{Z}^{(v)}, \mathbf{Z}^{(v)}\mathbf{1} = \mathbf{1}, \sum_{v=1}^{V} \alpha^{(v)} = 1, \alpha^{(v)} \geq 0 \tag{9}$$

where the non-negative $\alpha^{(v)}$ represents the normalized weighting value of the vth view.

C. Optimization

Inspired by the inexact augmented Lagrange multiplier (ALM) method [35], we introduce auxiliary variables $\mathcal{M}$ and $\mathbf{W}^{(v)}$, and rewrite the model (9) as the following unconstrained problem:

$$\mathcal{L}\left(\mathcal{M}, \left\{ \mathbf{Z}^{(v)} \right\}_{v=1}^{V}, \left\{ \mathbf{W}^{(v)} \right\}_{v=1}^{V}, \mathbf{F} \right)$$

$$= \|\mathcal{M}\|_{\textcircled{Sp}}^{p} + \beta \sum_{v=1}^{V} \frac{1}{\alpha^{(v)}} \text{tr}(\mathbf{F}^T\mathbf{L}_{\mathbf{W}^{(v)}}\mathbf{F})$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

XIA *et al.*: TENSOR COMPLETION

5

$$+ \frac{\mu}{2}\left\|\boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{M}} + \frac{\boldsymbol{\mathcal{Q}}_1}{\mu}\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{Z}^{(v)} - \mathbf{W}^{(v)} + \frac{\mathbf{Q}_2^{(v)}}{\mu}\right\|_F^2$$

$$\text{s.t. } C_\Omega\left(\mathbf{Z}^{(v)}\right) = C_\Omega\left(\bar{\mathbf{S}}^{(v)}\right), \ \mathbf{F}^{\mathbf{T}}\mathbf{F} = \mathbf{I}$$

$$\mathbf{W}^{(v)} \geq 0, \ \mathbf{W}^{(v)}\mathbf{1} = \mathbf{1}, \ \mathbf{Z}^{(v)} \geq 0, \ \mathbf{Z}^{(v)}\mathbf{1} = \mathbf{1} \qquad (10)$$

where $\boldsymbol{\mathcal{Q}}_1$ and $\mathbf{Q}_2^{(v)}$ represent the Lagrange multipliers and $\mu > 0$ is the adaptive penalty factor. Consequently, the optimization process could be separated into four steps.

1) $\mathbf{Z}^{(v)}$-*Subproblem:* For updating $\mathbf{Z}^{(v)}$, we solving the following problem by fixing the variables $\boldsymbol{\mathcal{M}}$, $\boldsymbol{\mathcal{Q}}_1$, $\mathbf{Q}_2^{(v)}$, $\mathbf{F}$, and $\mathbf{W}^{(v)}$:

$$\mathbf{Z}^{(v)*} = \arg\min_{\mathbf{Z}^{(v)}} \frac{\mu}{2}\left\|\mathbf{Z}^{(v)} - \mathbf{M}^{(v)} + \frac{\mathbf{Q}_1^{(v)}}{\mu}\right\|_F^2$$

$$+ \frac{\mu}{2}\left\|\mathbf{Z}^{(v)} - \mathbf{W}^{(v)} + \frac{\mathbf{Q}_2^{(v)}}{\mu}\right\|_F^2$$

$$+ \text{tr}\left(\mathbf{Q}_3^{(v)\mathbf{T}}\left(C_\Omega\left(\mathbf{Z}^{(v)} - \bar{\mathbf{S}}^{(v)}\right)\right)\right)$$

$$= \arg\min_{\mathbf{Z}^{(v)}} \mu\left\|\mathbf{Z}^{(v)} - \frac{1}{2}\mathbf{J}^{(v)} + \frac{1}{2\mu}P_\Omega\left(\mathbf{Q}_3^{(v)\mathbf{T}}\right)\right\|_F^2 \tag{11}$$

where $\mathbf{J}^{(v)} = \mathbf{M}^{(v)} - \mathbf{Q}_1^{(v)}/\mu + \mathbf{W}^{(v)} - \mathbf{Q}_2^{(v)}/\mu$; $\mathbf{Q}_3^{(v)}$ is a Lagrange multiplier. We take the derivative of (11) w.r.t. $\mathbf{Z}^{(v)}$ and set it to 0, we have

$$\mathbf{Z}^{(v)*} = \frac{1}{2}\mathbf{J}^{(v)} - \frac{1}{2\mu}C_\Omega\left(\mathbf{Q}_3^{(v)\mathbf{T}}\right). \tag{12}$$

2) $\boldsymbol{\mathcal{M}}$-*Subproblem:* In this case, model (10) becomes

$$\boldsymbol{\mathcal{M}}^* = \arg\min_{\boldsymbol{\mathcal{M}}} \|\boldsymbol{\mathcal{M}}\|_{\textcircled{Sp}}^p + \frac{\mu}{2}\left\|\boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{M}} + \frac{\boldsymbol{\mathcal{Q}}_1}{\mu}\right\|_F^2. \tag{13}$$

To solve the model (13), Theorem 1 is introduced.

Theorem 1 [16]: Suppose $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $h = \min(n_1, n_2)$, let $\boldsymbol{\mathcal{Z}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^T$. For the following model:

$$\arg\min_{\boldsymbol{\mathcal{M}}} \frac{1}{2}\|\boldsymbol{\mathcal{M}} - \boldsymbol{\mathcal{Z}}\|_F^2 + \tau\|\boldsymbol{\mathcal{M}}\|_{\textcircled{Sp}}^p \tag{14}$$

the optimal solution $\boldsymbol{\mathcal{M}}^*$ is

$$\boldsymbol{\mathcal{M}}^* = \Gamma_{\tau \cdot n_3}(\boldsymbol{\mathcal{Z}}) = \boldsymbol{\mathcal{U}} * \textit{ifft}\left(\mathbf{B}_{\tau \cdot n_3}\left(\overline{\boldsymbol{\mathcal{Z}}}\right)\right) * \boldsymbol{\mathcal{V}}^{\mathbf{T}} \tag{15}$$

where $\mathbf{B}_{\tau \cdot n_3}(\overline{\boldsymbol{\mathcal{Z}}})$ is a tensor, and the ith frontal slice of $\mathbf{B}_{\tau \cdot n_3}(\overline{\boldsymbol{\mathcal{Z}}})$ is $\mathbf{B}_{\tau \cdot n_3}(\overline{\boldsymbol{\mathcal{Z}}}^{(i)})$.

Thus, from Theorem 1, the solution to model (13) is

$$\boldsymbol{\mathcal{M}}^* = \Gamma_{\frac{1}{\rho}}\left(\boldsymbol{\mathcal{Z}} + \frac{1}{\mu}\boldsymbol{\mathcal{Q}}_1\right). \tag{16}$$

3) $\mathbf{W}^{(v)}$-*subproblem:* In this case, model (10) becomes

$$\min_{0 \leq \mathbf{W}^{(v)}, \mathbf{W}^{(v)}\mathbf{1}=\mathbf{1}} \beta \sum_{v=1}^V \frac{1}{\alpha^{(v)}}\text{tr}\left(\mathbf{F}^{\mathbf{T}}\mathbf{L}_{\mathbf{W}^{(v)}}\mathbf{F}\right)$$

$$+ \frac{\mu}{2}\left\|\mathbf{Z}^{(v)} - \mathbf{W}^{(v)} + \frac{\mathbf{Q}_2^{(v)}}{\mu}\right\|_F^2. \tag{17}$$

By simple algebra, we have

$$\min_{\mathbf{W}^{(v)}} \beta \frac{1}{\alpha^{(v)}}\text{tr}\left(\mathbf{F}^{\mathbf{T}}\mathbf{L}_{\mathbf{W}^{(v)}}\mathbf{F}\right)$$

$$+ \frac{\mu}{2}\left\|\mathbf{Z}^{(v)} - \mathbf{W}^{(v)} + \frac{\mathbf{Q}_2^{(v)}}{\mu}\right\|_F^2 + \varphi^{(v)\mathbf{T}}(\mathbf{W}^{(v)}\mathbf{1} - \mathbf{1})$$

$$= \min_{\mathbf{W}^{(v)}} \frac{\beta}{2\alpha^{(v)}}\sum_{i,j}\|\mathbf{F}_{i,:} - \mathbf{F}_{j,:}\|_2^2 w_{ij}^{(v)}$$

$$+ \frac{\mu}{2}\|\mathbf{W}^{(v)} - \mathbf{R}^{(v)}\|_F^2 + \varphi^{(v)\mathbf{T}}(\mathbf{W}^{(v)}\mathbf{1} - \mathbf{1}) \tag{18}$$

where $\mathbf{F}_{i,:}$ and $\mathbf{F}_{j,:}$ are the ith and jth rows of $\mathbf{F}$. We can see that problem (18) is independent to each row. Letting $\mathbf{R}^{(v)} = \mathbf{Z}^{(v)} + [\mathbf{Q}_2^{(v)}/\mu]$ and $\mathbf{D}_{ij} = \sum_{i,j}\|\mathbf{F}_{i,:} - \mathbf{F}_{j,:}\|_2^2$, then problem (18) can be simplified into the following problem:

$$\min_{\mathbf{W}^{(v)}} \frac{\beta}{2\alpha^{(v)}}\mathbf{W}_{i,:}^{(v)}\mathbf{D}_{i,:}^{(v)\mathbf{T}}$$

$$+ \frac{\mu}{2}\left\|\mathbf{W}_{i,:}^{(v)} - \mathbf{R}_{i,:}^{(v)}\right\|_F^2 - \varphi_i^{(v)T}\left(\mathbf{W}_{i,:}^{(v)}\mathbf{1} - \mathbf{1}\right)$$

$$= \min_{\mathbf{W}^{(v)}} \frac{\mu}{2}\left\|\mathbf{W}_{i,:}^{(v)} - \left(\mathbf{R}_{i,:}^{(v)} - \frac{\beta}{2\mu\alpha^{(v)}}\mathbf{D}_{i,:}^{(v)}\right)\right\|_F^2$$

$$- \varphi_i^{(v)T}\left(\mathbf{W}_{i,:}^{(v)}\mathbf{1} - \mathbf{1}\right) \tag{19}$$

where $\varphi_i^{(v)}$ is the ith element of vector $\varphi^{(v)}$. The optimal solution to problem (19) is

$$w_{ij}^{(v)} = \mathbf{R}_{ij}^{(v)} - \frac{\beta}{\mu\alpha^{(v)}}\mathbf{D}_{ij} + \frac{\varphi_i^{(v)}}{\mu} \tag{20}$$

Due to $\mathbf{W}^{(v)} = \max(\mathbf{W}^{(v)}, 0)$, all elements of matrix $\mathbf{W}^{(v)}$ are enforced to be not less than 0, that is, the elements less than 0 in $\mathbf{W}^{(v)}$ are set 0, and the remaining elements are preserved. Due to $\mathbf{W}^{(v)}\mathbf{1} = \mathbf{1}$, the Lagrange multiplier $\varphi^{(v)}$ can be updated by

$$\varphi_i^{(v)} = \mu\left(1 - \sum_{j=1}^n \mathbf{R}_{i,j}^{(v)} - \frac{\beta}{2\mu\alpha^{(v)}}\mathbf{D}_{i,j}^{(v)}\right)/(V - 1). \tag{21}$$

4) $\mathbf{F}$-*Subproblem:* The matrix $\mathbf{F}$ is optimized by

$$\min_{\mathbf{F}^{\mathbf{T}}\mathbf{F}=\mathbf{I}} \sum_{v=1}^V \frac{1}{\alpha^{(v)}}\text{tr}\left(\mathbf{F}^{\mathbf{T}}\mathbf{L}_{\mathbf{W}^{(v)}}\mathbf{F}\right)$$

$$= \min_{\mathbf{F}^{\mathbf{T}}\mathbf{F}=\mathbf{I}} \text{tr}\left(\mathbf{F}^{\mathbf{T}}\sum_{v=1}^V \frac{1}{\alpha^{(v)}}\mathbf{L}_{\mathbf{W}^{(v)}}\mathbf{F}\right). \tag{22}$$

The optimal solution $\mathbf{F}$ is formed by the K eigenvectors corresponding to the K smallest eigenvalues of $[\mathbf{L}_{\mathbf{W}^{(v)}}/\alpha^{(v)}]$.

5) $\alpha^{(v)}$-*Problem:* To solve $\alpha^{(v)}$, model (10) becomes

$$\min_{\alpha^{(v)}} \sum_{v=1}^V \frac{1}{\alpha^{(v)}}\text{tr}\left(\mathbf{F}^{\mathbf{T}}\mathbf{L}_{\mathbf{W}^{(v)}}\mathbf{F}\right)$$

$$\text{s.t. } \sum_{v=1}^V \alpha^{(v)} = 1, \alpha^{(v)} \geq 0. \tag{23}$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

6   IEEE TRANSACTIONS ON CYBERNETICS

Algorithm 1: Tensor Completion-Based IMC

Input: Data matrix $\{\mathbf{X}^{(v)}\}_{v=1}^{V}$, index matrix $\mathbf{G}^{(v)}$, parameter β.
Output: The indicator matrix $\mathbf{H}$.

1 Construct $\mathbf{S}^{(v)}$ like [36], $\bar{\mathbf{S}}^{(v)} = \mathbf{G}^{(v)\mathrm{T}}\mathbf{S}^{(v)}\mathbf{G}^{(v)}$;
2 Initialize $\mathcal{Q}_1 = \mathbf{0}$, $\mathbf{Q}_2^{(v)} = \mathbf{0}$, $\mathbf{Q}_3^{(v)} = \mathbf{0}$, $\rho = 1.1$, $\mu_0 = 10^{-5}$, $\alpha^{(v)} = \frac{1}{V}$;
3 **while** *not converge* **do**
4   Update $\mathbf{Z}^{(v)}$ by Eq. (12);
5   Update $\mathbf{W}^{(v)}$ by Eq. (20);
6   Update $\mathcal{M}$ by Eq. (16);
7   Update $\mathbf{F}$ by Eq. (22);
8   Update $\alpha^{(v)}$ by Eq. (26);
9   Update μ, $\mathcal{Q}_1$, $\mathbf{Q}_2^{(v)}$, $\mathbf{Y}^{(v)}$ by Eq. (27);
10 **end**
11 **return:** Indicator matrix $\mathbf{H}$. After obtained $\mathbf{F}$, we can use K-means to get indicator matrix $\mathbf{H}$.

Letting $\psi^{(v)} = \sqrt{\mathrm{tr}(\mathbf{F}^{\mathrm{T}}\mathbf{L}_{\mathbf{W}^{(v)}}\mathbf{F})}$, then (23) can be written as

$$\min_{\alpha^{(v)}} \sum_{v=1}^{V} \frac{1}{\alpha^{(v)}}\left(\psi^{(v)}\right)^2 \text{ s.t. } \sum_{v=1}^{V} \alpha^{(v)} = 1, \alpha^{(v)} \geq 0 \quad (24)$$

Due to $\sum_{v=1}^{V} \alpha^{(v)} = 1$, according to the Cauchy–Schwarz inequality, we have

$$\sum_{v=1}^{V} \frac{\psi^{(v)2}}{\alpha^{(v)}} = \left(\sum_{v=1}^{V} \frac{\psi^{(v)2}}{\alpha^{(v)}}\right)\left(\sum_{v=1}^{V} \alpha^{(v)}\right) \geq \left(\sum_{v=1}^{V} \psi^{(v)}\right)^2 \quad (25)$$

where the equation holds, if and only if $\sqrt{\alpha^{(v)}} \propto [\psi^{(v)}/\sqrt{\alpha^{(v)}}]$. Because the right-hand side in (25) is a constant, therefore $\forall v = 1, \ldots, V$, the optimal $\alpha^{(v)}$ is

$$\alpha^{(v)} = \psi^{(v)} / \sum_{v=1}^{V} \psi^{(v)}. \quad (26)$$

6) *Multipliers and Penalty Parameter Problem:* The Lagrange multipliers $\mathcal{Q}_1$, $\mathbf{Q}_2^{(v)}$, and $\mathbf{Q}_3^{(v)}$ and penalty parameter μ are updated by

$$\mathcal{Q}_1 = \mathcal{Q}_1 + \mu(\mathcal{Z} - \mathcal{M}), \ \mu = \min(\rho\mu, \ \mu_0)$$
$$\mathbf{Q}_2^{(v)} = \mathbf{Q}_2^{(v)} + \mu\left(\mathbf{Z}^{(v)} - \mathbf{W}^{(v)}\right)$$
$$\mathbf{Q}_3^{(v)} = \mathbf{Q}_3^{(v)} + \mu C_\Omega(\mathbf{Z}^{(v)} - \left(\mathbf{G}^{(v)}\right)^{\mathrm{T}}\mathbf{S}^{(v)}\mathbf{G}^{(v)}) \quad (27)$$

where $\rho > 1$ is a positive number.

Finally, the pseudocode of solving model (10) is reported in Algorithm 1. Note that $\mathbf{S}^{(v)}$ ($v = 1, 2, \ldots, V$) are initialized by the same way as in [36].

D. Computational Complexity

Since matrix $\mathbf{J}^{(v)}$ can be calculated in advance in (12) for solving $\mathbf{Z}^{(v)}$, the computational bottleneck of the proposed algorithm (Algorithm 1) only lies in solving two variable ($\mathcal{M}$ and $\mathbf{F}$). First, solving the $\mathcal{M}$-subproblem involves calculating the 3D FFT and 3D inverse FFT of an $N \times V \times N$ tensor and N SVDs of $N \times V$ matrices in the Fourier domain, both of which are with the complexity of $\mathcal{O}(2N^2V\log(N))$ and $\mathcal{O}(N^2V^2)$. The computation of updating $\mathbf{F}$ is $\mathcal{O}(N^3)$, since it is with the

eigendecomposition of the $N \times N$ matrix. Since in multiview scenarios, we have $N \gg V$ and $\log(N) > V$ [37], thus, the total complexity of the proposed method is $\mathcal{O}(N^3 + 2N^2V\log(N))$ for each iteration.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Experimental Setup

Datasets: We conduct different clustering tasks to investigate the superiority of the proposed method. These different tasks involves the following multiview benchmark datasets.

1) *Handwritten (HW) Digits Dataset [42]:* HW includes HW digits from 0 to 9. Each digit is composed of 200 images. Thus, this database has 2000 images. In the experiments, for each digit, we select two views for clustering. One view is composed of Fourier coefficients whose dimension is 76. Another view is profile correlations of digits whose dimension (D) is 216.

2) *BDGP Dataset [43]:* BDGP has 2500 drosophila embryos images, which are sampled from five different classes. Each class has two views: a) visual view and b) textual view. Each data are presented by 1750-D vector in visual view and 79-D vector in textual view. In the following experiments, all samples are used to evaluate clustering performance.

3) *ORL Dataset[1]:* This database has 40 distinct individuals. Each individuals have ten images. In the subsequent experiments, we reshape each image to be 32×32, and then extract three kinds of features, which are 1024-D LBP, 512-D GIST, and 1024-D pyramid of histogram of oriented gradients (PHOGs). After that, combining the aforementioned three types of features and original gray image, we have four views for multiview clustering.

4) *3 Sources Dataset[2]:* It consists of 984 news articles which are obtained from three sources: a) Reuters; b) BBC; and c) Guardian. In the subsequent experiments, we select 169 stories as gallery. These 169 stories belong to six different topical labels which are sport, business, health, technology, entertainment, and politics.

5) *BBCSport Databset [44]:* This database includes 737 sport news articles which are obtained from the BBC Sport Website and belong to five categories. Each document is presented by two–five views. In the following experiments, we select 116 documents[3] having four views to evaluate the clustering performance.

Two Kinds of Incomplete Multiview: The following two kinds of incomplete view data are constructed.

1) *Some Samples Have Complete Views:* In HW and BDGP datasets, to form the incomplete multiview datasets, we randomly choose 10%, 30%, 50%, and 70% paired data from the corresponding galleries. Then, for 50% of the rest data, we remove their 1st view, and for additional 50% of the remaining data, we remove their 2nd view.

---

[1]http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
[2]http://erdos.ucd.i.e.,/datasets/3sources.html
[3]https://github.com/GPMVCDummy/GPMVC/tree/master/partialMV/PVC/recreateResults/data

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

XIA *et al.*: TENSOR COMPLETION

7

TABLE I
CLUSTERING PERFORMANCES VERSUS DIFFERENT IMPARTIAL RATIO

Dataset	Handwritten digits (HW)											
Metrics	ACC (↑)				NMI (↑)				Purity (↑)			
Method / PER	0.1	0.3	0.5	0.7	0.1	0.3	0.5	0.7	0.1	0.3	0.5	0.7
SC1 [38]	0.4458	0.5097	0.5742	0.6432	0.3740	0.4419	0.5083	0.5717	0.4524	0.5175	0.5820	0.6511
SC2 [38]	0.2981	0.3279	0.3527	0.3766	0.2276	0.2614	0.2973	0.3296	0.3075	0.3366	0.3658	0.3914
ConSC [38]	0.4457	0.4825	0.5453	0.6454	0.4836	0.5202	0.5765	0.6661	0.4721	0.5148	0.5767	0.6587
AMGL [39]	0.3426	0.4221	0.5024	0.6037	0.2695	0.3576	0.4633	0.5840	0.3505	0.4373	0.5266	0.6222
RMSC [40]	0.4016	0.4625	0.5564	0.6330	0.3988	0.4461	0.5020	0.5650	0.4212	0.4782	0.5651	0.6404
GPVC [41]	0.3238	0.3077	0.3419	0.4236	0.3609	0.3787	0.4256	0.4993	0.3439	0.3410	0.3692	0.4425
IMG [24]	0.5350	0.5455	0.5457	0.5529	0.4359	0.4532	0.4590	0.4798	0.5469	0.5565	0.5585	0.5690
IMSC-AGL [15]	0.6724	0.7926	0.8342	0.8552	0.6701	0.6783	0.7297	0.7582	0.6978	0.7926	0.8342	0.8552
Ours	**0.7632**	**0.8032**	**0.8570**	**0.8701**	**0.7061**	**0.7173**	**0.7838**	**0.7904**	**0.7635**	**0.8071**	**0.8595**	**0.8701**
Dataset	BDGP											
Metrics	ACC (↑)				NMI (↑)				Purity (↑)			
Method / PER	0.1	0.3	0.5	0.7	0.1	0.3	0.5	0.7	0.1	0.3	0.5	0.7
SC1 [38]	0.3296	0.3539	0.3845	0.4103	0.1120	0.1347	0.1626	0.1884	0.3339	0.3572	0.3878	0.4122
SC2 [38]	0.4748	0.5169	0.5692	0.6139	0.2698	0.3158	0.3745	0.4364	0.4798	0.5235	0.5764	0.6194
ConSC [38]	0.2781	0.2230	0.2139	0.2106	0.0793	0.0228	0.0126	0.0098	0.2835	0.2268	0.2165	0.2126
AMGL [39]	0.2524	0.2357	0.2538	0.2807	0.0446	0.0260	0.0420	0.0756	0.2553	0.2386	0.2558	0.2826
RMSC [40]	0.3395	0.3683	0.3907	0.4233	0.1304	0.1487	0.1684	0.1962	0.3416	0.3712	0.3934	0.4246
GPVC [41]	0.5015	0.5424	0.6277	0.6833	0.3330	0.3907	0.4827	0.5741	0.5089	0.5615	0.6451	0.7062
IMG [24]	0.4373	0.4508	0.4868	0.5055	0.2764	0.2809	0.3147	0.3246	0.4431	0.4570	0.4916	0.5099
IMSC-AGL [15]	0.5625	0.6281	0.7258	0.7524	0.4006	0.3914	0.4851	0.5560	0.5625	0.6022	0.7258	0.7524
Ours	**0.6816**	**0.7200**	**0.7604**	**0.7884**	**0.4473**	**0.4971**	**0.5419**	**0.5808**	**0.6816**	**0.7200**	**0.7604**	**0.7884**

TABLE II
CLUSTERING RESULTS WHEN HANDLING THE DATA THAT NONE OF THE SAMPLES HAS COMPLETE VIEWS

Dataset	ORL			3 Sources			BBCSport		
Method	ACC (↑)	NMI (↑)	Purity (↑)	ACC (↑)	NMI (↑)	Purity (↑)	ACC (↑)	NMI (↑)	Purity (↑)
BSV [24]	0.3556	0.4203	0.4132	0.3598	0.2148	0.4936	0.4321	0.1860	0.4700
Concat [24]	0.4135	0.5403	0.4331	0.4268	0.2608	0.5513	0.4290	0.1948	0.4634
IMSC-AGL [15]	0.5602	0.7108	0.5891	0.4768	0.2862	0.5643	0.6914	0.5376	0.7810
Ours	**0.6010**	**0.7499**	**0.6260**	**0.7124**	**0.4900**	**0.7124**	**0.7759**	**0.6484**	**0.8534**

2) *None of the Samples Has Complete Views:* For ORL, 3 sources and BBCSport datasets, we randomly remove 55% data for each view and build three incomplete multiview gallery, which have no paired samples, for clustering.

Competitors: To evaluate the performances of the proposed method, the following methods are included, that is: 1) SC [38]; 2) SC with feature concatenation (ConSC) [38]; 3) autoweighted multiple graph learning (AMGL) [39]; 4) robust multiview SC (RMSC) [40]; 5) PMVC using graph regularized NMF (GPVC) [41]; 6) incomplete multimodal visual data grouping (IMG) [24]; 7) best single view (BSV) [24]; 8) BSV with concatenating all views into a single view (Concat) [24]; and 9) IMSC-AGL [15].

Evaluation Criteria: Three metrics are employed to evaluate the effectiveness of different methods, that is: 1) Purity; 2) normalized mutual information (NMI); and 3) Accuracy (ACC). For these three metrics, the higher the value, the better the clustering performance. In the following experiments, we repeat each experiment five times and list mean values of these three metrics.

Implementation Details: For the proposed method, we implement it in MATLAB R2018a. There exist two hyperparameters in the proposed method, that is, the power p of tensor Schatten p-norm, and parameter β of the connectivity constraint. We tune the value of p from 0.1 to 1.0 with step

size 0.1, and tune the value of β in the range of [0.0001, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 1.1, 1.5, 5] to obtain the best results.

B. Experimental Results and Analysis

We compare the clustering results of the proposed method with several methods w.r.t. under two kinds of incomplete view settings. To be specific, Table I reports the clustering results of different methods under the first kind of incomplete view data, Table II reports the clustering results of different methods under the second kind of incomplete view data. From Tables I and II, we have the following interesting observations.

1) In most cases, the clustering results of Single-view methods, that is, SC and ConSC, are overall inferior to multiview methods. The reason may be that multiview methods may leverage the complementary information embedded in multiview data, while the single-view method does not. Meanwhile. when using the concatenated view features, the performances of ConSC defeat the methods that using the single-view features on the HW dataset, even the results of ConSC are better than some methods with the multiview setting. In contrast, the proposed method obtains superior results than both SC and ConSC, the pivotal reason behind this is that the proposed model explores more informative knowledge
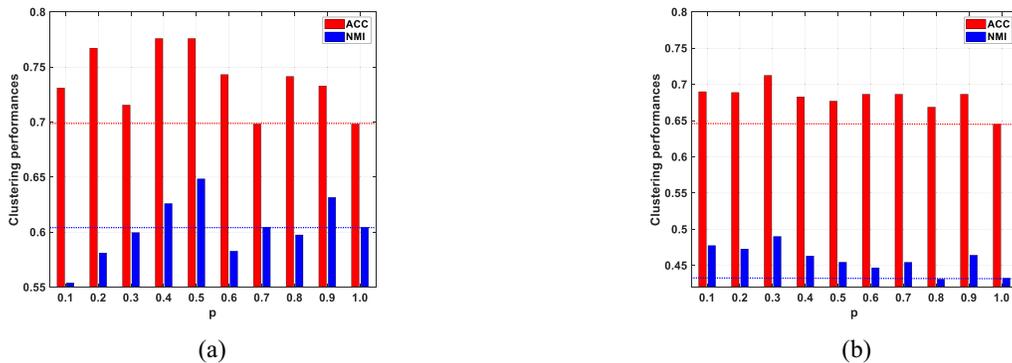
Fig. 2.   Clustering performances of our method with the varying value of p on (a) BBCSport and (b) 3 Sources datasets.

embedded in single-view features via the proposed tensor Schatten p-norm based complete technique.

2) When dealing with incomplete data, as shown in Table I, the clustering results of several multiview methods, for example, AMGL and RMSC, are inferior to single-view methods. These results show that data missing can easily affect the performance of these methods. Moreover, to handle the incomplete multiview clustering problem, these methods uses the average instances to fill the missing instances, these results also demonstrate that the clustering results obtained under such a data completion technique are inferior.

3) Comparatively, the proposed method overall outperforms the second-best competitor, that is, IMSC-AGL. The reason maybe that we employ the tensor Schatten p-norm based complete technique to learn the consensus graph, which helps to characterize the similarity structure of interview graphs. Thus, the learned graph well exploits both the spatial structure and complementary information embedded in graphs of different views. Meanwhile, the connectivity constraint on the learned graph helps to characterize the cluster structure of data.

4) When handling data that none of the samples has complete views, the performance of the proposed method is still satisfactory. For example, in Table II, comparing with the strongest competitor, that is, IMSC-AGL on the 3 Sources dataset, the ACC, NMI, and Purity are improved about 23.56%, 20.38%, and 14.81%, respectively, the ACC, NMI, and Purity are improved about 8.45%, 11.08%, and 7.24% on the BBCSport dataset, respectively. This is because we leverage the tensor Schatten p-norm-based complete technique to construct the graph of entire data including missing view, the learned graph via such a strategy well preserves the relationship between the unmissing data. These results strongly indicate that the proposed method is very effective in clustering the multiview data that none of the samples has complete views.

5) For the results based on the second kind of incomplete multiview data, the proposed model significantly and consistently outperforms the single-view clustering method, for example, BSV and Concat. Specifically, on the ORL dataset, in terms of Purity and NMI, it

TABLE III
CLUSTERING RESULTS W./W.O. ADAPTIVE WEIGHTING STRATEGY, WHERE ✗ MEANS WITHOUT THE ADAPTIVE WEIGHTING STRATEGY

Dataset	3 Sources		
Method	ACC (↑)	NMI (↑)	Purity (↑)
✗	0.6852	0.4611	0.6852
✓	**0.7124**	**0.4900**	**0.7124**
Dataset	BBCSport		
Method	ACC (↑)	NMI (↑)	Purity (↑)
✗	0.7414	0.6131	0.8276
✓	**0.7759**	**0.6484**	**0.8534**

improves over the second-best single-view Concat by 19.29% and 20.96%, respectively. This is because the proposed method takes into account the contributions of different views by the adaptive weighting strategy, which helps encode the discriminant information embedded in multiple graphs.

C. Ablation Studies

In this section, we conduct ablation studies to further verify the effectiveness of each component, that is, the tensor Schatten p-norm and adaptive weighting strategy, in the proposed method.

1) Effect of Tensor Schatten p-Norm: To this end, Fig. 2 plots the clustering results, that is, ACC and NMI, versus parameter p on the BBCSport and 3 Sources datasets. From Fig. 2, we can see that the performance of our method has large fluctuation with varying p. When $p \neq 1$, the clustering performance of our method is overall superior to that under $p = 1$. Note that when $p = 1$, tensor Schatten p-norm degenerates to tensor nuclear norm [32]. On the BBCSport and 3 Sources datasets, when p equals 0.5 and 0.3, respectively, our method obtains the best ACC and NMI that are remarkably superior to that under $p = 1$. These results indicate that tensor Schatten p-norm makes the rank of the learned graph approximate the target rank well.

2) Effect of Adaptive Weighting Strategy: Moreover, we compare the clustering results w./w.o. the adaptive weighting strategy. It should be pointed out that when we remove the adaptive weighting strategy from (9), $\alpha^{(v)} = 1/V$. As can be seen in Table III, the adaptive weighting strategy plays an indispensable role in multiview clustering.
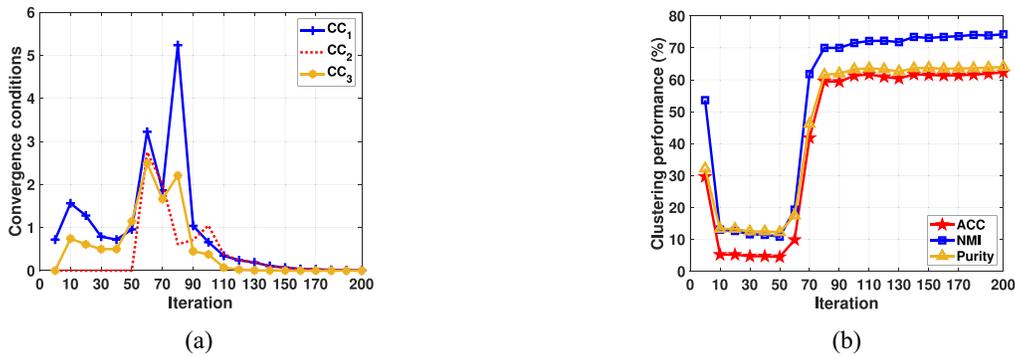
Fig. 3. Convergence results on the ORL dataset. (a) Error convergence curves. (b) Clustering results convergence curves.

D. Convergence Analysis

In [35], the convergence of ALM has been well proved when the number of blocks is at most two, whereas its convergence is still an open problem when number of blocks is greater than or equal to three [45]. Since there are four variables $(\mathcal{M}, \{\mathbf{Z}^{(v)}\}_{v=1}^{V}, \{\mathbf{W}^{(v)}\}_{v=1}^{V})$ and $\mathbf{F}$ in our proposed method, and the objective function is not smooth, it would be difficult to prove the convergence in theory. Fortunately, experimental results on real datasets indicate that the proposed method has a good convergence property. In Fig. 3(a), we show the errors of variables of the proposed method in each iteration on the ORL dataset. Here, the errors of variables are $\text{CC}_1 = \|\mathbf{Z}_{t+1}^{(v)} - \mathbf{Z}_t^{(v)}\|_{\infty}$, $\text{CC}_2 = \|\mathcal{M}_{t+1} - \mathcal{M}_t\|_{\infty}$ and $\text{CC}_3 = \|\mathbf{M}_t^{(v)} - \mathbf{Z}_t^{(v)}\|_{\infty}$. According to Fig. 3(a), the results verify that the proposed method can achieve convergence within a few iterations. Moreover, in Fig. 3(b), we show the clustering results versus iteration number. According to Fig. 3(b), we can see that the proposed method obtains relatively stable clustering performance within a few iterations. These results well indicate that the proposed method can converges quickly.

V. CONCLUSION

In this article, we completed the incomplete graph with missing data referring to tensor complete, and presented an effective incomplete multiview clustering model. To take the similarity structure of interview graphs into account, the tensor Schatten *p*-norm-based completion technique was leveraged to complete the entire graph, which can guarantee that the learned entire graph not only has low-rank structure but also well preserves the relationship between the unmissing data. After that, the connectivity constraint was employed on the learned graph to make sure that the connected components approximately indicate clusters. This also helps guide the tensor completion. Extensive experimental results indicated the efficacy of the proposed incomplete multiview clustering method on kinds of datasets in terms of three evaluation metrics. In the future, we plan to investigate the deep learning-based tensor completion method for incomplete multiview clustering.

REFERENCES

[1] Y. Xie, W. Zhang, Y. Qu, L. Dai, and D. Tao, "Hyper-Laplacian regularized multilinear multiview self-representations for clustering and semisupervised learning," *IEEE Trans. Cybern.*, vol. 50, no. 2, pp. 572–586, Feb. 2020.

[2] C. Zhang *et al.*, "Generalized latent multi-view subspace clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 1, pp. 86–99, Jan. 2020.

[3] F. Nie, Z. Wang, R. Wang, and X. Li, "Submanifold-preserving discriminant analysis with an auto-optimized graph," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp. 3682–3695, Aug. 2020.

[4] Y. Pang, J. Xie, F. Nie, and X. Li, "Spectral clustering by joint spectral embedding and spectral rotation," *IEEE Trans. Cybern.*, vol. 50, no. 1, pp. 247–258, Jan. 2020.

[5] G. Sun, Y. Cong, J. Dong, Y. Liu, Z. Ding, and H. Yu, "What and how: Generalized lifelong spectral clustering via dual memory," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Feb. 11, 2021, doi: 10.1109/TPAMI.2021.3058852.

[6] W. Xia, X. Zhang, Q. Gao, X. Shu, J. Han, and X. Gao, "Multiview subspace clustering by an enhanced tensor nuclear norm," *IEEE Trans. Cybern.*, early access, Feb. 26, 2021, doi: 10.1109/TCYB.2021.3052352.

[7] Q. Wang, J. Cheng, Q. Gao, G. Zhao, and L. Jiao, "Deep multi-view subspace clustering with unified and discriminative learning," *IEEE Trans. Multimedia*, vol. 23, pp. 3483–3493, 2021.

[8] A. Trivedi, P. Rai, H. Daumé III, and S. L. DuVall, "Multiview clustering with incomplete views," in *Proc. NeurIPS Workshop*, 2010, pp. 1–7.

[9] S.-Y. Li, Y. Jiang, and Z.-H. Zhou, "Partial multi-view clustering," in *Proc. AAAI*, 2014, pp. 1968–1974.

[10] W. Shao, L. He, and P. S. Yu, "Multiple incomplete views clustering via weighted nonnegative matrix factorization with l_2, 1 regularization," in *Proc. ECML*, 2015, pp. 318–334.

[11] J. Guo and W. Zhu, "Partial multi-view outlier detection based on collective learning," in *Proc. AAAI*, 2018, pp. 298–305.

[12] X. Liu *et al.*, "Efficient and effective regularized incomplete multi-view clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 8, pp. 2634–2646, Aug. 2021.

[13] D. Xie, X. Zhang, Q. Gao, J. Han, S. Xiao, and X. Gao, "Multiview clustering by joint latent representation and similarity learning," *IEEE Trans. Cybern.*, vol. 50, no. 11, pp. 4848–4854, Nov. 2020.

[14] C. Zhang, Q. Hu, H. Fu, P. Zhu, and X. Cao, "Latent multi-view subspace clustering," in *Proc. IEEE CVPR*, 2017, pp. 4333–4341.

[15] J. Wen, Y. Xu, and H. Liu, "Incomplete multiview spectral clustering with adaptive graph learning," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1418–1429, Apr. 2020.

[16] Q. Gao, P. Zhang, W. Xia, D. Xie, X. Gao, and D. Tao, "Enhanced tensor RPCA and its application," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 6, pp. 2133–2140, Jun. 2021.

[17] H. Gao, F. Nie, X. Li, and H. Huang, "Multi-view subspace clustering," in *Proc. IEEE ICCV*, 2015, pp. 4238–4246.

[18] D. Xie, W. Xia, Q. Wang, Q. Gao, and S. Xiao, "Multi-view clustering by joint manifold learning and tensor nuclear norm," *Neurocomputing*, vol. 380, pp. 105–114, Mar. 2020.

[19] G. Sun, Y. Cong, Y. Zhang, G. Zhao, and Y. Fu, "Continual multiview task learning via deep matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 139–150, Jan. 2021.

[20] J. Wen, Z. Zhang, Y. Xu, and Z. Zhong, "Incomplete multi-view clustering via graph regularized matrix factorization," in *Proc. ECCV Workshops*, 2018, pp. 593–608.

[21] Z. Huang, P. Hu, J. T. Zhou, J. Lv, and X. Peng, "Partially view-aligned clustering," in *Proc. NeurIPS*, 2020.

[22] J. Wen, Z. Zhang, Z. Zhang, L. Fei, and M. Wang, "Generalized incomplete multiview clustering with flexible locality structure diffusion," *IEEE Trans. Cybern.*, vol. 51, no. 1, pp. 101–114, Jan. 2021.

[23] X. Liu *et al.*, "Late fusion incomplete multi-view clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 10, pp. 2410–2423, Oct. 2019.

[24] H. Zhao, H. Liu, and Y. Fu, "Incomplete multi-modal visual data grouping," in *Proc. IJCAI*, 2016, pp. 2392–2398.

[25] W. Zhou, H. Wang, and Y. Yang, "Consensus graph learning for incomplete multi-view clustering," in *Proc. PAKDD*, 2019, pp. 529–540.

[26] H. Wang, L. Zong, B. Liu, Y. Yang, and W. Zhou, "Spectral perturbation meets incomplete multi-view data," in *Proc. IJCAI*, 2019, pp. 3677–3683.

[27] J. Wen *et al.*, "DIMC-net: Deep incomplete multi-view clustering network," in *Proc. ACM MM*, 2020, pp. 3753–3761.

[28] J. Wen, Z. Zhang, Y. Xu, B. Zhang, L. Fei, and G.-S. Xie, "CDIMC-net: Cognitive deep incomplete multi-view clustering network," in *Proc. IJCAI*, 2020, pp. 3230–3236.

[29] Y. Lin, Y. Gou, Z. Liu, B. Li, J. Lv, and X. Peng, "COMPLETER: Incomplete multi-view clustering via contrastive prediction," in *Proc. IEEE CVPR*, 2021, pp. 11174–11183.

[30] Q. Wang, H. Lian, G. Sun, Q. Gao, and L. Jiao, "iCmSC: Incomplete cross-modal subspace clustering," *IEEE Trans. Image Process.*, vol. 30, pp. 305–317, 2021.

[31] Q. Wang, Z. Ding, Z. Tao, Q. Gao, and Y. Fu, "Generative partial multi-view clustering with adaptive fusion and cycle consistency," *IEEE Trans. Image Process.*, vol. 30, pp. 1771–1783, 2021.

[32] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis with a new tensor nuclear norm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 925–938, Apr. 2020.

[33] F. R. Chung and F. C. Graham, *Spectral Graph Theory*. vol. 92. Providence, RI, USA: Amer. Math. Soc., 1997.

[34] K. Fan, "On a theorem of Weyl concerning eigenvalues of linear transformations I," *Proc. Nat. Acad. Sci. USA*, vol. 35, no. 11, pp. 652–655, 1949.

[35] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," 2010, *arXiv:1009.5055*.

[36] F. Nie, G. Cai, J. Li, and X. Li, "Auto-weighted multi-view learning for image clustering and semi-supervised classification," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1501–1511, Mar. 2018.

[37] Y. Xie, D. Tao, W. Zhang, Y. Liu, L. Zhang, and Y. Qu, "On unifying multi-view self-representations for clustering by tensor multi-rank minimization," *Int. J. Comput. Vis.*, vol. 126, no. 11, pp. 1157–1179, 2018.

[38] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. NeurIPS*, 2001, pp. 849–856.

[39] F. Nie, J. Li, and X. Li, "Parameter-free auto-weighted multiple graph learning: A framework for multiview clustering and semi-supervised classification," in *Proc. IJCAI*, 2016, pp. 1881–1887.

[40] R. Xia, Y. Pan, L. Du, and J. Yin, "Robust multi-view spectral clustering via low-rank and sparse decomposition," in *Proc. AAAI*, 2014, pp. 2149–2155.

[41] N. Rai, S. Negi, S. Chaudhury, and O. Deshmukh, "Partial multi-view clustering using graph regularized NMF," in *Proc. ICPR*, 2016, pp. 2192–2197.

[42] M. van Breukelen, R. P. W. Duin, D. M. J. Tax, and J. E. den Hartog, "Handwritten digit recognition by combined classifiers," *Kybernetika*, vol. 34, no. 4, pp. 381–386, 1998.

[43] X. Cai, H. Wang, H. Huang, and C. H. Q. Ding, "Joint stage recognition and anatomical annotation of *drosophila* gene expression patterns," *Bioinformatics*, vol. 28, no. 12, pp. 16–24, 2012.

[44] D. Greene and P. Cunningham, "Practical solutions to the problem of diagonal dominance in kernel document clustering," in *Proc. ICML*, vol. 148, 2006, pp. 377–384.

[45] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

Wei Xia (Graduate Student Member, IEEE) received the B.Eng. degree in communication engineering from the Lanzhou University of Technology, Lanzhou, China, in 2018. He is currently pursuing the Ph.D. degree in communication and information system in Xidian University, Xi'an, China.

His research interests include pattern recognition, machine learning, and deep learning.

Quanxue Gao received the B.Eng. degree from Xi'an Highway University, Xi'an, China, in 1998, the M.S. degree from the Gansu University of Technology, Lanzhou, China, in 2001, and the Ph.D. degree from Northwestern Polytechnical University, Xi'an, in 2005.

He was an Associate Research with the Biometrics Center, The Hong Kong Polytechnic University, Hong Kong, from 2006 to 2007. From 2015 to 2016, he was a Visiting Scholar with the Department of Computer Science, The University of Texas at Arlington, Arlington, TX, USA. He is currently a Professor with the School of Telecommunications Engineering, Xidian University, Xi'an, and also a Key Member of State Key Laboratory of Integrated Services Networks. He has authored around 80 technical articles in refereed journals and proceedings, including IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON CYBERNETICS, CVPR, AAAI, and IJCAI. His current research interests include pattern recognition and machine learning.

Qianqian Wang received the B.Eng. degree in communication engineering from the Lanzhou University of Technology, Lanzhou, China, in 2014, and the Ph.D. degree from Xidian University, Xi'an, China, in 2019.

She is currently a Lecturer with the School of Telecommunications Engineering, Xidian University. Her research interests include pattern recognition, dimensionality reduction, sparse representation, and face recognition.

Xinbo Gao (Senior Member, IEEE) received the B.Eng., M.Sc., and Ph.D. degrees in electronic engineering, signal and information processing from Xidian University, Xi'an, China, in 1994, 1997, and 1999, respectively.

From 1997 to 1998, he was a Research Fellow with the Department of Computer Science, Shizuoka University, Shizuoka, Japan. From 2000 to 2001, he was a Postdoctoral Research Fellow with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong. Since 2001, he has been with the School of Electronic Engineering, Xidian University. He is currently a Cheung Kong Professor with the Ministry of Education of China, a Professor of Pattern Recognition and Intelligent System with Xidian University, and a Professor of Computer Science and Technology with Chongqing University of Posts and Telecommunications, Chongqing, China. He has published six books and around 300 technical articles in refereed journals and proceedings. His current research interests include image processing, computer vision, multimedia analysis, machine learning, and pattern recognition.

Prof. Gao is on the Editorial Boards of several journals, including *Signal Processing* (Elsevier) and *Neurocomputing* (Elsevier). He served as the General Chair/Co-Chair, a Program Committee Chair/Co-Chair, or a PC Member for around 30 major international conferences. He is a Fellow of the Institute of Engineering and Technology and the Chinese Institute of Electronics.