

Received May 11, 2020, accepted May 21, 2020, date of publication May 26, 2020, date of current version June 8, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2997755

Multi-View Spectral Clustering With Incomplete Graphs

WENZHANG ZHUGE¹, TINGJIN LUO¹, HONG TAO¹, CHENPING HOU¹, (Member, IEEE), AND DONGYUN YI^{1,2}

¹College of Liberal Arts and Science, National University of Defense Technology, Changsha 410073, China

²School of Mathematics and Computing Science, Hunan First Normal University, Changsha 410205, China

Corresponding author: Chenping Hou (hcpnudt@hotmail.com)

This work was supported in part by the NSF of China under Grant 61922087 and Grant 61906201, and in part by the NSF for Distinguished Young Scholars of Hunan Province under Grant 2019JJ20020.

ABSTRACT Traditional multi-view learning usually assumes each instance appears in all views. However, in real-world applications, it is not an uncommon case that a number of instances suffer from some view samples missing. How to effectively cluster this kind of partial multi-view data has attracted much attention. In this paper, we propose an incomplete multi-view clustering method, namely Multi-view Spectral Clustering with Incomplete Graphs (MSCIG), which connects processes of spectral embedding and similarity matrix completion to achieve better clustering performance. Specifically, MSCIG recovers missing entries of each similarity matrix based on multiplications of a common representation matrix and corresponding view-specific representation matrix, and in turn learns these representation matrices based on the complete similarity matrices. Besides, MSCIG adopts the p -th root integration strategy to incorporate losses of multiple views, which characterizes the contributions of different views. Moreover, we develop an iterative algorithm with proved convergence to solve the resultant problem of MSCIG, which updates the common representation matrix, view-specific representation matrices, similarity matrices, and view weights alternatively. We conduct extensive experiments on 9 benchmark datasets to compare the proposed algorithm with existing state-of-the-art incomplete multi-view clustering methods. Experimental results validate the effectiveness of the proposed algorithm.

INDEX TERMS Partial multi-view data, multi-view spectral clustering, incomplete graphs, similarity matrix completion.

I. INTRODUCTION

With the development of data collection techniques, in many practical applications such as image retrieval and cross-language document categorization, data appear in multiple modalities or naturally come from multiple sources, which are named as multi-view data. To effectively deal with multi-view data, multi-view learning has become into a hot area of research in last decades [1]–[6]. A common assumption adopted by conventional multi-view learning methods is that each data point appears in all views. However, in real applications, it is not a uncommon case that a number of instances suffer from some view representations missing, which results in partial multi-view data. For example, in web image retrieval, there may be some images with no text descriptions, and due to the invalid URL, the image itself

may be inaccessible. Another example is cross-language document categorization, it is often the case that a document is translated into several but not all languages. Since it is an often case that every view of this kind of data suffers from some samples missing, traditional single-view or multi-view clustering methods may fail to obtain the clustering results of all data points directly. Therefore, how to effectively cluster partial multi-view data has become an important research direction in recent years.

The early researches focus on the clustering task of incomplete two-view data. The work in [7] requires that there is at least a complete view, and proposes a method which constructs a full kernel on an incomplete view with the help of the complete view. To deal with data with two incomplete views, the work in [8] factorizes view-specific instances and complete instances into a common learned latent subspace to obtain their homogeneous and comparable representations. Based on [8], the algorithms proposed in [9], [10]

The associate editor coordinating the review of this manuscript and approving it for publication was Corrado Mencar¹.

introduce the adaptive graph regularization terms to capture local structure in the common representation learning. The works proposed in [11]–[13] utilize the similar strategy with [8] to design incomplete multi-view clustering algorithms. One limitation of this strategy is that it requires that each instance is complete or only appears in one view. However, for data with more than two incomplete views, it is a common case that a number of instances are present in more than one view but not all views.

To cluster multi-view data with arbitrary incomplete views, several algorithms are designed based on weighted matrix factorization. The method proposed in [14] first fills in the missing samples with average values of present samples on each view independently, and then learns a common representation matrix based on weighted nonnegative matrix factorization which assigns smaller weights for these filled samples. The approach proposed in [15] introduces a zero-one diagonal matrix for each view to eliminates the influence of its missing samples, and learns a common representation matrix based on semi-nonnegative matrix factorization. Moreover, the methods proposed in [16], [17] introduce zero-one weight matrices for multiple views to distinguishes their present elements from missing elements, and thus they can deal with a more general incomplete problem. Although these matrix factorization-based methods adopt different norms, constraints, or regularization terms, they are essentially linear methods, which limits their ability to disclose the non-linear structure of data. To address this issue, several kernel-based and graph-based methods are proposed. The work [18] recovers the missing entries of incomplete kernels by measuring both between-view and within-view relationships among kernel values. The work [19] extends multiple kernel k -means and considers mutual kernel completion to integrate kernel imputation and representation learning. The approach proposed in [20] first fills in the missing entries of graph matrices with the average of the columns, and then uses a co-training strategy to recover the representations of missing samples of each view and learn a common representation matrix. The method proposed in [21] jointly performs graph construction and common representation learning based on incomplete view representations. The work in [22] performs representation learning and clustering simultaneously with multiple incomplete similarity matrices. The approach proposed in [23] first fills in the missing entries of similarity matrices with the average of corresponding certain entries, and then learns views weights to combine a common graph matrix.

In this paper, we propose Multi-view Spectral Clustering with Incomplete Graphs (MSCIG) to handle incomplete multi-view clustering problem. Different from previous graph-based methods [20], [23] which adopt separate steps to fill in the missing entries of similarity matrices and learn a common representation matrix, the proposed MSCIG aims to achieve better clustering performance by integrating the processes of both imputation and representation learning seamlessly, which makes low dimensional

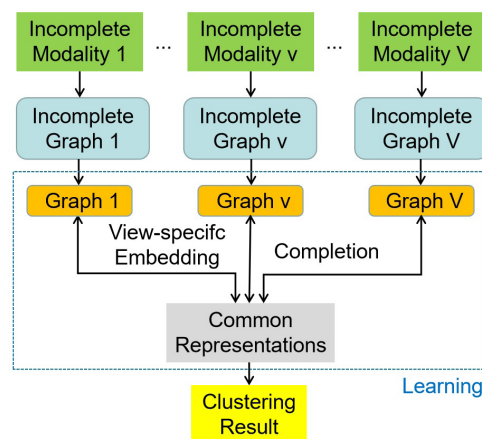


FIGURE 1. The flowchart of MSCIG.

representations guide the imputation of missing graph elements, and the completed graphs in turn influence the subsequent representation learning. Concretely, MSCIG employs the common representation matrix to take part in the imputation of missing entries of similarity matrices. To enable the imputation to explore the diverse information of views, a view-specific representation matrix is introduced to help the imputation of missing entries of each similarity matrix, and a regularization term is design to maximize the alignment between the common representation matrix and the view-specific representation matrix. To characterize contributions of multiple views, MSCIG adopts the p -th root integration to incorporate the losses of multiple views. To solve the resultant non-convex problem of MSCIG, we introduce another problem with explicit view weight factors. And we propose an iterative and alternative algorithm for optimization. In our proposed algorithm, the representation learning, imputation of missing graph entries, and view weight learning are alternatively performed until convergence. Fig. 1 displays the flowchart of MSCIG. The clustering performance of the proposed algorithm is evaluated by systematical experimental study on nine multi-view datasets. The experimental results indicate that our proposed algorithm achieves better clustering performance than the compared state-of-the-art incomplete multi-view clustering methods.

The rest of the paper is organized as follows. Section II introduces the background. Section III and IV introduce the formulation and the optimization algorithm of the proposed MSCIG, respectively. Section V gives some analysis about the proposed algorithm. Experimental results on nine multi-view datasets are displayed in Section VI. Finally, we give the conclusion of this paper in Section VII.

II. BACKGROUND

Throughout the paper, matrices are written as boldface uppercase letters, and vectors are written as boldface lowercase letters. For matrix $\mathbf{M} = [m_{ij}]$, its i -th row is denoted by \mathbf{m}_i . The transpose, the Frobenius norm, and the trace of matrix \mathbf{M} are denoted by \mathbf{M}^T , $\|\mathbf{M}\|_F$, and $tr(\mathbf{M})$, respectively. The 2-norm of vector \mathbf{m}_i is denoted by $\|\mathbf{m}_i\|$.

A. PROBLEM SETTING

A dataset with n instances can be represented by $\mathbf{X} = [\mathbf{x}_1; \dots; \mathbf{x}_n] \in \mathbb{R}^{n \times d}$, where $\mathbf{x}_i \in \mathbb{R}^{1 \times d}$ is the i -th instance. Suppose they are collected from V views, and therefore, each instance has V view samples, i.e., $\mathbf{x}_i = [\mathbf{x}_i^{(1)}, \dots, \mathbf{x}_i^{(V)}]$, where $\mathbf{x}_i^{(v)} \in \mathbb{R}^{1 \times d^{(v)}}$ is the i -th sample of the v -th view and $d = \sum_{v=1}^V d^{(v)}$. We use $\mathbf{X}^{(v)} = [\mathbf{x}_1^{(v)}; \dots; \mathbf{x}_n^{(v)}] \in \mathbb{R}^{n \times d^{(v)}}$ to collect all v -th view samples.

For an incomplete multi-view dataset, every view may suffer from some view samples missing. To identify the present samples from missing samples of each views, we introduce a indicator matrix $\mathbf{M} \in \{0, 1\}^{n \times V}$, and its (v, i) -th element m_{vi} is defined as follows:

$$m_{vi} = \begin{cases} 0, & \text{if } \mathbf{x}_i^{(v)} \text{ is missing.} \\ 1, & \text{otherwise.} \end{cases} \quad (1)$$

Incomplete multi-view clustering aims to group $\{\mathbf{x}_i\}_{i=1}^n$ into C clusters by utilizing multiple views.

B. CLASSIC NORMALIZED CUT AND MOTIVATION

Given an one-view dataset $\mathbf{X} = [\mathbf{x}_1; \dots; \mathbf{x}_n] \in \mathbb{R}^{n \times d}$, we can construct a graph similarity matrix $\mathbf{W} = [w_{ij}] \in \mathbb{R}^{n \times n}$ to measure the relationships between each pair of data points, where w_{ij} reflects the similarity between \mathbf{x}_i and \mathbf{x}_j . Based on \mathbf{W} , the Laplacian matrix $\mathbf{L} \in \mathbb{R}^{n \times n}$ is calculated by $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where \mathbf{D} is a diagonal matrix with the i -th entry $d_{ii} = \sum_{j=1}^n w_{ij}$. The objective of relaxed Normalized Cut (Ncut) is

$$\min_{\mathbf{F}^T \mathbf{F} = \mathbf{I}_C} \text{tr}(\mathbf{F}^T \mathbf{D}^{-\frac{1}{2}} \mathbf{L} \mathbf{D}^{-\frac{1}{2}} \mathbf{F}) \quad (2)$$

where $\mathbf{F} \in \mathbb{R}^{n \times C}$ denotes the relaxed variable. Since $\mathbf{L} = \mathbf{D} - \mathbf{W}$, the problem (2) is equivalent to

$$\max_{\mathbf{F}^T \mathbf{F} = \mathbf{I}_C} \text{tr}(\mathbf{F}^T \mathbf{D}^{-\frac{1}{2}} \mathbf{W} \mathbf{D}^{-\frac{1}{2}} \mathbf{F}) \quad (3)$$

The problem (3) is further equivalent to

$$\min_{\mathbf{F}^T \mathbf{F} = \mathbf{I}_C} \|\mathbf{A} - \mathbf{F} \mathbf{F}^T\|_F^2 \quad (4)$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is the normalized similarity matrix and computed by $\mathbf{A} = \mathbf{D}^{-\frac{1}{2}} \mathbf{W} \mathbf{D}^{-\frac{1}{2}}$.

In (4), the normalized similarity matrix \mathbf{A} is complete. For an incomplete V -view dataset, we can construct $\{\mathbf{A}^{(v)}\}_{v=1}^V$ on each view independently, but each of $\mathbf{A}^{(v)}$ may suffer from some information missing. Inspired by the objective of (4), in next section, we propose a method, which predict the missing values of $\{\mathbf{A}^{(v)}\}_{v=1}^V$ and learn a common representation matrix \mathbf{F} simultaneously.

III. PROPOSED METHOD

To explore the complementary and diverse information of multiple views, we construct a graph similarity matrix on each view independently to reflect the similarity relationships between pair of view samples. The v -th view graph matrix is denoted by $\mathbf{A}^{(v)} = [a_{ij}^{(v)}] \in \mathbb{R}^{n \times n}$ where its (i, j) -th element

$a_{ij}^{(v)}$ measures the similarity between $\mathbf{x}_i^{(v)}$ and $\mathbf{x}_j^{(v)}$. Since the v -th view may suffer from some samples missing, $a_{ij}^{(v)}$ can be computed as follows

$$a_{ij}^{(v)} = \begin{cases} f(\mathbf{x}_i^{(v)}, \mathbf{x}_j^{(v)}), & \text{if } m_{vi}m_{vj} = 1. \\ NaN, & \text{otherwise.} \end{cases} \quad (5)$$

where $f(\mathbf{x}_i^{(v)}, \mathbf{x}_j^{(v)})$ can be a traditional similarity computation method such as [24], [25]. *NaN* denotes ‘‘not a number’’, and can be regarded as a invalid number, which means the information of $a_{ij}^{(v)}$ is inaccessible. To eliminate the influence of *NaN* in the subsequent calculating, we define that $0 \cdot NaN = NaN \cdot 0 = 0$. Since $n^{(v)}$ samples appear in the v -th view, $n^{(v)} \times n^{(v)}$ elements of $\mathbf{A}^{(v)}$ are certain, and the rest elements are uncertain.

To integrate processes of both similarity matrix completion and clustering, based on these incomplete similarity matrices $\{\mathbf{A}^{(v)}\}_{v=1}^V$, we aim to learn complete similarity matrices $\{\mathbf{S}^{(v)}\}_{v=1}^V$ and a common relaxed cluster indicator matrix $\mathbf{F} = [\mathbf{f}_1; \dots; \mathbf{f}_n] \in \mathbb{R}^{n \times C}$ simultaneously, where $\mathbf{S}^{(v)} \in \mathbb{R}^{n \times n}$ is the complete similarity matrix of $\mathbf{A}^{(v)}$. $\mathbf{f}_i = [f_{i1}, \dots, f_{iC}]$ is the label vector of the i -th instance \mathbf{x}_i , and its c -th element f_{ic} satisfy the following constraint

$$f_{ic} = \begin{cases} \frac{1}{\sqrt{n(c)}}, & \text{if } \mathbf{x}_i \text{ belongs to the } c - \text{th cluster.} \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

where $n(c)$ is the number of the instances belonging to the c -th cluster ($c = 1, \dots, C$). Based on \mathbf{F} , the similarity information can be reflected by $\mathbf{F} \mathbf{F}^T$. Therefore, the optimization problem can be written as

$$\begin{aligned} \min_{\mathbf{F}, \{\mathbf{S}^{(v)}\}_{v=1}^V} & \sum_{v=1}^V \|\mathbf{S}^{(v)} - \mathbf{F} \mathbf{F}^T\|_F^2 \\ \text{s.t. } & \mathbf{F} \in \mathcal{C}_F, \Theta^{(v)} \odot \mathbf{S}^{(v)} = \Theta^{(v)} \odot \mathbf{A}^{(v)}, \quad (\forall v) \end{aligned} \quad (7)$$

where \mathcal{C}_F is the constraint of \mathbf{F} in (6), $\Theta^{(v)} = \mathbf{m}_v^T \mathbf{m}_v \in \{0, 1\}^{n \times n}$ distinguish certain elements of $\mathbf{A}^{(v)}$ from its uncertain elements, and \odot denotes the element-wise product between two matrices. The constraint $\Theta^{(v)} \odot \mathbf{S}^{(v)} = \Theta^{(v)} \odot \mathbf{A}^{(v)}$ makes sure that $\mathbf{S}^{(v)}$ maintains the known entries of $\mathbf{A}^{(v)}$. However, solving the optimization problem (7) under such constraint of \mathbf{F} is NP-hard. Thus, we relax the constraint of \mathbf{F} as $\mathbf{F}^T \mathbf{F} = \mathbf{I}_C$, and the resultant optimization problem becomes

$$\begin{aligned} \min_{\{\mathbf{S}^{(v)}\}_{v=1}^V, \mathbf{F}} & \sum_{v=1}^V \|\mathbf{S}^{(v)} - \mathbf{F} \mathbf{F}^T\|_F^2 \\ \text{s.t. } & \mathbf{F}^T \mathbf{F} = \mathbf{I}_C, \Theta^{(v)} \odot \mathbf{S}^{(v)} = \Theta^{(v)} \odot \mathbf{A}^{(v)}, \\ & \mathbf{S}^{(v)} \geq 0, \quad (\forall v) \end{aligned} \quad (8)$$

The nonnegative constraint $\mathbf{S}^{(v)} \geq 0$ is imposed to ensure that $\mathbf{S}^{(v)}$ is a similarity matrix. Although (8) connects processes of both graph imputation and common representation learning, its performance can be improved due to the following reasons: 1) all uncertain elements of $\{\mathbf{A}^{(v)}\}_{v=1}^V$ are imputed

based on the corresponding elements of $\mathbf{F}\mathbf{F}^T$, which limits the flexibility of $\{\mathbf{S}^{(v)}\}_{v=1}^V$ to exploit the view-specific information; 2) different views are treated equally, and they play the same important roles in the learning stage of \mathbf{F} , which may suffer from performance degeneration when there are unreliable views.

To design a more reasonable model, we introduce view-specific representation matrices $\{\mathbf{F}^{(v)}\}_{v=1}^V$, where the v -th view representation matrix $\mathbf{F}^{(v)} = [\mathbf{f}_1^{(v)}; \dots; \mathbf{f}_n^{(v)}] \in \mathbb{R}^{n \times C}$ satisfies $(\mathbf{F}^{(v)})^T \mathbf{F}^{(v)} = \mathbf{I}_C$, and estimate the uncertain elements of $\mathbf{A}^{(v)}$ based on $\mathbf{F}^{(v)}\mathbf{F}^T$. The v -th view loss $\mathcal{J}(\mathbf{S}^{(v)}, \mathbf{F}^{(v)}, \mathbf{F})$ are defined as follows

$$\mathcal{J}(\mathbf{S}^{(v)}, \mathbf{F}^{(v)}, \mathbf{F}) = \|\mathbf{S}^{(v)} - \mathbf{F}^{(v)}\mathbf{F}^T\|_F^2 + \lambda \mathcal{R}(\mathbf{F}^{(v)}, \mathbf{F}) \quad (9)$$

where $\lambda > 0$ is a balanced parameter, where $\mathcal{R}(\mathbf{F}^{(v)}, \mathbf{F})$ is a regularization term, which enforces $\mathbf{F}^{(v)}$ and \mathbf{F} to be approximated. In this paper, we utilize $\mathcal{R}(\mathbf{F}^{(v)}, \mathbf{F}) = \|\mathbf{F}^{(v)} - \mathbf{F}\|_F^2$. And for this regularization term, we give an explanation by the following proposition.

Proposition 1: With constraints $(\mathbf{F}^{(v)})^T \mathbf{F}^{(v)} = \mathbf{I}_C$ and $\mathbf{F}^T \mathbf{F} = \mathbf{I}_C$, to minimize the objective $\|\mathbf{F}^{(v)} - \mathbf{F}\|_F^2$ is equivalent to maximize the objective $tr(\mathbf{F}^T \mathbf{F}^{(v)})$, which makes $\mathbf{F}^{(v)}$ and \mathbf{F} maximally align with each other.

Proof: Since $tr(\mathbf{F}^T \mathbf{F}^{(v)}) = tr[(\mathbf{F}^{(v)})^T \mathbf{F}]$, we have

$$\begin{aligned} \|\mathbf{F}^{(v)} - \mathbf{F}\|_F^2 &= tr[(\mathbf{F}^{(v)} - \mathbf{F})^T (\mathbf{F}^{(v)} - \mathbf{F})] \\ &= tr[(\mathbf{F}^{(v)})^T \mathbf{F}^{(v)}] + tr(\mathbf{F}^T \mathbf{F}) - 2tr(\mathbf{F}^T \mathbf{F}^{(v)}) \\ &= 2tr(\mathbf{I}_C) - 2tr(\mathbf{F}^T \mathbf{F}^{(v)}) \end{aligned} \quad (10)$$

Since $tr(\mathbf{I}_C) = C$ is a constant, we complete the proof. \square

After proposing the loss functions of multiple views, a rough and simple integration strategy to obtain the multi-view formulation is to sum up them directly, i.e., $\sum_{v=1}^V \mathcal{J}(\mathbf{S}^{(v)}, \mathbf{F}^{(v)}, \mathbf{F})$. However, this strategy ignores the various importance of views. A more reasonable integration way is $\sum_{v=1}^V \omega^{(v)} \mathcal{J}(\mathbf{S}^{(v)}, \mathbf{F}^{(v)}, \mathbf{F})$, where $\omega^{(v)}$ is the v -th view weight factor. Since the view weight factors $\{\omega^{(v)}\}_{v=1}^V$ are usually difficult to be predetermined, inspired by [26], we adopt the p -th root integration strategy with $0 < p \leq 1$ to generate the common objective function, i.e., $\sum_{v=1}^V \mathcal{J}(\mathbf{S}^{(v)}, \mathbf{F}^{(v)}, \mathbf{F})^p$, which utilizes an implicit way to automatically assign suitable view weights for each view according their losses. As a result, the proposed MSCIG can be formulated as follows

$$\begin{aligned} \min_{\Upsilon} \sum_{v=1}^V \sqrt[q]{\|\mathbf{S}^{(v)} - \mathbf{F}^{(v)}\mathbf{F}^T\|_F^2 + \lambda \|\mathbf{F}^{(v)} - \mathbf{F}\|_F^2} \\ \text{s.t. } \mathbf{F}^T \mathbf{F} = \mathbf{I}_C, \Theta^{(v)} \odot \mathbf{S}^{(v)} = \Theta^{(v)} \odot \mathbf{A}^{(v)}, \\ \mathbf{S}^{(v)} \geq 0, (\mathbf{F}^{(v)})^T \mathbf{F}^{(v)} = \mathbf{I}_C, (\forall v) \end{aligned} \quad (11)$$

where $\Upsilon = \{\{\mathbf{S}^{(v)}, \mathbf{F}^{(v)}\}_{v=1}^V, \mathbf{F}\}$ collects all variables, and $q = \frac{1}{p}$ is a parameter. The model of MSCIG in (11) admits the following advantages: 1) our objective function well targets the ultimate goal, i.e., clustering, by integrating imputation and spectral embedding into a unified learning framework;

2) our formulation utilizes $\{\mathbf{F}^{(v)}\}_{v=1}^V$ to help \mathbf{F} to complete uncertain elements of each $\mathbf{S}^{(v)}$, which enables $\{\mathbf{S}^{(v)}\}_{v=1}^V$ to capture view-specific information on their estimated elements; 3) our formulation weights different views automatically, which makes views with smaller losses play more important roles in the learning of common representation matrix \mathbf{F} ; 4) our method does not require a complete view or an instance present in all views, which is different from some previous methods.

IV. OPTIMIZATION ALGORITHM

When $0 < p < 1$, the p -th root integration strategy makes the problem (11) difficult to solve directly. In this paper, we obtain the solution of the problem (11) by solving the following problem with a new introduced variable vector $\alpha = [\alpha_1, \dots, \alpha_V]^T \in \mathbb{R}^V$:

$$\begin{aligned} \min_{\Phi} \sum_{v=1}^V (\alpha_v)^\gamma [\|\mathbf{S}^{(v)} - \mathbf{F}^{(v)}\mathbf{F}^T\|_F^2 + \lambda \|\mathbf{F}^{(v)} - \mathbf{F}\|_F^2] \\ \text{s.t. } \Theta^{(v)} \odot \mathbf{S}^{(v)} = \Theta^{(v)} \odot \mathbf{A}^{(v)}, \quad \mathbf{S}^{(v)} \geq 0, \alpha_v \geq 0 \\ (\mathbf{F}^{(v)})^T \mathbf{F}^{(v)} = \mathbf{I}_C, (\forall v), \quad \mathbf{F}^T \mathbf{F} = \mathbf{I}_C, \quad \alpha^T \mathbf{1}_V = 1 \end{aligned} \quad (12)$$

where $\Phi = \{\{\mathbf{S}^{(v)}, \mathbf{F}^{(v)}\}_{v=1}^V, \alpha, \mathbf{F}\}$ collects all variables of the problem (12), and $\gamma = \frac{p-1}{p}$ is a parameter. In (12), the explicit v -th view weight factor $\omega^{(v)}$ can be calculated as $\omega^{(v)} = (\alpha_v)^\gamma$, which is different from (11), where no view weight factors are explicitly defined.

A. UPDATING RULES

To solve the problem (12), an effective algorithm is designed to update four groups of variables iteratively and alternatively.

1) UPDATE $\{\mathbf{F}^{(v)}\}_{v=1}^V$ AND FIX OTHERS

With $\{\mathbf{S}^{(v)}\}_{v=1}^V$, α , and \mathbf{F} fixed, the objective function (12) w.r.t $\{\mathbf{F}^{(v)}\}_{v=1}^V$ is additive and the constraints w.r.t $\{\mathbf{F}^{(v)}\}_{v=1}^V$ are separable, and thus each $\mathbf{F}^{(v)}$ can be updated by solving the following problem

$$\min_{(\mathbf{F}^{(v)})^T \mathbf{F}^{(v)} = \mathbf{I}_C} \|\mathbf{S}^{(v)} - \mathbf{F}^{(v)}\mathbf{F}^T\|_F^2 + \lambda \|\mathbf{F}^{(v)} - \mathbf{F}\|_F^2 \quad (13)$$

Considering the constraint $(\mathbf{F}^{(v)})^T \mathbf{F}^{(v)} = \mathbf{I}_C$ and removing the constant terms, the problem (13) is equivalent to the following problem

$$\max_{(\mathbf{F}^{(v)})^T \mathbf{F}^{(v)} = \mathbf{I}_C} tr[(\mathbf{F}^{(v)})^T (\mathbf{S}^{(v)} \mathbf{F} + \lambda \mathbf{F})] \quad (14)$$

To solve the problem (14), we introduce the following proposition to obtain its close-form solution.

Proposition 2: Suppose the compact SVD of matrix $\mathbf{B}^{(v)} = \mathbf{S}^{(v)} \mathbf{F} + \lambda \mathbf{F}$ is $\mathbf{B}^{(v)} = \mathbf{U}_B^{(v)} \Sigma_B^{(v)} (\mathbf{V}_B^{(v)})^T \in \mathbb{R}^{n \times C}$, where $\mathbf{U}_B^{(v)} \in \mathbb{R}^{n \times C}$, $\Sigma_B^{(v)} \in \mathbb{R}^{C \times C}$, and $\mathbf{V}_B^{(v)} \in \mathbb{R}^{C \times C}$. Then the optimal $\mathbf{F}^{(v)}$ to the problem (14) is

$$\mathbf{F}^{(v)} = \mathbf{U}_B^{(v)} (\mathbf{V}_B^{(v)})^T \quad (15)$$

Proof: Let the full SVD of $\mathbf{B}^{(v)} = \mathbf{S}^{(v)}\mathbf{F} + \lambda\mathbf{F}$ be $\mathbf{B}^{(v)} = \mathbf{U}^{(v)}\Sigma^{(v)}(\mathbf{V}^{(v)})^T$ with $\mathbf{U}^{(v)} \in \mathbb{R}^{n \times n}$, $\Sigma^{(v)} \in \mathbb{R}^{n \times C}$, and $\mathbf{V}^{(v)} \in \mathbb{R}^{C \times C}$. Therefore, the objective of (14) can be represented as

$$\begin{aligned} \text{tr}[(\mathbf{F}^{(v)})^T \mathbf{B}^{(v)}] &= \text{tr}[(\mathbf{F}^{(v)})^T \mathbf{U}^{(v)} \Sigma^{(v)} (\mathbf{V}^{(v)})^T] \\ &= \text{tr}[\Sigma^{(v)} (\mathbf{V}^{(v)})^T (\mathbf{F}^{(v)})^T \mathbf{U}^{(v)}] \\ &= \text{tr}[\Sigma^{(v)} \mathbf{Z}^{(v)}] \\ &= \sum_i \sigma_{ii}^{(v)} z_{ii}^{(v)} \end{aligned} \quad (16)$$

where $\mathbf{Z}^{(v)} = (\mathbf{V}^{(v)})^T (\mathbf{F}^{(v)})^T \mathbf{U}^{(v)} \in \mathbb{R}^{C \times n}$, $\sigma_{ii}^{(v)}$ and $z_{ii}^{(v)}$ are the (i, i) -th elements of $\Sigma^{(v)}$ and $\mathbf{Z}^{(v)}$, respectively. Since $\mathbf{Z}^{(v)}(\mathbf{Z}^{(v)})^T = \mathbf{I}_C$, we can conclude that each element of $\mathbf{Z}^{(v)}$ satisfies $-1 \leq z_{ij}^{(v)} \leq 1$. On the other hand, since $\sigma_{ii}^{(v)}$ is a singular value, there is $\sigma_{ii}^{(v)} \geq 0$. As a result, we arrive that

$$\text{tr}[(\mathbf{F}^{(v)})^T \mathbf{B}^{(v)}] = \sum_i \sigma_{ii}^{(v)} z_{ii}^{(v)} \leq \sum_i \sigma_{ii}^{(v)} \quad (17)$$

The equality in (17) holds when all $z_{ii}^{(v)} = 1$. That is to say, $\text{tr}[(\mathbf{F}^{(v)})^T \mathbf{B}^{(v)}]$ reaches its maximum when $\mathbf{Z}^{(v)} = [\mathbf{I}_C, \mathbf{0}] \in \mathbb{R}^{C \times n}$. Since $\mathbf{Z}^{(v)} = (\mathbf{V}^{(v)})^T (\mathbf{F}^{(v)})^T \mathbf{U}^{(v)}$, the optimal $\mathbf{F}^{(v)}$ to the problem (14) can be written as

$$\mathbf{F}^{(v)} = \mathbf{U}^{(v)}[\mathbf{I}_C; \mathbf{0}](\mathbf{V}^{(v)})^T \quad (18)$$

Since (18) is based upon the full SVD of matrix $\mathbf{B}^{(v)}$, it can be rewritten as $\mathbf{F}^{(v)} = \mathbf{U}_B^{(v)}(\mathbf{V}_B^{(v)})^T$ via the compact SVD of matrix $\mathbf{B}^{(v)} = \mathbf{U}_B^{(v)}\Sigma_B^{(v)}(\mathbf{V}_B^{(v)})^T$. \square

2) UPDATE $\{\mathbf{S}^{(v)}\}_{v=1}^V$ AND FIX OTHERS

With α , \mathbf{F} , and $\{\mathbf{F}^{(v)}\}_{v=1}^V$ fixed, the relations among multiple views are decoupled. By removing the constant terms, each $\mathbf{S}^{(v)}$ can be updated individually by solving the following problem

$$\begin{aligned} \min_{\mathbf{S}^{(v)}} \|\mathbf{S}^{(v)} - \mathbf{F}^{(v)}\mathbf{F}^T\|_{\mathbf{F}}^2 \\ \text{s.t. } \Theta^{(v)} \odot \mathbf{S}^{(v)} = \Theta^{(v)} \odot \mathbf{A}^{(v)}, \quad \mathbf{S}^{(v)} \geq \mathbf{0} \end{aligned} \quad (19)$$

To solve the problem (19), we introduce a matrix $\mathbf{E}^{(v)} \in \mathbb{R}^{n \times n}$ calculated by $\mathbf{E}^{(v)} = \max(\mathbf{F}^{(v)}\mathbf{F}^T, \mathbf{0})$. And then we can write the optimal solution to the problem (19) as the following form

$$\mathbf{S}^{(v)} = \mathbf{E}^{(v)} + \Theta^{(v)} \odot (\mathbf{A}^{(v)} - \mathbf{E}^{(v)}) \quad (20)$$

From (20), we can see that the optimal $\mathbf{S}^{(v)}$ is a duplicated of $\mathbf{A}^{(v)}$ on its known entries, and the uncertain elements are imputed by those of $\mathbf{E}^{(v)}$.

3) UPDATE \mathbf{F} AND FIX OTHERS

With $\{\mathbf{F}^{(v)}\}_{v=1}^V$, $\{\mathbf{S}^{(v)}\}_{v=1}^V$, and α fixed, by considering the constraint $\mathbf{F}^T \mathbf{F} = \mathbf{I}_C$ and removing the constant term, the problem (12) w.r.t. \mathbf{F} can be transformed into

$$\max_{\mathbf{F}^T \mathbf{F} = \mathbf{I}_C} \text{tr}[\mathbf{F}^T \sum_{v=1}^V (\alpha_v)^\gamma [(\mathbf{S}^{(v)})^T \mathbf{F}^{(v)} + \lambda \mathbf{F}^{(v)}]] \quad (21)$$

Denote $\mathbf{C} = \sum_{v=1}^V (\alpha_v)^\gamma [(\mathbf{S}^{(v)})^T \mathbf{F}^{(v)} + \lambda \mathbf{F}^{(v)}] \in \mathbb{R}^{n \times C}$. Suppose that the compact SVD of \mathbf{C} is $\mathbf{C} = \mathbf{U}_C \Sigma_C (\mathbf{V}_C)^T$

with $\mathbf{U}_C \in \mathbb{R}^{n \times C}$, $\Sigma_C \in \mathbb{R}^{C \times C}$, and $\mathbf{V}_C \in \mathbb{R}^{C \times C}$. According to Proposition 2, the optimal \mathbf{F} to the problem (21) can be obtained by

$$\mathbf{F} = \mathbf{U}_C (\mathbf{V}_C)^T \quad (22)$$

4) UPDATE α AND FIX OTHERS

With \mathbf{F} , $\{\mathbf{F}^{(v)}\}_{v=1}^V$, and $\{\mathbf{S}^{(v)}\}_{v=1}^V$ fixed, we can compute the losses $\{\mathcal{J}(\mathbf{S}^{(v)}, \mathbf{F}^{(v)}, \mathbf{F})\}_{v=1}^V$ of multiple views by (9) accordingly. And the problem (12) w.r.t. α can be rewritten as the following form

$$\min_{\alpha_v \geq 0, \alpha^T \mathbf{1}_V = 1} \sum_{v=1}^V (\alpha_v)^\gamma \mathcal{J}(\mathbf{S}^{(v)}, \mathbf{F}^{(v)}, \mathbf{F}) \quad (23)$$

When $\gamma = 0$, i.e., $p = 1$, the p -root integration strategy degenerates into the simple adding integration strategy, and we can set all $\alpha_v = 1/V$. When $\gamma < 0$, i.e., $0 < p < 1$, the Lagrangian function of the problem (23) can be written as $\mathcal{L}_\mu = \sum_{v=1}^V (\alpha_v)^\gamma \mathcal{J}(\mathbf{S}^{(v)}, \mathbf{F}^{(v)}, \mathbf{F}) - \mu(\alpha^T \mathbf{1}_V - 1)$, where μ is the Lagrange multiplier. Setting the derivative of \mathcal{L}_μ with respect to α_v to zero and combining the constraint $\alpha^T \mathbf{1}_V = 1$, the closed-form solution of the problem (23) can be rewritten as follows

$$\alpha_v = \frac{[\mathcal{J}(\mathbf{S}^{(v)}, \mathbf{F}^{(v)}, \mathbf{F})]^{-\frac{1}{1-\gamma}}}{\sum_{v=1}^V [\mathcal{J}(\mathbf{S}^{(v)}, \mathbf{F}^{(v)}, \mathbf{F})]^{-\frac{1}{1-\gamma}}} \quad (24)$$

B. INITIALIZATION

According to the above four steps, we alternatively update $\{\mathbf{F}^{(v)}\}_V$, $\{\mathbf{S}^{(v)}\}_V$, \mathbf{F} as well as α , and repeat these procedures iteratively until the objective of the problem (12) converges. Since we adopt an alternative and iterative updating strategy, it is important to have a reasonable initialization. We initialize all $\alpha_v = 1/V$, which treats all views equally. For all $\mathbf{S}^{(v)}$, we initialize its element $s_{ij}^{(v)}$ according to the following way

$$s_{ij}^{(v)} = \begin{cases} a_{ij}^{(v)}, & \text{if } \theta_{ij}^{(v)} = 1. \\ \frac{\sum_{v=1}^V \theta_{ij}^{(v)} a_{ij}^{(v)}}{\sum_{v=1}^V \theta_{ij}^{(v)}}, & \text{otherwise.} \end{cases} \quad (25)$$

where $\theta_{ij}^{(v)}$ is the (i, j) -th element of $\Theta^{(v)}$. From (25), if $a_{ij}^{(v)}$ is uncertain, $s_{ij}^{(v)}$ is initialized by the average of corresponding certain elements of $\{a_{ij}^{(v)}\}_{v=1}^V$. After obtaining initialized $\{\mathbf{S}^{(v)}\}_{v=1}^V$, we obtain a fusion graph by adding them up directly, i.e., $\sum_{v=1}^V \mathbf{S}^{(v)}$, and then initialize \mathbf{F} by solving the following problem

$$\max_{\mathbf{F}^T \mathbf{F} = \mathbf{I}_C} \text{tr}[\mathbf{F}^T (\sum_{v=1}^V \mathbf{S}^{(v)}) \mathbf{F}] \quad (26)$$

In this paper, we solve the problem (26) by an iterative algorithm proposed in [27].

The entire optimization procedure of the optimization problems (11), and (12) are summarized in Algorithm 1.

Algorithm 1 Optimization of MSCIG

Input: Incomplete V view data $\{\mathbf{X}^{(v)}\}_{v=1}^V$, indicator matrix \mathbf{M} , parameters p , and λ .
Output: \mathbf{F} , α , $\{\mathbf{F}^{(v)}\}_{v=1}^V$, and $\{\mathbf{S}^{(v)}\}_{v=1}^V$.
Initialization:
 Construct $\{\mathbf{A}^{(v)}\}_{v=1}^V$ by (5).
 Initialize all $\alpha_v = 1/V$.
 Initialize $\{\mathbf{S}^{(v)}\}_{v=1}^V$ according to (25).
 Initialize \mathbf{F} by solving the problem (26).
While not converged **do**
 1: Update $\{\mathbf{F}^{(v)}\}_{v=1}^V$ with (15).
 2: Update $\{\mathbf{S}^{(v)}\}_{v=1}^V$ with (20).
 3: Update \mathbf{F} with (22).
 4: Update α with (24).
End While

V. ANALYSIS

In this section, we first analyze the convergence behavior of the proposed Algorithm 1, and then analyze its computational complexity.

A. CONVERGENCE BEHAVIOR

Proposition 3: Algorithm 1 will monotonically decrease the objective of problem (12) until it converges to a stationary point of the problem (12).

Proof: In each iteration, denote the updated $\mathbf{F}^{(v)}$, $\mathbf{S}^{(v)}$, \mathbf{F} , and α as $\tilde{\mathbf{F}}^{(v)}$, $\tilde{\mathbf{S}}^{(v)}$, $\tilde{\mathbf{F}}$, and $\tilde{\alpha}$, respectively. In Algorithm 1, the original problem (12) is divided into four groups of subproblems w.r.t. $\{\mathbf{F}^{(v)}\}_{v=1}^V$, $\{\mathbf{S}^{(v)}\}_{v=1}^V$, \mathbf{F} , and α , respectively. Since $\tilde{\mathbf{F}}^{(v)}$, $\tilde{\mathbf{S}}^{(v)}$, $\tilde{\mathbf{F}}$, and $\tilde{\alpha}$ are the optimal solutions to their corresponding subproblems, Algorithm 1 decreases the objective of the problem (12) alternatively and iteratively. Since the objective of the problem (12) is lower bounded by 0, Algorithm will converge. Denote the converged $\mathbf{F}^{(v)}$, $\mathbf{S}^{(v)}$, \mathbf{F} , and α as $\hat{\mathbf{F}}^{(v)}$, $\hat{\mathbf{S}}^{(v)}$, $\hat{\mathbf{F}}$, and $\hat{\alpha}$, respectively. In the convergence, $\{\{\hat{\mathbf{F}}^{(v)}, \hat{\mathbf{S}}^{(v)}\}_{v=1}^V, \hat{\mathbf{F}}, \hat{\alpha}\}$ satisfies the KKT conditions of the problem (12), and thus it is a stationary point of the problem (12). \square

Proposition 4: Each $\{\{\tilde{\mathbf{F}}^{(v)}, \tilde{\mathbf{S}}^{(v)}\}_{v=1}^V, \tilde{\mathbf{F}}\}$ in Algorithm 1 will monotonically decrease the objective of problem in each iteration, which makes $\{\{\hat{\mathbf{F}}^{(v)}, \hat{\mathbf{S}}^{(v)}\}_{v=1}^V, \hat{\mathbf{F}}\}$ be at least a stationary point of the problem (11).

Proof: According to Proposition 3, we obtain inequality $\sum_{v=1}^V \alpha_v^\gamma \mathcal{J}(\tilde{\mathbf{F}}^{(v)}, \tilde{\mathbf{S}}^{(v)}, \tilde{\mathbf{F}}) \leq \sum_{v=1}^V \alpha_v^\gamma \mathcal{J}(\mathbf{F}^{(v)}, \mathbf{S}^{(v)}, \mathbf{F})$. Considering (24) and $\gamma = \frac{p-1}{p}$, we can concluded that

$$\sum_{v=1}^V \frac{\mathcal{J}(\tilde{\mathbf{F}}^{(v)}, \tilde{\mathbf{S}}^{(v)}, \tilde{\mathbf{F}})}{\mathcal{J}(\mathbf{F}^{(v)}, \mathbf{S}^{(v)}, \mathbf{F})^{1-p}} \leq \sum_{v=1}^V \mathcal{J}(\mathbf{F}^{(v)}, \mathbf{S}^{(v)}, \mathbf{F})^p \quad (27)$$

On the other hand, we define function $g(x) = x^p$ and the supergradient of $g(x)$ is $g'(x) = px^{p-1}$. When $0 < p < 1$, $g(\mathcal{J}(\mathbf{F}^{(v)}, \mathbf{S}^{(v)}, \mathbf{F}))$ is a concave function in the domain of $\mathcal{J}(\mathbf{F}^{(v)}, \mathbf{S}^{(v)}, \mathbf{F})$, and $g'(\mathcal{J}(\mathbf{F}^{(v)}, \mathbf{S}^{(v)}, \mathbf{F})) = p\mathcal{J}(\mathbf{F}^{(v)}, \mathbf{S}^{(v)}, \mathbf{F})^{p-1}$. Based on the definition of supergradient

of a concave function, the following inequality holds:

$$g(\tilde{x}) - g(x) \leq g'(x)(\tilde{x} - x) \quad (28)$$

Replacing \tilde{x} , and x with $\mathcal{J}(\tilde{\mathbf{F}}^{(v)}, \tilde{\mathbf{S}}^{(v)}, \tilde{\mathbf{F}})$, and $\mathcal{J}(\mathbf{F}^{(v)}, \mathbf{S}^{(v)}, \mathbf{F})$, respectively, we can infer that

$$\begin{aligned} & \sum_{v=1}^V \mathcal{J}(\tilde{\mathbf{F}}^{(v)}, \tilde{\mathbf{S}}^{(v)}, \tilde{\mathbf{F}})^p - \sum_{v=1}^V \frac{p\mathcal{J}(\tilde{\mathbf{F}}^{(v)}, \tilde{\mathbf{S}}^{(v)}, \tilde{\mathbf{F}})}{\mathcal{J}(\mathbf{F}^{(v)}, \mathbf{S}^{(v)}, \mathbf{F})^{1-p}} \\ & \leq \sum_{v=1}^V \mathcal{J}(\mathbf{F}^{(v)}, \mathbf{S}^{(v)}, \mathbf{F})^p - \sum_{v=1}^V p\mathcal{J}(\mathbf{F}^{(v)}, \mathbf{S}^{(v)}, \mathbf{F})^p \end{aligned} \quad (29)$$

Based on (27) and (29), we arrive that

$$\sum_{v=1}^V \mathcal{J}(\tilde{\mathbf{F}}^{(v)}, \tilde{\mathbf{S}}^{(v)}, \tilde{\mathbf{F}})^p \leq \sum_{v=1}^V \mathcal{J}(\mathbf{F}^{(v)}, \mathbf{S}^{(v)}, \mathbf{F})^p \quad (30)$$

Therefore, each $\{\{\tilde{\mathbf{F}}^{(v)}, \tilde{\mathbf{S}}^{(v)}\}_{v=1}^V, \tilde{\mathbf{F}}\}$ will monotonically decrease the objective of the problem (11).

According to Proposition 3, $\{\{\hat{\mathbf{F}}^{(v)}, \hat{\mathbf{S}}^{(v)}\}_{v=1}^V, \hat{\mathbf{F}}, \hat{\alpha}\}$ is a stationary point of the problem (12). By replacing α_v according to (24), $\{\{\hat{\mathbf{F}}^{(v)}, \hat{\mathbf{S}}^{(v)}\}_{v=1}^V, \hat{\mathbf{F}}\}$ is a stationary point of the following problem

$$\min_{\Upsilon \in \mathcal{C}} \left[\sum_{v=1}^V \mathcal{J}(\mathbf{F}^{(v)}, \mathbf{S}^{(v)}, \mathbf{F})^p \right]^{\frac{1}{p}} \quad (31)$$

where \mathcal{C} collects the constraints of Υ in (11). When $0 < p \leq 1$, the problem (31) and the problem (11) are equivalent. Thus, $\{\{\hat{\mathbf{F}}^{(v)}, \hat{\mathbf{S}}^{(v)}\}_{v=1}^V, \hat{\mathbf{F}}\}$ satisfies the KKT conditions of the problem (11). \square

B. COMPUTATIONAL COMPLEXITY

In this subsection, we analyze the computational complexity of the proposed Algorithm 1.

In initialization, the computational complexity of the construction of $\{\mathbf{A}^{(v)}\}_{v=1}^V$ and the initialization of $\{\mathbf{S}^{(v)}\}_{v=1}^V$ is upper bounded by $O(n^2d)$; the initialization of \mathbf{F} costs $O(n^2V + (n^2C + C^3)\tau)$, where τ is the number of iterations of the algorithm proposed in [27] to solve (26).

In each iteration, the computational complexity to update $\{\mathbf{F}^{(v)}\}_{v=1}^V$ is $O(n^2CV + C^3V)$; the computational complexity to update $\{\mathbf{S}^{(v)}\}_{v=1}^V$ is upper bounded by $O(n^2CV)$; the computational complexity to update \mathbf{F} is $O(n^2CV + C^3)$; the computational complexity to update α is upper bounded by $O(n^2V + nCV)$.

Since $C \ll n$, the overall computational complexity of Algorithm 1 is $O(n^2(d + C\tau + CVt))$, where t is the number of iterations of the proposed Algorithm 1.

VI. EXPERIMENT

In this section, we conduct experiments to evaluate the performance of the proposed algorithm. First, we evaluate the effectiveness of MSCIG by comparing its clustering performance with some baselines. Then, we present experimental results about convergence behavior. Finally, we study the parameter sensitivity.

A. DATASET DESCRIPTION

The experiments are conducted on 9 datasets, i.e., 3Source,¹ BBCSport,² BBC,³ Dermatology,⁴ Forest,⁵ WebKB,⁶ Yale,⁷ MSRC-v1,⁸ and Digits.⁹ The detailed descriptions of these datasets are listed as follows.

- **3Source** includes 416 news stories collected from three online news sources: BBC, Reuters, and The Guardian. The stories belong to 6 classes, i.e., 104 business stories, 60 entertainment stories, 54 health stories, 49 politics stories and 89 sport stories and 60 tech stories. Since some news may not be reported by all sources, three views have 352, 302, and 294 present samples.
- **BBCSport** is composed of 737 news documents [28] divided into 5 classes: 101 athletics documents, 124 cricket documents, 265 football documents, 147 rugby documents, and 100 tennis documents. In [29], each raw document is split into 1–3 segments by merging consecutive paragraphs, and this process ensures that each segment has at least 200 words. Each segment is assigned to at most one view, and three views have 519, 531, and 513 present samples, respectively.
- **BBC** consists of 2225 news documents [28] belonging to 5 classes: 510 business documents, 386 entertainment documents, 417 politics documents, 511 sport documents, and 401 tech documents. In [29], each raw document is split into 1–3 segments, and each segment is assigned to one view. Its three views have 1828, 1832, and 1845 present samples, respectively.
- **Dermatology** is about erythemato-squamous diseases [30] which contains 336 instances from 6 classes: psoriasis (112 patients), seboreic dermatitis (61 patients), lichen planus (72 patients), pityriasis rosea (49 patients), cronic dermatitis (52 patients), and pityriasis rubra pilaris (20 patients). Each patient has 11 complete clinical attributes and 22 histopathological attributes.
- **Forest** includes multi-temporal remote sensing data for a forested area [31]. It is composed of 523 data points belonging to 4 classes, i.e., ‘Sugi’ forest, ‘Hinoki’ forest, ‘Mixed deciduous’ forest and ‘Other’ non-forest land. Each data point has 9 ASTER image bands features and 18 predicted spectral value features.
- **WebKB** is composed of 1051 web documents [32] divided into 2 classes: 230 Course pages and 821 Non-Course pages. Each instance has two representations: 2949 Fulltext features describe the textual contents on

the web page, while 334 Inlinks features records the anchor information on the hyperlinks.

- **Yale** consists of 165 face images belonging to 15 persons, and each person has 11 images. We extract three features for each image, i.e., 512 GIST, 256 Local Binary Pattern (LBP) and 168 Pyramid HOG (PHOG).
- **MSRC-v1** is composed of 240 images belonging to 8 categories. Following [33], we select 7 widely utilized categories, i.e., tree, building, airplane, cow, face, car, bicycle, and each category has 30 images. For each image, we extract six features, i.e., 1302 CENTRIST, 256 LBP, 48 Color Moment (CMT), 100 Histogram of Oriented Gradient (HOG), 200 SIFT and 512 GIST.
- **Digits** includes 2,000 instances for 0 to 9 ten digit classes, and each class has 200 instances. Six public features are available, i.e., 76 Fourier coefficients of the character shapes (FOU), 216 profile correlations (FAC), 64 Karhunen-love coefficients (KAR), 240 pixel averages in 2×3 windows (PIX), 47 Zernike moment (ZER) and 6 morphological (MOR) features.

B. EXPERIMENTAL SETUP

1) DATA PRECESSING

Since six datasets Dermatology, Forest, WebKB, Yale, MSRC-v1, and Dights are originally complete, we randomly delete some view samples to simulate the incomplete multi-view setting. Concretely, there is a probability to delete each $\mathbf{x}_i^{(v)}$, and we ensure that each data point \mathbf{x}_i has at least one $\mathbf{x}_i^{(v)}$ remaining. In the experiments, we tune the probability from 10% to 50% with a step 10% on these six datasets. And we can regard the probability as the incomplete example ratio (IER) of all views of the dataset.

2) BASELINES AND EXPERIMENTAL ENVIRONMENT

In the experiments, we compare the proposed MSCIG with six state-of-the-art methods: Multiple Incomplete views Clustering (MIC) [14], Multi-view Learning with Incomplete Views (MVL-IV) [16], Incomplete Multi-modality Grouping (IMG) [9], Doubly Aligned Incomplete Multi-view Clustering (DAIMC) [15], Incomplete Multiple Kernel K-means Algorithm with Mutual Kernel Completion [19] (IMKK-MKC) and Perturbation-oriented Incomplete multi-view Clustering (PIC) [23]. Since IMG is originally designed for incomplete two-view data, we extend it by [22], and thus the extended version can deal with data with arbitrary incomplete views. For fair comparison, all compared methods apply K-means on their corresponding common representations to extract clustering results. All the experiments are conducted by Matlab2016a on a laptop with Intel(R) Core(TM) i7-7500U CPU (2.9GHz) and 8.0GB RAM memory on the Windows 10 operating system.

3) PARAMETER DETERMINATION

In the experiments, we determine all hyper-parameters by grid-search, and record the clustering results with the best

¹<http://mlg.ucd.ie/datasets/3sources.html>

²<http://mlg.ucd.ie/datasets/segment.html>

³<http://mlg.ucd.ie/datasets/segment.html>

⁴<https://archive.ics.uci.edu/ml/datasets/dermatology>

⁵<https://archive.ics.uci.edu/ml/datasets/Forest+type+mapping>

⁶<http://www.cs.cmu.edu/afs/cs/project/theo-11/www/wwkb/>

⁷<http://vision.ucsd.edu/content/yale-face-database>

⁸<https://www.microsoft.com/en-us/research/project/>

⁹<https://archive.ics.uci.edu/ml/datasets/Multiple+Features>

TABLE 1. ACC (%) and NMI (%) comparisons on 3 datasets. STD (%) is in the parentheses. The first highest score is in bold. Symbols ‘●/○/◊’ denote that MSCIG is better/tied/worse than the corresponding method by the paired t-test with confidence level 0.05, respectively.

Data set	MIC	MVL-IV	IMG	DAIMC	IMKK-MKC	PIC	MSCIG
ACC							
3Source	57.74(4.06)●	76.20(1.76)●	78.65(0.75)●	73.13(2.28)●	84.09(0.86)●	89.28(1.12)○	89.76(0.13)
BBCSport	65.92(3.89)●	80.60(6.02)●	88.49(0.62)●	60.87(5.19)●	88.74(0.00)●	75.98(5.76)●	92.43(0.36)
BBC	76.00(6.35)●	85.38(5.59)●	86.72(0.36)●	85.36(1.50)●	91.64(0.11)●	79.07(6.72)●	92.58(0.10)
NMI							
3Source	56.97(1.48)●	64.19(1.24)●	66.79(0.92)●	60.66(2.11)●	71.36(1.13)●	74.87(1.54)○	75.95(0.22)
BBCSport	52.53(4.14)●	63.34(5.93)●	72.06(1.37)●	42.16(7.18)●	72.48(0.00)●	73.64(0.00)●	79.02(0.76)
BBC	58.10(6.05)●	68.01(7.06)●	69.61(0.48)●	68.77(1.12)●	76.76(0.21)●	70.85(0.00)●	78.83(0.20)

tuned parameters. For the baselines, we download the source codes from the authors’ websites and determine the searching ranges of their parameters according to the corresponding papers. For the proposed MSCIG, we construct the partial similarity matrices based on [24], tune the parameter p from 0.1 to 0.9 with a step 0.2, and tune the parameter λ from in the range of $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$. In the construction of partial similarity matrix, we set the neighbor number to be 15 for 3Source, BBC, and BBCSport, and we fix the neighbor number as 5 for Dermatology, Forest, WebKB, Yale, MSRC-v1, and Dights. For all compared algorithms which adopt iterative optimization strategy, we adopt the following stop criteria

$$\frac{f(t-1) - f(t)}{f(t-1)} < 10^{-5} \quad (32)$$

where $f(t)$ is the objective value in the t -th iteration.

4) EVALUATION METRIC

The clustering performance is measured in terms of accuracy (ACC) and the normalized mutual information (NMI). On datasets 3Source, BBC, and BBCSport, we run each compared method for 5 independent times. On each dataset of Dermatology, Forest, WebKB, Yale, MSRC-v1, and Dights, every time we create incomplete datasets with different IERs and repeat 5 independent times. And we report the average result with standard deviation (STD).

C. CLUSTERING PERFORMANCE

Table 1 displays the clustering results of seven compared methods with respect to both ACC and NMI on three incomplete multi-view datasets. Table 2 and Table 3 show the ACC and NMI comparisons of all seven compared methods on six datasets in incomplete multi-view setting with respect to different incomplete example ratios. According to results, we have the following observations.

As shown from the clustering results on Table 1 and the win/tie/loss counts in last rows of Table 2 and Table 3, the proposed MSCIG consistently achieves better or comparable performance than other methods in terms of both ACC and NMI. This may be because the integration of missing element imputation and representation learning enables the completed similarity matrices to measure the relationships between pair of view samples more accurately, and the common representation matrix to better reflect the underlying clustering structure of data.

MIC achieves the worst results on datasets 3Source, BBCSport, and BBC, and its performance degenerates more significantly than other methods with the increase of IER on datasets MSRC-v1, Digits, Yale, and Forest. This might be because that MIC simply utilizes the feature average of each view to fill the missing samples, which leads to a deviation especially when IER is large.

Comparing the performance of matrix factorization-based methods MVL-IV, IMG, and DAIMC, each of them achieves better results on several datasets but performs worse on other datasets. This can be owing to that these methods adopt different norms, regularization terms and constraints, which makes them good at clustering several datasets but bad at grouping others.

PIC achieves comparable results with MSCIG in a few cases but performs significantly worse on some datasets such as BBCSport, BBC, and WebKB, indicating that using the average of certain elements to fill the corresponding missing entries is not always an advisable manner to deal with incomplete similarity matrices. On datasets BBCSport, BBC, and Yale, IMKK-MKC achieves better performance than matrix factorization-based methods, and the possible reason is that IMKK-MKC can utilize the non-linear information. On the other hand, IMMK-MKC achieves the worst performance on Dights in some cases, and this might be because the construction of kernels ignores the local structure of data.

With the increase of IER, each method tends to achieve worse clustering performance in most cases. An exception is the clustering results on WebKB. Since both views of WebKB are sparse, the lack of certain samples may make methods fail to achieve reasonable clustering results. This also possibly explains why STD of MVL-IV, PIC, and MSCIG are high in some cases.

D. CONVERGENCE ANALYSIS

In Section V-A, we have proved the proposed Algorithm 1 will monotonically decrease the objective values of both (11) and (12) in each iteration until convergence.

In this subsection, we verify the convergence behavior of Algorithm 1 by conducting experiments on datasets 3Source, BBCSport3, Dermatology, and Forest. On Dermatology, and Forest, we set $IER = 30\%$. For MSCIG, we fix $p = 0.9$ and $\gamma = 10^{-2}$. We plot the curves of objective values of (11) and (12) in Fig. 2. As we can see, on these four datasets, both the objective values of (11) and (12) decrease as the iteration round increases and converge to fixed values within

TABLE 2. ACC (%) comparisons on 6 datasets with different IERs. STD (%) is in the parentheses. The first highest score is in bold. Symbols ‘●/○/◊’ denote that MSCIG is better/tied/worse than the corresponding method by the paired t-test with confidence level 0.05, respectively. The win/tie/loss counts are reported in the last row.

Data set	IER	MIC	MVL-IV	IMG	DAIMC	IMKK-MKC	PIC	MSCIG
Dermatology	10%	71.26(3.03)●	61.86(9.15)●	86.50(3.91)●	79.73(6.10)●	78.42(0.93)●	86.61(3.55)●	92.35(0.51)
	20%	65.08(4.68)●	52.02(11.76)●	86.34(3.68)◊	80.33(6.71)●	79.29(1.45)●	85.36(4.01)●	90.44(1.48)
	30%	65.57(6.06)●	57.60(16.63)●	83.01(3.73)●	79.23(1.64)●	80.66(1.88)●	83.44(4.67)●	90.55(1.25)
	40%	63.77(6.72)●	53.61(10.26)●	76.23(4.68)●	76.94(5.59)●	79.51(1.14)●	79.02(5.55)●	89.84(0.93)
	50%	59.56(4.12)●	54.59(14.39)●	72.51(3.52)●	75.85(5.30)●	77.76(0.69)●	81.75(3.33)◊	85.85(1.30)
Forest	10%	68.87(4.29)●	59.24(8.32)●	78.05(0.60)●	78.89(1.56)●	78.16(0.76)●	82.75(1.01)●	85.89(0.48)
	20%	68.76(4.48)●	67.80(4.97)●	77.86(0.44)●	77.17(1.14)●	76.90(2.05)●	82.41(0.73)●	85.70(0.56)
	30%	58.28(1.56)●	64.86(7.77)●	77.25(1.11)●	75.95(1.96)●	74.23(2.10)●	83.33(1.84)◊	85.85(1.81)
	40%	58.93(3.84)●	67.00(7.25)●	77.59(0.64)●	76.25(1.03)●	71.09(2.07)●	84.40(0.90)◊	85.62(0.91)
	50%	51.51(4.16)●	60.31(8.74)●	76.75(1.32)●	73.38(1.58)●	69.25(3.29)●	83.75(1.80)●	85.58(1.34)
WebKB	10%	79.71(2.44)●	79.39(8.72)◊	75.53(0.99)●	78.33(0.92)●	67.36(3.92)●	79.83(2.65)●	88.47(1.59)
	20%	78.14(0.04)●	70.81(10.54)●	76.14(0.82)●	80.06(0.65)●	64.32(6.00)●	78.63(1.93)●	88.28(2.25)
	30%	78.14(0.04)●	78.31(10.73)◊	78.61(1.07)●	80.40(0.65)●	66.26(5.80)●	77.93(0.00)●	87.67(3.46)
	40%	78.12(0.00)●	81.52(5.34)◊	78.63(0.79)●	81.71(0.28)●	67.12(4.61)●	77.98(0.23)●	87.90(1.85)
	50%	78.14(0.04)●	78.04(2.59)●	77.51(2.28)●	81.83(0.88)●	62.72(2.21)●	77.75(0.29)●	86.77(2.05)
Yale	10%	52.73(3.15)●	54.55(3.81)●	66.06(2.31)●	66.06(2.35)●	70.18(2.55)◊	65.33(2.89)●	70.18(1.17)
	20%	47.39(5.02)●	55.76(6.05)●	61.21(2.94)●	62.67(3.39)●	66.42(1.64)●	65.21(2.29)●	69.45(1.10)
	30%	42.30(3.98)●	52.85(5.41)●	56.61(3.87)●	55.52(3.05)●	64.36(2.69)◊	62.67(1.18)●	66.91(0.69)
	40%	37.70(2.66)●	46.18(4.99)●	50.91(3.59)●	50.67(5.31)●	58.91(4.36)◊	55.76(5.05)●	61.45(3.25)
	50%	31.27(1.10)●	42.79(3.96)●	48.36(4.87)●	44.85(4.41)●	51.27(4.12)●	51.27(2.45)●	57.09(3.13)
MSRC-v1	10%	71.33(3.00)●	83.24(7.09)●	78.19(3.52)●	76.19(2.71)●	85.14(2.37)●	87.62(4.19)●	95.24(1.12)
	20%	55.71(5.60)●	81.81(4.68)●	75.33(2.88)●	76.86(2.44)●	85.33(3.13)●	89.14(8.01)●	94.38(1.19)
	30%	40.67(4.35)●	74.48(2.04)●	72.19(4.28)●	70.95(3.05)●	83.90(5.09)●	91.43(5.43)●	93.33(0.48)
	40%	29.90(2.06)●	76.29(5.10)●	66.48(4.20)●	67.43(3.34)●	83.81(3.63)◊	88.76(2.71)◊	90.10(3.25)
	50%	23.90(1.23)●	67.52(8.39)●	64.57(4.47)●	65.43(3.31)●	75.14(4.62)●	82.29(4.32)◊	87.24(2.92)
Digits	10%	70.69(3.07)●	77.93(4.70)●	77.79(4.20)●	83.37(3.39)●	63.19(2.38)●	86.17(0.60)●	88.00(0.29)
	20%	62.90(3.21)●	75.99(4.95)●	73.30(5.82)●	83.81(3.82)◊	59.36(2.17)●	86.01(0.25)●	87.68(0.31)
	30%	55.58(1.56)●	76.94(2.29)●	73.54(4.95)●	82.46(3.63)◊	55.57(4.78)●	85.59(0.47)●	87.44(0.62)
	40%	47.37(1.87)●	73.94(4.04)●	78.68(2.88)●	80.60(4.21)●	50.73(5.23)●	85.26(2.16)◊	86.67(1.01)
	50%	40.58(2.92)●	74.57(5.25)●	70.23(2.23)●	70.37(4.89)●	46.87(5.94)●	82.35(2.46)●	85.32(1.20)
win/tie/lose		30/0/0	27/3/0	29/1/0	28/2/0	26/4/0	24/6/0	-

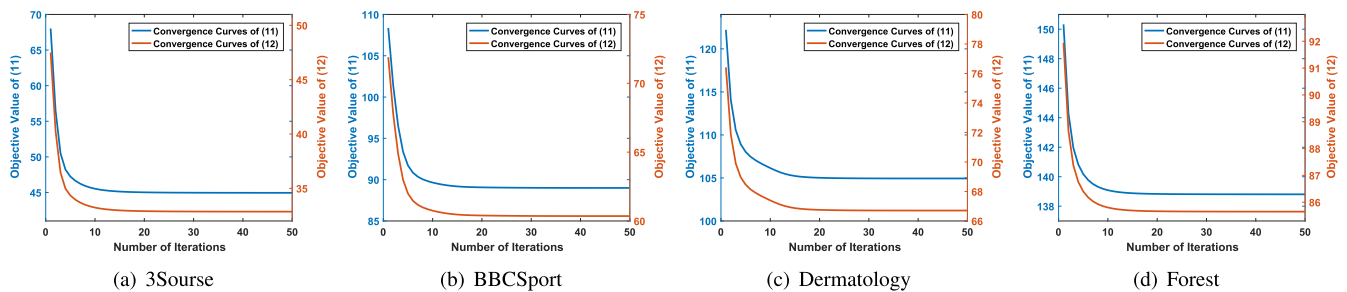


FIGURE 2. The convergence curves of (11) and (12) on 4 datasets.

30 iterations, indicating that Algorithm 1 has fine convergence speed.

E. PARAMETER STUDY

In this section, we study the influence of hyper-parameters on the performance of the proposed MSCIG. There are two nonnegative parameters in MSCIG, i.e., p and γ . We tune p in the range of $\{0.1, 0.3, 0.5, 0.7, 0.9\}$ while we vary γ from $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$. We conduct experiments on 4 datasets, i.e., 3Source, BBCSport, Dermatology, and Forest, and we set $IER = 30\%$ on datasets Dermatology, and Forest. Since NMI has similar tendency with ACC, Fig. 3 displays ACC results with varying p and γ .

As seen from the results, we have the following observations. On 3Source, MSCIG achieves acceptable performance in a wide range of parameters. On the other three

datasets, the value variation of γ has larger effect on the performance of MSCIG than that of p . Besides, MSCIG has different optimal parameters on 4 datasets, indicating the optimal parameters are data independent. These 4 datasets have different optimal combination of p and γ since their data characteristics are different.

F. SUMMARY OF EXPERIMENTAL RESULTS

In this section, we verify the effectiveness of the proposed MSCIG by comparing with six state-of-art incomplete multi-view clustering methods. The experiments are conducted on nine datasets, and both ACC and NMI results demonstrate the advantage of MSCIG. Besides, we plot objective value curves of (11), and (12) on four datasets. The results show that Algorithm 1 monotonically decreases the objective values of both (11), and (12) until convergence

TABLE 3. NMI (%) comparisons on 6 datasets with different IERs. STD (%) is in the parentheses. The first highest score is in bold. Symbols ‘●/○/◊’ denote that MSCIG is better/tied/worse than the corresponding method by the paired t-test with confidence level 0.05, respectively. The win/tie/loss counts are reported in the last row.

Data set	IER	MIC	MVL-IV	IMG	DAIMC	IMKK-MKC	PIC	MSCIG
Dermatology	10%	73.22(4.58)●	57.25(19.27)●	84.31(0.79)◊	82.34(1.10)●	64.14(1.26)●	80.25(1.42)●	84.45(1.21)
	20%	68.48(5.06)●	44.93(20.66)●	78.07(1.74)◊	78.92(1.39)◊	64.95(1.37)●	78.78(2.48)●	81.28(2.85)
	30%	66.61(5.15)●	46.96(19.18)●	74.04(2.02)●	76.26(1.26)●	66.56(1.54)●	78.02(3.57)●	81.65(1.98)
	40%	65.10(1.87)●	42.98(14.36)●	68.38(1.61)●	71.05(2.20)●	66.06(0.97)●	74.88(3.18)●	79.54(1.16)
	50%	59.34(1.71)●	42.92(15.56)●	63.33(2.36)●	67.73(3.24)●	65.36(1.08)●	75.59(2.64)◊	75.98(2.52)
Forest	10%	44.64(2.64)●	31.70(7.13)●	53.63(0.84)●	55.01(3.03)●	52.65(1.02)●	60.93(1.09)●	64.07(0.86)
	20%	44.91(3.64)●	39.91(5.65)●	53.30(1.31)●	51.19(2.44)●	50.59(2.38)●	60.27(0.86)●	63.61(1.21)
	30%	36.60(2.54)●	35.52(8.90)●	52.88(1.46)●	49.53(2.74)●	47.23(2.80)●	60.85(2.86)◊	63.47(2.99)
	40%	35.35(3.72)●	38.85(6.86)●	53.17(0.54)●	50.06(3.00)●	44.84(2.77)●	62.46(1.67)◊	63.15(1.28)
	50%	27.35(2.81)●	34.23(4.59)●	51.89(1.25)●	46.42(1.53)●	41.62(3.42)●	61.70(2.65)◊	63.59(2.38)
WebKB	10%	14.42(5.74)●	15.97(20.07)◊	5.41(1.46)●	6.00(4.32)●	5.31(1.34)●	9.36(12.00)●	39.66(6.17)
	20%	10.97(3.56)●	8.80(13.93)●	5.98(1.26)●	11.30(3.62)●	6.09(6.26)●	5.96(7.94)●	37.47(7.23)
	30%	8.26(1.28)●	16.80(13.73)◊	8.33(1.58)●	12.99(4.34)●	6.33(6.06)●	0.61(0.08)●	36.05(13.75)
	40%	4.23(2.16)●	19.06(10.22)●	8.13(1.36)●	17.90(2.27)●	6.85(3.06)●	1.28(1.20)●	37.71(4.90)
	50%	3.70(1.24)●	9.08(5.17)●	7.67(3.25)●	15.36(3.75)●	4.53(1.53)●	1.03(0.56)●	31.93(7.28)
Yale	10%	59.24(1.58)●	60.70(2.39)●	71.09(2.73)◊	70.52(1.04)●	73.12(1.56)◊	69.26(2.99)●	72.92(1.25)
	20%	53.23(3.55)●	60.30(5.20)●	66.85(2.01)●	66.96(2.28)◊	68.69(1.30)◊	68.04(1.55)●	70.24(1.08)
	30%	48.97(2.17)●	56.33(4.60)●	62.00(3.17)●	60.87(2.09)●	66.06(1.72)◊	65.19(1.73)●	68.19(1.67)
	40%	43.43(2.23)●	51.33(3.62)●	56.08(3.66)●	55.86(5.09)◊	61.74(4.00)◊	59.48(4.77)◊	62.92(3.57)
	50%	37.03(2.28)●	48.54(2.24)●	54.85(4.10)◊	52.00(4.57)●	53.87(4.43)●	56.47(1.79)●	58.99(2.57)
MSRC-v1	10%	63.48(3.07)●	74.58(4.75)●	66.93(3.96)●	63.64(1.71)●	78.62(2.15)●	84.28(2.94)●	91.42(1.47)
	20%	44.53(4.94)●	71.33(4.48)●	63.06(2.73)●	63.34(3.58)●	77.57(3.36)●	87.05(7.00)◊	90.22(0.98)
	30%	29.43(4.34)●	65.59(3.60)●	60.80(2.84)●	58.95(2.80)●	75.40(5.20)●	86.56(4.94)◊	87.32(0.87)
	40%	17.70(2.98)●	64.82(4.38)●	54.00(3.55)●	54.78(3.22)●	74.30(3.39)●	81.95(3.14)◊	83.56(3.11)
	50%	11.05(2.32)●	59.53(4.27)●	51.36(6.00)●	52.33(3.37)●	65.36(5.69)●	75.81(3.23)◊	79.20(2.60)
Digits	10%	64.14(1.57)●	70.87(5.44)●	72.05(1.05)●	74.83(2.97)●	58.42(0.94)●	86.22(0.67)●	90.14(0.32)
	20%	56.79(2.34)●	71.53(2.85)●	69.98(2.41)●	74.32(3.46)●	54.47(1.16)●	85.65(0.69)●	89.56(0.27)
	30%	50.69(1.54)●	70.59(1.60)●	68.92(2.94)●	72.81(1.93)●	50.43(2.63)●	84.89(0.66)●	88.89(0.37)
	40%	42.65(1.79)●	67.33(1.93)●	69.18(1.86)●	70.01(2.07)●	46.71(2.54)●	85.21(1.18)●	86.68(0.68)
	50%	35.89(3.47)●	63.61(4.67)●	63.02(1.38)●	63.40(2.90)●	44.39(2.83)●	83.35(1.37)●	84.30(1.10)
win/tie/lose	30/0/0	28/2/0	26/4/0	27/3/0	26/4/0	21/9/0	-	

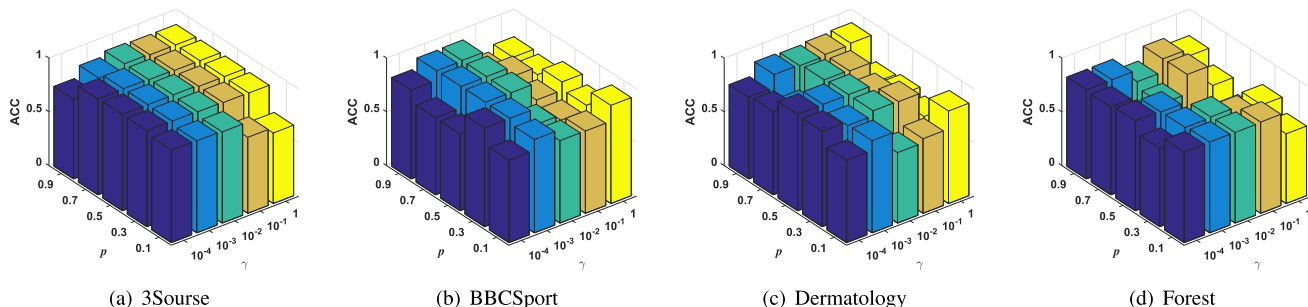


FIGURE 3. Sensitivity analysis on p and γ .

and has fast convergence property. Moreover, we conduct experiments on four datasets to study the influence of the parameters on the performance of MSCIG. From the results, we can see that the optimal parameters are data independent, and how to determine them is still an open problem.

VII. CONCLUSION

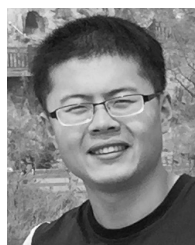
In this paper, we propose MSCIG to cluster incomplete multi-view data, which integrates processes of both graph imputation and spectral embedding seamlessly to achieve better clustering performance. To solve the resultant problem of MSCIG, we design an optimization algorithm with proved convergence which updates common representation matrix, view-specific representation matrices, similarity matrices, and view weights alternatively and iteratively. The proposed

MSCIG is evaluated on nine real-world datasets, and experimental results demonstrate its effectiveness. Inspired by [34], in the future, we plan to design a new regularization term on view-specific representations to meet the inconsistency and further improve the performance of MSCIG. Also, we want to improve the proposed method by taking the correlations of similarity matrices into account. Besides, extending the proposed MSCIG for incomplete multi-view semi-supervised classification is an interesting task.

REFERENCES

[1] J. Zhao, X. Xie, X. Xu, and S. Sun, “Multi-view learning overview: Recent progress and new challenges,” *Inf. Fusion*, vol. 38, pp. 43–54, Nov. 2017.
 [2] H. Tao, C. Hou, F. Nie, J. Zhu, and D. Yi, “Scalable multi-view semi-supervised classification via adaptive regression,” *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4283–4296, Sep. 2017.

- [3] 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.
- [4] M. Karasuyama and H. Mamitsuka, "Multiple graph label propagation by sparse integration," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 12, pp. 1999–2012, Dec. 2013.
- [5] S. Sun, "A survey of multi-view machine learning," *Neural Comput. Appl.*, vol. 23, nos. 7–8, pp. 2031–2038, Dec. 2013.
- [6] C. Hou, C. Zhang, Y. Wu, and F. Nie, "Multiple view semi-supervised dimensionality reduction," *Pattern Recognit.*, vol. 43, no. 3, pp. 720–730, Mar. 2010.
- [7] A. Trivedi, P. Rai, H. Daumé, III, and S. L. DuVall, "Multiview clustering with incomplete views," in *Proc. NIPS Workshop*, vol. 224, 2010, pp. 1–7.
- [8] S.-Y. Li, Y. Jiang, and Z.-H. Zhou, "Partial multi-view clustering," in *Proc. AAAI*, 2014, pp. 1968–1974.
- [9] H. Zhao, H. Liu, and Y. Fu, "Incomplete multi-modal visual data grouping," in *Proc. IJCAI*, 2016, pp. 2392–2398.
- [10] N. Xu, Y. Guo, X. Zheng, Q. Wang, and X. Luo, "Partial multi-view subspace clustering," in *Proc. ACM Multimedia Conf. (MM)*, 2018, pp. 1794–1801.
- [11] Q. Yin, S. Wu, and L. Wang, "Incomplete multi-view clustering via subspace learning," in *Proc. 24th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, 2015, pp. 383–392.
- [12] Q. Yin, S. Wu, and L. Wang, "Unified subspace learning for incomplete and unlabeled multi-view data," *Pattern Recognit.*, vol. 67, pp. 313–327, Jul. 2017.
- [13] L. Zhao, Z. Chen, Y. Yang, Z. J. Wang, and V. C. M. Leung, "Incomplete multi-view clustering via deep semantic mapping," *Neurocomputing*, vol. 275, pp. 1053–1062, Jan. 2018.
- [14] 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 PKDD*, 2015, pp. 318–334.
- [15] M. Hu and S. Chen, "Doubly aligned incomplete multi-view clustering," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 2262–2268.
- [16] C. Xu, D. Tao, and C. Xu, "Multi-view learning with incomplete views," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5812–5825, Dec. 2015.
- [17] H. Tao, C. Hou, D. Yi, and J. Zhu, "Unsupervised maximum margin incomplete multi-view clustering," in *Proc. ICAI*, 2018, pp. 13–25.
- [18] S. Bhadra, S. Kaski, and J. Rousu, "Multi-view kernel completion," *Mach. Learn.*, vol. 106, no. 5, pp. 713–739, May 2017.
- [19] X. Liu, W. Gao, X. Zhu, M. Li, L. Wang, E. Zhu, T. Liu, M. Kloft, D. Shen, and J. Yin, "Multiple kernel k-means with incomplete kernels," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 5, pp. 1191–1204, May 2020.
- [20] H. Gao, Y. Peng, and S. Jian, "Incomplete multi-view clustering," in *Proc. ICIP*, 2016, pp. 245–255.
- [21] 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.
- [22] W. Zhuge, C. Hou, X. Liu, H. Tao, and D. Yi, "Simultaneous representation learning and clustering for incomplete multi-view data," in *Proc. IJCAI*, Aug. 2019, pp. 4482–4488.
- [23] H. Wang, L. Zong, B. Liu, Y. Yang, and W. Zhou, "Spectral perturbation meets incomplete multi-view data," in *Proc. IJCAI*, Aug. 2019, pp. 3677–3683.
- [24] L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 17, 2004, pp. 1601–1608.
- [25] F. Nie, X. Wang, M. I. Jordan, and H. Huang, "The constrained Laplacian rank algorithm for graph-based clustering," in *Proc. AAAI*, 2016, pp. 1969–1976.
- [26] L. Tian, F. Nie, and X. Li, "A unified weight learning paradigm for multi-view learning," in *Proc. AISTATS*, 2019, pp. 2790–2800.
- [27] F. Nie, R. Zhang, and X. Li, "A generalized power iteration method for solving quadratic problem on the stiefel manifold," *Sci. China Inf. Sci.*, vol. 60, no. 11, pp. 146–155, Nov. 2017.
- [28] D. Greene and P. Cunningham, "Practical solutions to the problem of diagonal dominance in kernel document clustering," in *Proc. 23rd Int. Conf. Mach. Learn. (ICML)*, 2006, pp. 377–384.
- [29] D. Greene and P. Cunningham, "A matrix factorization approach for integrating multiple data views," in *Proc. ECML PKDD*, 2009, pp. 423–438.
- [30] H. A. Güvenir, G. Demiröz, and N. Ilter, "Learning differential diagnosis of erythematous-squamous diseases using voting feature intervals," *Artif. Intell. Med.*, vol. 13, no. 3, pp. 147–165, Jul. 1998.
- [31] B. Johnson, R. Tateishi, and Z. Xie, "Using geographically weighted variables for image classification," *Remote Sens. Lett.*, vol. 3, no. 6, pp. 491–499, Nov. 2012.
- [32] V. Sindhwani, P. Niyogi, and M. Belkin, "Beyond the point cloud: From transductive to semi-supervised learning," in *Proc. ICML*, 2005, pp. 824–831.
- [33] Y. J. Lee and K. Grauman, "Foreground focus: Unsupervised learning from partially matching images," *Int. J. Comput. Vis.*, vol. 85, no. 2, pp. 143–166, Nov. 2009.
- [34] Y. Liang, D. Huang, and C.-D. Wang, "Consistency meets inconsistency: A unified graph learning framework for multi-view clustering," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2019, pp. 1204–1209.



WENZHANG ZHUGE received the B.S. degree from Shandong University, Jinan, China, in 2015, and the M.S. degree from the National University of Defense Technology, Changsha, China, in 2017. He is currently pursuing the Ph.D. degree with the National University of Defense Technology, Changsha. His research interests include machine learning, system science, and data mining.

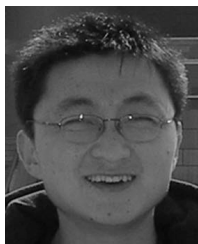


TINGJIN LUO received the B.S., master's and Ph.D. degrees from the National University of Defense Technology, Changsha, China, in 2011, 2013, and 2018, respectively. He is currently a Lecturer with the College of Science, National University of Defense Technology. He was a visiting Ph.D. student with the University of Michigan, Ann Arbor, USA, from 2015 to 2017. He has authored more than 15 articles in journals and conferences, such as the IEEE TRANSACTIONS ON

KNOWLEDGE AND DATA ENGINEERING, the IEEE TRANSACTIONS ON CYBERNETICS, the IEEE TRANSACTIONS ON IMAGE PROCESSING, KDD, and so on. He has been a Program Committee member of several conferences including IJCAI, AAAI, and so on. His research interests include machine learning, multi-media analysis, optimization, and computer vision.



HONG TAO received the Ph.D. degree from the National University of Defense Technology, Changsha, China, in 2019. She is currently a Lecturer with the College of Liberal Arts and Science, National University of Defense Technology. Her research interests include machine learning, system science, and data mining.



CHENPING HOU (Member, IEEE) received the Ph.D. degree from the National University of Defense Technology, Changsha, China, in 2009. He is currently a Full Professor with the Department of Systems Science, National University of Defense Technology. He has authored more than 80 peer-reviewed articles in journals and conferences, such as the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING

SYSTEMS/IEEE TRANSACTIONS ON NEURAL NETWORKS, the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS/TCB, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IJCAI, and AAAI. His current research interests include machine learning, data mining, and computer vision.



DONGYUN YI received the B.S. degree from Nankai University, Tianjin, China, and the M.S. and Ph.D. degrees from the National University of Defense Technology, Changsha, China. He was a Visiting Researcher with the University of Warwick, Coventry, U.K., in 2008. He is currently a Professor with the College of Science, National University of Defense Technology. His current research interests include statistics, systems science, and data mining.

• • •