IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 14, 2020, accepted July 23, 2020, date of publication July 28, 2020, date of current version August 7, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.3012500*

# Auto-Weighted Incomplete Multi-View Clustering

**WANYU DENG, (Associate Member, IEEE), LIXIA LIU, JIANQIANG LI, AND YIJUN LIN**
School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an 710121, China
Corresponding author: Wanyu Deng (dengwanyu@126.com)

**ABSTRACT** Nowadays, multi-view clustering has attracted more and more attention, which provides a way to partition multi-view data into their corresponding clusters. Previous studies assume that each data instance appears in all views. However, in real-world applications, it is common that each view may contain some missing data instances, resulting in incomplete multi-view data. To address the incomplete multi-view clustering problem, we will propose an auto-weighted incomplete multi-view clustering method in this paper, which learns a common representation of the instances and an affinity matrix of the learned representation simultaneously in a unified framework. Learning the affinity matrix of the representation guides to learn a more discriminative and compact consensus representation for clustering. Moreover, by considering the impact of the significance of different views, an adaptive weighting strategy is designed to measure the importance of each view. An efficient iterative algorithm is proposed to optimize the objective function. Experimental results on various real-world datasets show that the proposed method can improve the clustering performance in comparison with the state-of-the-art methods in most cases.

**INDEX TERMS** Adaptive weighting strategy, affinity matrix, common representation, incomplete multi-view clustering.

## I. INTRODUCTION

In recent years, many real-world datasets naturally come from multiple sources or comprise of multiple modalities, which are called multi-view data. For instance, in web page clustering, the web page content and its linkage information can be regarded as two views; in disease diagnosis, the blood test, CT, and the neuroimage can be regarded as three views of each individual; in bi-lingual documents grouping, the two languages can be seen as two views. Usually, in multi-view data, these multiple views provide consistent and complementary information. By exploiting the information present in the different views, multi-view learning methods have been proposed for tasks such as clustering and classification.

In all tasks of multi-view learning [1], [2], multi-view clustering [3], which exploits multiple views to effectively learn from unlabeled data, has attracted more and more attention. The goal of multi-view clustering is partitioning data points into their corresponding clusters. In the past few years, a number of multi-view clustering methods have been proposed. Among these methods, there are two main clustering categories: spectral based and subspace based. With the help of some similarity measure between examples, spectral clustering [4] has been extended to multi-view data. de Sa [5] creates a bipartite graph and proposes the spectral clustering algorithm based on the minimizing-disagreement idea. Based on the spectral clustering, [6] boosts the multi-view clustering by exploring the complementary information of multi-view representations. Wang *et al.* [7] propose a structured low-rank matrix factorization-based method to learn low-dimensional data-cluster representations for all views by preserving the flexible manifold structures and introducing the divergence constraint term jointly. Subspace based methods learn a latent subspace for different views. Canonical Correlation Analysis (CCA) [8] is a typical subspace-based approach, which has been proposed for dimension reduction [9] and multi-view subspace learning [10], [11]. Gao *et al.* [12] learn a common and consistent representation of all views. Liu *et al.* [13] formulate a joint matrix factorization process with the constraint that pushes the clustering solution of each view towards a common consensus instead of fixing it directly.

It is noteworthy that existing multi-view clustering methods make a common assumption that all the views are complete, i.e., each instance presents in all views. However, in real-world applications, multi-view data tend to be incomplete. For example, in bi-lingual documents grouping, the two languages can be treated as two views and many documents have only a single language part. Another example is in disease diagnosis, the blood test, CT, and MRI can be regarded as three views of everyone, and it often occurs that some

The associate editor coordinating the review of this manuscript and approving it for publication was Mu-Yen Chen.

individuals would only like to take one test. The above-mentioned cases lead to the incomplete multi-view data. The multi-view clustering approaches can not be exploited to the scene of incomplete multi-view data, and thus can not obtain a satisfactory clustering performance. So how to make use of the complementary information of the different views and reduce the impact of incomplete views become the most challenging problems in incomplete multi-view learning. Recently, different incomplete multi-view clustering methods have been proposed. Trivedi *et al.* [14] propose to deal with incomplete views by utilizing information from one complete view to refer to the kernel of incomplete views. [15], [16] attempt to solve multiple incomplete views clustering based on nonnegative matrix factorization(NMF). Li *et al.* [15] propose PVC to learn a common latent subspace for the different views based on NMF. Zhao *et al.* [17] propose IMG to learn the common latent subspace for all the views by integrating a graph Laplacian term. However, PVC and IMG can only deal with the two-view incomplete multi-view clustering problem. Shao *et al.* [16] propose MIC to learn a latent subspace for more than two incomplete views based on weighted NMF and $L_{2,1}$-Norm regularization [18], [19]. However, there are some limitations to these methods and incomplete multi-view clustering still faces significant challenges. The challenges of incomplete multi-view clustering are: 1) Previous incomplete multi-view clustering methods exploit the available information of the non-missing views to learn a common representation for all views, ignoring the underlying semantic information of the missing views. 2) The previous incomplete multi-view clustering methods, which are offline, require that the multi-view data can be fitted into the memory. However, in the data explosion age, the size of individual views data is often huge. For example, in Web scale data mining, one may encounter billions of Web pages and the dimension of the features may be as large as $O(10^6)$. Data in such scale is hard to store in the memory and process in offline way. Therefore, existing offline incomplete multi-view clustering methods can not handle large-scale data.

In this paper, we propose an Auto-Weighted Incomplete Multi-View Clustering, dubbed as AWIMVC, to address incomplete multi-view clustering problem. Firstly, we learn a common representation for all the views in the latent subspace where the instances belonging to the same example are close to each other. Secondly, we propose to learn the affinity matrix of the learned representation which reveals the global subspace structure of data and guides to learn a more discriminative and compact common representation. Thirdly, an adaptively learned weight vector is introduced to measure the importance of each view. Experimental results demonstrate the advantages of the proposed method. In brief, the proposed AWIMVC has the following contributions.

1) We propose to learn a common representation of the views in the latent space as well as preserve the global structure simultaneously. Moreover, due to the affinity matrix learning, a more discriminative and compact common representation is obtained which is beneficial to clustering.
2) Considering the importance of different views, an adaptive weight vector is proposed to measure the significance of different views.
3) An iterative optimization algorithm for AWIMVC with a convergent guarantee is proposed. Experimental results on real-world datasets demonstrate its advantages.

## II. RELATED WORK

Li *et al.* [15] propose Partial Multi-View Clustering (PVC) which is able to address the case that each view suffers from missing information. PVC learns a latent subspace based on NMF. The objective function of PVC is designed as follows:

$$\begin{aligned}\min_{\{U^{(v)},\bar{P}^{(v)}\}_{v=1}^{2}} O = &\left\| \begin{bmatrix} X_c^{(1)} \\ \hat{X}^{(1)} \end{bmatrix} - \begin{bmatrix} P_c \\ \hat{P}^{(1)} \end{bmatrix} U^{(1)} \right\|_F^2 + \lambda \left\| \bar{P}^{(1)} \right\|_1 \\ &+ \left\| \begin{bmatrix} X_c^{(2)} \\ \hat{X}^{(2)} \end{bmatrix} - \begin{bmatrix} P_c \\ \hat{P}^{(2)} \end{bmatrix} U^{(2)} \right\|_F^2 + \lambda \left\| \bar{P}^{(2)} \right\|_1 \\ s.t. &U^{(1)} \geq 0, U^{(2)} \geq 0, \bar{P}^{(1)} \geq 0, \bar{P}^{(2)} \geq 0. \end{aligned} \tag{1}$$

where $\lambda$ is the penalty parameter of the views, $U^{(1)}$, $U^{(2)}$ are the basis matrix for each view's latent space, $\bar{P}^{(1)} = [P_c; \hat{P}^{(1)}]$, $\bar{P}^{(2)} = [P_c; \hat{P}^{(2)}]$ are the latent representation of instances for two views. The homogeneous feature representation for all examples can be obtained as $P = [P_c; \hat{P}^{(1)}; \hat{P}^{(2)}]$. Clustering approaches such as k-means can be applied on such representation.

## III. PROPOSED METHOD

In this section, we present the proposed approach (AWIMVC), which simultaneously learns a common representation in the latent subspace and an affinity matrix of the learned representation. In the following, we propose our model in three aspects and then give a unified objective function for implementing AWIMVC.

### A. COMMON REPRESENTATION LEARNING

Inspired by nonnegative matrix factorization(NMF) [20], we propose to learn latent representations in the latent subspace of different views by matrix factorization:

$$\begin{aligned}\min_{U^{(v)},V^{(v)}} &\sum_{v=1}^{n_v} \left\| X^{(v)} - U^{(v)} V^{(v)} \right\|_F^2 \\ s.t. \quad &U^{(v)T} U^{(v)} = I, \end{aligned} \tag{2}$$

where $X^{(v)} \in R^{m_v \times n}$ denotes the instances from $v$th view, $U^{(v)} \in R^{m_v \times c}$ is the basis matrix for the $v$th view's latent subspace, $V^{(v)} \in R^{c \times n}$ is the latent representation of the $v$th view in the latent subspace, where $c$ is the dimension of the latent representation or cluster number. In (2), the orthogonal constraint $U^{(v)T} U^{(v)} = I$ is introduced to make the basis matrix independent. Moreover, we impose the orthogonal constraint on $U^{(v)} \in R^{m_v \times c}$ to avoid a trivial solution.

However, (2) only independently decomposes different views without considering their consistency information. To address this problem, we assume that different views have distinct basis matrices $\{U^{(v)}\}_{v=1}^{n_v}$, but share the same latent representation $V$. As a result, (2) is rewritten as follows:

$$\begin{aligned} &\min_{U^{(v)},V} \sum_{v=1}^{n_v} \left\| X^{(v)} - U^{(v)}V \right\|_F^2 \\ &s.t. \quad U^{(v)T}U^{(v)} = I, \end{aligned} \tag{3}$$

In common representation learning, the previous works ignore the hidden information of the missing views [21]. In this paper, we propose to learn a consensus representation not only with the available information of the non-missing views but also with the underlying information of the missing views:

$$\begin{aligned} \min_{U^{(v)},V,E^{(v)}} &\sum_{v=1}^{n_v} (\left\| X^{(v)} + E^{(v)}M^{(v)} - U^{(v)}V \right\|_F^2 \\ &+ \frac{\lambda_1}{2} \sum_{j=1}^{m_v} \sum_{i=1}^{m_v} \left\| E_{i,:}^{(v)} - E_{j,:}^{(v)} \right\|_2^2 W_{i,j}^{(v)}) \\ s.t. \quad & U^{(v)T}U^{(v)} = I, \end{aligned} \tag{4}$$

where $\lambda_1$ is a positive penalty parameter. $E^{(v)} \in R^{m_v \times n_v^m}$ denotes the error matrix, which is used to model the missing instances of the $v$th view, $n_v^m$ is the number of missing instances of the $v$th view, $E_{i,:}^{(v)}$ and $E_{j,:}^{(v)}$ denote the $i$th and $j$th row vector of matrix $E^{(v)}$. $M^{(v)} \in R^{n_v^m \times n}$ is an index matrix of the $v$th view and is defined as follows:

$$M_{i,j}^{(v)} = \begin{cases} 1, & if\ the\ j-th\ instance \\ & is\ the\ i-th \\ & missing\ instance \\ & in\ the\ \text{v-th}\ view \\ 0, & otherwise. \end{cases} \tag{5}$$

In (4), $W^{(v)} \in R^{m_v \times m_v}$ is the nearest neighbor graph of features from the $v$th view, which is pre-constructed as:

$$W_{i,j}^{(v)} = \begin{cases} 1, & if\ \bar{X}_{i,:}^{(v)} \in \psi(\bar{X}_{j,:}^{(v)})\ or\ \bar{X}_{j,:}^{(v)} \in \psi(\bar{X}_{i,:}^{(v)}) \\ 0, & otherwise, \end{cases} \tag{6}$$

where $\psi(\bar{X}_{j,:}^{(v)})$ denotes the set of $k$ nearest neighbors of the $j$th feature.

The problem (4) can be transformed into the following equivalent formula:

$$\begin{aligned} \min_{U^{(v)},V,E^{(v)}} &\sum_{v=1}^{n_v} (\left\| X^{(v)} + E^{(v)}M^{(v)} - U^{(v)}V \right\|_F^2 \\ &+ \lambda_1 Tr(E^{(v)T}L^{(v)}E^{(v)})) \\ s.t. \quad & U^{(v)T}U^{(v)} = I, \end{aligned} \tag{7}$$

where $L^{(v)}$ is the Laplacian matrix of the graph $W^{(v)}$ and is calculated as $L^{(v)} = D^{(v)} - W^{(v)}$, $D^{(v)}$ is a diagonal matrix and its $i$th diagonal element is calculated by $D_{i,j}^{(v)} = \sum_{j=1}^{m_v} W_{i,j}^{(v)}$.

### B. AFFINITY MATRIX LEARNING

Affinity matrix reveals the global subspace structure of data. The clustering results highly depend on the learned affinity matrix, so it is significant to preserve the global structure of data in clustering. In addition, global structure preserving is beneficial to improve the compactness of the learned low-dimensional representation. For multi-view learning problem, global constraints can be easily achieved because of the complete view setting. While in the incomplete multi-view learning problem, the global constraints cannot be easily achieved due to the missing instances of views. To preserve the compact global structure of data, we propose to learn the affinity matrix of representation, which is defined as follows:

$$\begin{aligned} \min_{G} \quad & \frac{\lambda_2}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (\left\| V_{:,i} - V_{:,j} \right\|_2^2 G_{i,j}) + \lambda_3 \left\| G \right\|_F^2 \\ s.t. \quad & \forall i, G_i^T \mathbf{1} = 1, G_i \geq 0, \end{aligned} \tag{8}$$

where $G \in R^{n \times n}$ denotes the affinity matrix with each element representing the similarity degree between the corresponding two instances. $\mathbf{1}$ is a vector with all 1s. Introducing the constraints $G_i\mathbf{1} = 1$ and $G_i \geq 0$ can guarantee the probability property of $G_i$. Minimizing $||G||_F^2$ can avoid the trivial solution.

For the Eq. (8), we can rewrite it as follows:

$$\begin{aligned} \min_{G} \quad & \lambda_2 Tr(VL_GV^T) + \lambda_3 \left\| G \right\|_F^2 \\ s.t. \quad & \forall i, G_i^T \mathbf{1} = 1, G_i \geq 0, \end{aligned} \tag{9}$$

where $L_G \in R^{n \times n}$ is the Laplacian matrix of affinity matrix $G$, defined by $L_G = D - G$, in which the degree matrix is the diagonal matrix with $D_{ii} = \sum_{j=1}^{n} G_{ij}$.

### C. ADAPTIVE WEIGHT VECTOR LEARNING

In multi-view learning, it is crucial to assign a reasonable weight to each view according to the importance of different views, which is taken into account in [22], [23]. For multi-view clustering task, assigning a suitable weight for each view is beneficial to improve the clustering performance. It is necessary to employ a group of meaningful weights to measure the importance of each view. In the case of incomplete multi-view clustering, the available information of different views will have huge differences because of the number of the missing instances of multiple views and the different feature dimensions. It is essential to assign a reasonable weight to each view according to the importance of the view. To this end, we introduce a hyperparameter to learn the weight vector of the views:

$$\begin{aligned} &\min_{\mu^{(v)}} \sum_{v=1}^{n_v} (\mu^{(v)})^s \Lambda^{(v)} \\ &s.t. \sum_{v=1}^{n_v} \mu^{(v)} = 1, \mu^{(v)} \geq 0 \end{aligned} \tag{10}$$

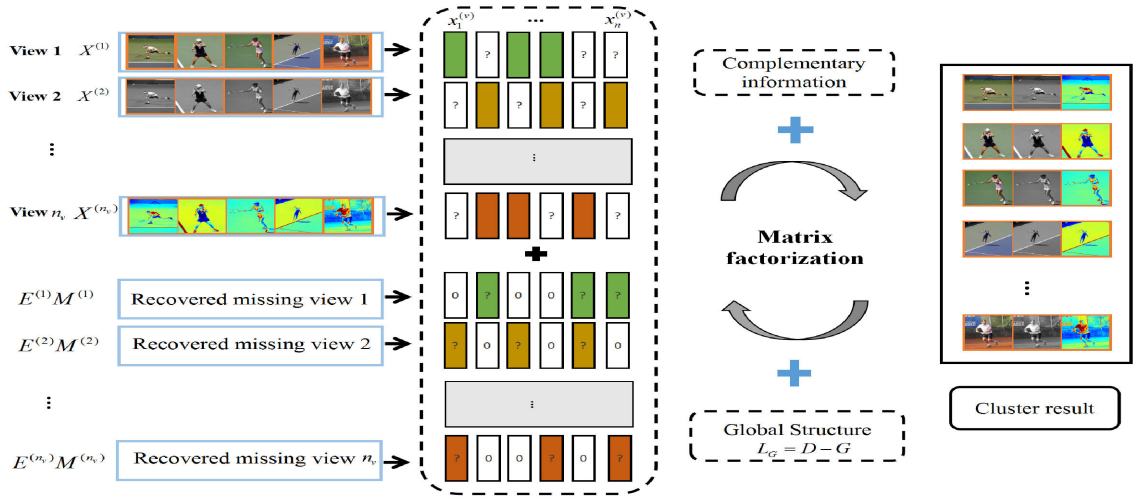



**FIGURE 1.** Framework of the proposed approach.

where $\mu^{(v)}$ is a weight to measure the importance of the $v$th view. $\Lambda^{(v)}$ denotes the learning model of the $v$th view. Parameter $s > 1$ is used to smooth the weight distribution.

Finally, by integrating the common representation learning, affinity matrix learning, and adaptive weight vector learning into a unified framework, we obtain the final model as follows:

$$\begin{aligned}\min_{U^{(v)},V,E^{(v)},G,\mu^{(v)}} &\sum_{v=1}^{n_v}(\mu^{(v)})^s \\ &\times \begin{pmatrix} \left\|X^{(v)}+E^{(v)}M^{(v)}-U^{(v)}V\right\|_F^2 \\ +\lambda_1 Tr(E^{(v)T}L^{(v)}E^{(v)}) \\ +\lambda_2 Tr(VL_GV^T)+\lambda_3\left\|G\right\|_F^2 \end{pmatrix} \\ s.t. &U^{(v)T}U^{(v)}=I, \forall i, G_i^T\mathbf{1}=1, G_i\geq 0, \\ &\sum_{v=1}^{n_v}\mu^{(v)}=1, \mu^{(v)}\geq 0 \end{aligned} \tag{11}$$

The framework of the proposed approach is outlined in Fig.1.

## IV. OPTIMIZATION ALGORITHM

In this section, we first solve the objective function of AWIMVC. Then, we will give the convergence and computational complexity analysis. The algorithm flow will be presented in detail.

### A. OPTIMIZATION

To address the optimization problem (11), we introduce an alternative optimization algorithm [24].

Step 1: Update variable $U^{(v)}$. Fixing all of the other variables, the sub-problem to solve variable $U^{(v)}$ is reformulated as:

$$\min_{U^{(v)T}U^{(v)}=I}\left\|X^{(v)}+E^{(v)}M^{(v)}-U^{(v)}V\right\|_F^2 \tag{12}$$

We can obtain the optimal solution of $U^{(v)}=LR^T$ by solving (12), where $L$ and $R$ are the left and right singular matrices of the matrix $((X^{(v)}+E^{(v)}M^{(v)})V^T)$.

Step 2: Update variable $V$. Fixing the other variables, the optimization problem to calculate $V$ is formulated as follows:

$$\begin{aligned}\varphi(V)=\sum_{v=1}^{n_v}(\mu^{(v)})^s(&\left\|X^{(v)}+E^{(v)}M^{(v)}-U^{(v)}V\right\|_F^2 \\ &+\lambda_2 Tr(VL_GV^T))\end{aligned} \tag{13}$$

Then the derivative of $\varphi(V)$ with respect to $V$ is:

$$\begin{aligned}\frac{\partial\varphi(V)}{\partial V}=\sum_{v=1}^{n_v}(\mu^{(v)})^s[&2U^{(v)T}(U^{(v)}V-Y^{(v)}) \\ &+\lambda_2(VL_G^T+VL_G)]\end{aligned} \tag{14}$$

Setting the derivative of $\varphi(V)$ to be 0 with respect to $V$, we can obtain variable $V$ as follows:

$$V=(\sum_{v=1}^{n_v}(\mu^{(v)})^sU^{(v)T}Y^{(v)})(I+\lambda_2L_G)^{-1}\Big/\sum_{v=1}^{n_v}(\mu^{(v)})^s \tag{15}$$

where $Y^{(v)}=X^{(v)}+E^{(v)}M^{(v)}$.

Step 3: Update variable $E^{(v)}$. When fixing the other variables, the sub-problem to solve variable $E^{(v)}$ is as follows:

$$\min_{E^{(v)}}\left\|X^{(v)}+E^{(v)}M^{(v)}-U^{(v)}V\right\|_F^2+\lambda_1 Tr(E^{(v)T}L^{(v)}E^{(v)}) \tag{16}$$

Since all instances of $X^{(v)}$ corresponding to $E^{(v)}$ are zeros, problem (16) can be rewritten as follows:

$$\varphi(E^{(v)})=\left\|E^{(v)}-U^{(v)}VM^{(v)T}\right\|_F^2+\lambda_1 Tr(E^{(v)T}L^{(v)}E^{(v)}) \tag{17}$$

**Algorithm 1** AWIMVC(solving (11))

**Input:** Incomplete multi-view data $X = \{X^{(v)}\}_{v=1}^{n_v}$ with missing views filled in 0s, index matrix of missing views $M = \{M^{(v)}\}_{v=1}^{n_v}$, nearest neighbor graph $W = \{W^{(v)}\}_{v=1}^{n_v}$, parameters $\lambda_1, \lambda_2, \lambda_3, s$.
**Initialization:** $\mu^{(s)} = 1/n_v$, $U^{(v)}$ is an orthogonal matrix with random values, random matrix $E^{(v)}$, random graph $G$.
**while** not converged **do**
1. Update $V$ using (15);
2. Update $G$ using (20);

**for** $v$ from 1 to $n_v$
3. Update $U^{(v)}$ by solving (12);
4. Update $E^{(v)}$ by solving (18);

**end**
5. Update $\mu^{(v)}$ using (22).

**end while**
**Output:** $V$, $U^{(v)}$, $G$, $E^{(v)}$

Then we can obtain $E^{(v)}$ by setting the derivative of $\varphi(E^{(v)})$ with respect to $E^{(v)}$ to 0 as follows:

$$\begin{aligned}\frac{\partial\varphi(E^{(v)})}{\partial E^{(v)}} &= 2(E^{(v)} - U^{(v)}VM^{(v)T}) \\ &\quad +\lambda_1(L^{(v)}E^{(v)} + L^{(v)T}E^{(v)}) = 0 \\ \Leftrightarrow E^{(v)} &= (I + \lambda_1 L^{(v)})^{-1}U^{(v)}VM^{(v)T} \end{aligned} \tag{18}$$

Step 4: Update variable $G(L_G)$. Fixing the other variables, the sub-problem of the variable $G$ is in the following form:

$$\begin{aligned}&\min_{G} \lambda_2 Tr(VL_GV^T) + \lambda_3 \|G\|_F^2 \\ &s.t. \forall i, G_i^T\mathbf{1} = 1, G_i \geq 0\end{aligned} \tag{19}$$

We divide problem (19) into a set of sub-problems $G^i$ according to samples index i as

$$G^i = \underset{G^i\in\{\alpha|\alpha^T\mathbf{1}=1;\alpha\geq 0\}}{\arg\min} \left\|G^i + S^i\right\|_F^2, \tag{20}$$

where $S^i$ is a vector with its element j defined as $S^{ij} = \frac{\lambda_2\|V^{:,i}-V^{:,j}\|_F^2}{4\lambda_3}$. The detailed deduction is described in [25].

Step 5: Update variable $\mu^{(v)}$. Fixing all of the other variables, we can calculate $\mu^{(v)}$ by minimizing the following problem:

$$\min_{\mu^{(v)}>0, \sum_{v=1}^{n_v}\mu^{(v)}=1} \sum_{v=1}^{n_v} (\mu^{(v)})^s f^{(v)}, \tag{21}$$

where $f^{(v)} = \left\|X^{(v)} + E^{(v)}M^{(v)} - U^{(v)}V\right\|_F^2 + \lambda_1 Tr(E^{(v)T}L^{(v)}E^{(v)}) + \lambda_2 Tr(VL_GV^T) + \lambda_3\|G\|_F^2$.

The optimal solution of $\mu^{(v)}$ is given by [26]:

$$\mu^{(v)} = (f^{(v)})^{1/(1-s)} \Big/ \sum_{v=1}^{n_v} (f^{(v)})^{1/(1-s)} \tag{22}$$

The entire algorithm of the proposed method is summarized in Algorithm 1.

### B. CONVERGENCE AND COMPUTATIONAL COMPLEXITY

Convergence Analysis. As shown by Algorithm 1, an iterative algorithm is exploited to optimize the optimization problem. The optimization of AWIMVC is divided into five sub-problems, which are all convex *w.r.t.* each parameter. And the five sub-problems have the closed-form solution *w.r.t.* each parameter. It is concluded that the objective function is monotonically decreasing towards a stationary point. Moreover, the objective function is lower bounded. The above two factors ensure the proposed optimization method to finally find the local optimal point of the objective problem, which guarantees the convergence property.

Computational Complexity. For the calculation of the variables in Steps 1 and 5, it can be found that the highest computational costs are the singular value decomposition (SVD) in Step 1 and inverse operation in Steps 2 and 3. For an $m \times n$ matrix, the computational complexity operation of SVD is $O(mn^2)$. For an $n \times n$ matrix, the computational complexity of the inverse operation is $O(n^3)$. For Step 3, the computational complexity can be ignored since the inverse operation of $(I + \lambda_1 L^{(v)})^{-1}$ can be calculated outside the loop. For Step 4, the computational complexity of updating $G$ is $O(n^3)$. Thus the total computational complexity of Algorithm 1 is about $O(\sum_{v=1}^{n_v} tm_v c^2 + 2tn^3)$, where $t$ is the iteration times.

## V. EXPERIMENTS AND ANALYSIS

In this section, we perform experiments on five datasets to prove the effectiveness and efficiency of the proposed method.

### A. EXPERIMENTAL SETTINGS

In this section, we first give the detailed information of the datasets and the incomplete multi-view data construction. The baseline methods which are used to compare with the proposed method are presented. Finally, we choose three criteria to evaluate the clustering performances of different methods.

#### 1) DATASETS

We conduct experiments on five real-world datasets. Table 1 describes the statistics of the five datasets. The detailed information of the datasets is as follows.

**TABLE 1. Statistics of the datasets.**

| Dataset | # Classes | # Views | # Size | # Features |
|---|---|---|---|---|
| Handwritten | 10 | 2 | 2000 | 240/76 |
| BUAA | 10 | 2 | 90 | 100/100 |
| 3 Sources | 6 | 3 | 169 | 3560/3631/3068 |
| BBCSport | 5 | 4 | 116 | 1991/2063/2113/2158 |
| Cornell | 5 | 2 | 195 | 195/1703 |

1) Handwritten digit dataset [27]: The handwritten digit dataset consists of 2000 handwritten digits (0-9). Two views are used in our experiments: 1) 240 pixel averages in windows, 2) 76 Fourier coefficients of the character shapes.

2) BUAA-visnir face dataset (BUAA) [28]: Following the experimental settings in [17], we conduct our experiments on a subset of BUAA which is composed of 90 visual images and 90 near infrared images of the first 10 classes.

3) 3 Sources dataset: This dataset is collected from three online news sources: 1) BBC; 2) Reuters; and 3) The Guardian. Each source can be regarded as a view. There are 948 news articles in total. A subset of 3 Sources dataset is extracted for evaluation in our experiments, which contains 169 stories of six topical labels and the 169 stories are reported in all of the three sources.

4) BBCSport dataset: The original BBCSport dataset contains 737 documents about the sport news articles collected from BBCSport website. These documents are described by 2–4 views and categorized into five classes. We select a subset to conduct our experiments which consists of 116 samples described by all of the four views.

5) Cornell dataset: The Cornell dataset is one of the popular WebKB datasets, which contains 195 documents over the five labels, i.e., student, project, course, staff, and faculty. Each document is described by two views: the content view and the citation view.

#### 2) INCOMPLETE MULTI-VIEW DATA CONSTRUCTION

Following [21], we construct two types of incomplete multi-view data in our experiments.

1) For the BUAA, Handwritten, and Cornell datasets, we randomly select 10%, 30%, 50%, and 70% samples from the datasets as the paired samples. For the half of the remaining samples, we remove the first view. While for the other half of the samples, we remove the second view to construct the incomplete multi-view data.

2) For the BBCSport and 3 Sources datasets, we randomly remove 10%, 30%, and 50% instances of each view to construct the incomplete multi-view data.

#### 3) BASELINE METHODS

We select the following methods to compare with the proposed method.

1) BSV (Best Single View): Following [16], BSV fills in the missing views with the average of instances of the corresponding view, and then performs k-means on each view independently, and reports the best result.

2) Concat: Same as BSV, Concat fills in the missing views with the average of instances of the corresponding view. Differently, Concat concatenates all views into a single view, and then performs k-means to obtain the clustering results.

3) Partial multi-view clustering (PVC) [15]: PVC learns a common representation for all views based on the NMF(non-negative matrix factorization), and then performs k-means on the learned representation to obtain the clustering results.

4) Partial multi-view clustering using graph regularized NMF(GPMVC) [29]: GPMVC extends PVC in two directions. First, GPMVC extends PVC for the k partial-view scenario. Second, GPMVC extends the algorithm to include view specific graph Laplacian regularization. Same as PVC, GPMVC learns a common representation for all views, and then performs k-means on the learned representation to obtain the clustering results.

5) Incomplete multi-modality grouping (IMG) [17]: IMG uses the matrix factorization technique to learn a common latent representation for all views. Moreover, IMG learns a similarity matrix for the learned representation.

6) MIC [16]: MIC first fills the missing instances in each incomplete view with the average features, and then based on weighted non-negative matrix factorization with $L_{2,1}$-norm regularization, MIC learns a common latent representation.

7) Doubly aligned incomplete multi-view clustering (DAIMC) [30]: DAIMC uses the weighted semi-NMF to learn a common representation for all views, and then performs k-means on the learned representation to obtain the clustering results.

8) Online multi-view clustering (OMVC) [31]: OMVC processes the data chunk by chunk and learns a common feature matrix simultaneously.

9) UEAF [21]: Based on recovering the missing views and reverse graph regularization, UEAF learns a common representation for all views. And then UEAF performs k-means on the learned representation to obtain the clustering results.

#### 4) EVALUATION

In our experiments, we choose the clustering accuracy (ACC), normalized mutual information (NMI), and purity as the criterion to compare these methods [27]. For the above baseline methods except BSV and Concat, the grid search approach is exploited to find the optimal parameters of these methods and the best clustering results of these methods are reported, respectively. For each dataset, we perform all methods on the same 5 randomly constructed incomplete cases and report their average results for fair comparison.

### *B. EXPERIMENTAL RESULTS AND ANALYSIS*

Experimental results on the above datasets are shown in Tables 2-6 and Figs. 2-6. From these tables and figures, we can obtain the following points.

1) From Tables 2-6 and Figs. 2-6, it can be observed that, compared with the other methods, BSV and Concat baseline methods perform worst in most cases. The experimental results of BSV and Concat prove that simply fills in the missing views with the average of instances of the corresponding view is not a good method to solve the incomplete multi-view clustering problem.

2) It can be found that, by exploiting the complementary information of the views to learn a consensus representation, PVC, GPMVC, IMG, DAIMC, UEAF, and the proposed method can achieve much better performance than BSV and Concat in most cases from Tables 2-6 and Figs. 2-6.

3) From Tables 2-4 and Figs. 2-4, we can find that IMG, GPMVC, UEAF, and AWIMVC outperform PVC in most cases, which proves the effectiveness of preserving the geometric structure of data in incomplete multi-view clustering.

**TABLE 2.** Mean NMIs (%), ACCs (%), and purities (%) of different methods on the handwritten dataset.

| | NMI | | | | ACC | | | | Purity | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method\Paired rate | 0.1 | 0.3 | 0.5 | 0.7 | 0.1 | 0.3 | 0.5 | 0.7 | 0.1 | 0.3 | 0.5 | 0.7 |
| BSV [21] | 37.04 | 44.48 | 51.50 | 58.61 | 43.08 | 50.45 | 57.39 | 64.44 | 43.81 | 51.08 | 58.31 | 65.92 |
| Concat [21] | 47.71 | 54.43 | 61.12 | 70.30 | 46.01 | 57.46 | 66.45 | 78.64 | 48.20 | 58.57 | 66.73 | 78.72 |
| PVC [21] | 55.13 | 60.85 | 64.88 | 68.54 | 63.81 | 70.90 | 73.44 | 75.20 | 65.41 | 72.14 | 75.01 | 76.96 |
| GPMVC [21] | 60.99 | 63.99 | 72.23 | 73.68 | 65.60 | 74.04 | 76.94 | 79.06 | 66.58 | 74.85 | 76.21 | 79.70 |
| IMG [21] | 58.05 | 62.38 | 64.91 | 68.21 | 69.22 | 75.41 | 76.34 | 77.54 | 70.76 | 75.44 | 76.45 | 77.55 |
| MIC [21] | 44.59 | 51.09 | 57.41 | 66.32 | 49.13 | 59.99 | 64.98 | 74.70 | 50.25 | 60.56 | 66.73 | 75.70 |
| DAIMC [21] | 40.99 | 55.81 | 62.68 | 66.63 | 46.75 | 67.32 | 75.09 | 77.63 | 49.48 | 68.12 | 75.12 | 78.00 |
| OMVC [21] | 46.36 | 54.12 | 62.11 | 66.65 | 54.57 | 61.80 | 73.45 | 77.89 | 55.83 | 63.18 | 73.90 | 77.89 |
| UEAF | 60.46 | 66.77 | 70.86 | 74.14 | **69.59** | 77.95 | 82.79 | 84.80 | **69.66** | 77.95 | 82.79 | 84.80 |
| AWIMVC(ours) | **61.35** | **68.21** | **73.29** | **75.56** | 69.05 | **79.54** | **83.93** | **85.70** | 69.15 | **79.54** | **83.93** | **85.70** |

**TABLE 3.** Mean NMIs (%), ACCs (%), and purities (%) of different methods on the BUAA dataset.

| | NMI | | | | ACC | | | | Purity | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method\Paired rate | 0.1 | 0.3 | 0.5 | 0.7 | 0.1 | 0.3 | 0.5 | 0.7 | 0.1 | 0.3 | 0.5 | 0.7 |
| BSV [21] | 43.10 | 53.03 | 61.78 | 69.91 | 48.33 | 56.96 | 64.26 | 70.81 | 50.19 | 58.66 | 65.96 | 72.35 |
| Concat [21] | 51.22 | 51.95 | 52.43 | 56.51 | 45.62 | 46.61 | 47.46 | 52.34 | 47.99 | 49.41 | 49.68 | 54.81 |
| PVC [21] | 61.35 | 67.07 | 71.97 | 78.70 | 57.41 | 66.46 | 70.01 | 75.92 | 60.58 | 68.79 | 71.98 | 77.95 |
| GPMVC [21] | 62.12 | 70.25 | 74.33 | 81.63 | 58.98 | 68.75 | 74.28 | 78.28 | 61.26 | 71.66 | 75.36 | 81.71 |
| IMG [21] | 54.72 | 67.53 | 76.74 | 82.83 | 53.95 | 67.39 | 76.14 | 79.36 | 57.32 | 68.65 | 78.23 | 82.42 |
| MIC [21] | 59.92 | 72.59 | 69.56 | 74.11 | 57.77 | 70.00 | 68.22 | 74.89 | 61.11 | 73.33 | 70.44 | 75.33 |
| DAIMC [21] | 60.16 | 74.63 | 79.13 | 83.32 | 57.56 | 71.55 | 79.78 | 80.67 | 60.44 | 74.00 | 79.33 | 82.00 |
| OMVC [21] | 60.69 | 63.99 | 64.32 | 66.91 | 60.59 | 62.89 | 65.74 | 68.37 | 62.15 | 65.57 | 67.48 | 69.78 |
| UEAF | **68.47** | 80.53 | 88.97 | 90.14 | **69.11** | 81.78 | 89.78 | **92.22** | 69.78 | 82.22 | 90.22 | **92.22** |
| AWIMVC(ours) | 67.91 | **83.64** | **89.98** | **90.68** | 68.22 | **84.44** | **91.56** | 92.00 | **70.67** | **85.11** | **91.56** | 92.00 |

**TABLE 4.** Mean NMIs (%), ACCs (%), and purities (%) of different methods on the CORNELL dataset.

| | NMI | | | | ACC | | | | Purity | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method\Paired rate | 0.1 | 0.3 | 0.5 | 0.7 | 0.1 | 0.3 | 0.5 | 0.7 | 0.1 | 0.3 | 0.5 | 0.7 |
| BSV | 8.11 | 9.34 | 7.01 | 10.83 | 36.92 | 34.36 | 31.79 | 35.39 | 46.67 | 48.21 | 46.15 | 46.67 |
| Concat | 6.75 | 8.48 | 7.93 | 11.05 | 36.92 | 34.87 | 33.33 | 35.38 | 44.62 | 46.67 | 48.21 | 45.13 |
| PVC | 12.15 | 16.21 | 20.91 | 21.96 | 42.56 | 45.69 | 43.79 | 45.90 | 49.69 | 53.54 | 55.64 | 53.03 |
| GPMVC | 13.90 | 16.07 | 18.99 | 15.03 | 40.39 | 43.86 | **46.53** | 44.56 | 47.90 | 48.00 | 48.89 | 49.13 |
| IMG | 11.51 | 16.23 | 16.77 | 20.33 | 45.19 | 45.45 | 46.42 | 46.16 | 48.12 | 49.84 | 51.24 | 54.70 |
| MIC | 10.05 | 12.34 | 17.69 | 17.90 | 40.10 | 37.23 | 41.79 | 42.05 | 48.67 | 48.51 | 52.15 | 53.90 |
| DAIMC | 12.73 | 16.90 | 16.39 | 18.75 | 39.95 | 40.87 | 40.31 | 40.21 | 49.44 | 51.28 | 52.26 | 52.41 |
| OMVC | 14.25 | 15.99 | 16.83 | 18.60 | 43.32 | 43.56 | 46.14 | 47.47 | 46.79 | 46.38 | 47.35 | 50.82 |
| UEAF | 15.67 | 18.61 | 27.36 | 26.78 | 45.13 | 46.15 | 43.08 | **50.26** | 50.26 | 52.31 | 61.44 | 62.05 |
| AWIMVC(ours) | **15.87** | **18.79** | **27.99** | **30.71** | **46.66** | **46.67** | 43.59 | 47.59 | **53.33** | **54.36** | **62.05** | **63.59** |

**TABLE 5.** Mean NMIs (%), ACCs (%), and purities (%) of different methods on the bbcsport dataset.

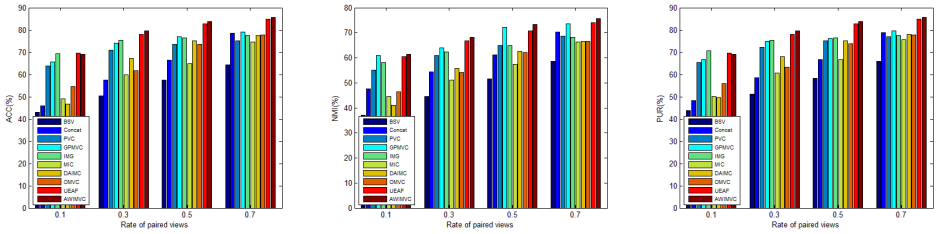
| | NMI | | | ACC | | | Purity | | |
|---|---|---|---|---|---|---|---|---|---|
| Method\Missing rate | 0.1 | 0.3 | 0.5 | 0.1 | 0.3 | 0.5 | 0.1 | 0.3 | 0.5 |
| BSV [21] | 43.73 | 31.03 | 21.40 | 58.62 | 51.31 | 44.03 | 65.79 | 55.07 | 47.59 |
| Concat [21] | 61.69 | 38.92 | 18.61 | 70.62 | 58.72 | 33.21 | 80.59 | 63.24 | 37.00 |
| GPMVC [21] | 28.23 | 20.04 | 15.48 | 51.44 | 46.89 | 43.91 | 58.39 | 52.76 | 45.29 |
| MIC [21] | 29.90 | 25.84 | 24.01 | 51.21 | 46.21 | 46.03 | 55.00 | 51.72 | 52.41 |
| DAIMC [21] | 56.62 | 50.17 | 37.89 | 68.62 | 63.45 | 56.89 | 76.90 | 71.72 | 61.03 |
| OMVC [21] | 30.64 | 41.57 | 40.63 | 53.33 | 51.38 | 48.79 | 56.49 | 59.20 | 57.47 |
| UEAF | 68.15 | 69.88 | 48.66 | 77.41 | 78.28 | **64.48** | 86.38 | 87.43 | **73.79** |
| AWIMVC(ours) | **70.14** | **70.70** | **49.19** | **78.79** | **79.66** | 64.14 | **87.59** | **87.76** | 73.28 |

4) From Tables 5-6 and Figs. 5-6, AWIMVC, UEAF and DAIMC achieve much better performance than MIC and OMVC, which proves that AWIMVC, UEAF and DAIMC can capture more complementary and compatible information from the incomplete multiple views.

5) From Tables 2-4 and Figs. 2-4, AWIMVC and UEAF perform much better than IMG, which demonstrates that

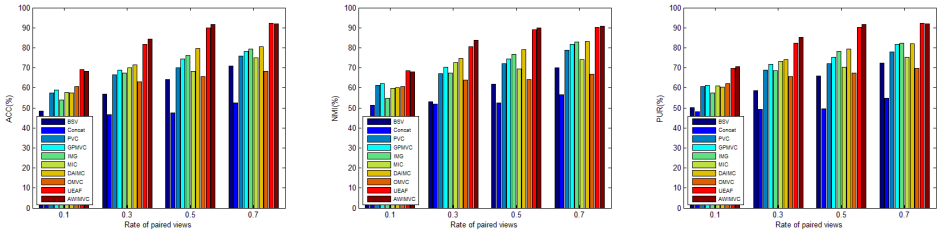
**TABLE 6.** Mean NMIs (%), ACCs (%), and purities (%) of different methods on the 3sources dataset.

| Method\Missing rate | NMI | | | ACC | | | Purity | | |
|---|---|---|---|---|---|---|---|---|---|
| | 0.1 | 0.3 | 0.5 | 0.1 | 0.3 | 0.5 | 0.1 | 0.3 | 0.5 |
| BSV [21] | 50.07 | 34.46 | 22.34 | 56.90 | 47.38 | 39.24 | 68.14 | 57.63 | 48.99 |
| Concat [21] | 51.98 | 37.87 | 18.32 | 53.54 | 46.79 | 37.68 | 69.78 | 58.51 | 46.48 |
| GPMVC [21] | 34.82 | 30.44 | 28.15 | 48.24 | 44.50 | 42.01 | 60.47 | 58.58 | 57.40 |
| MIC [21] | 37.23 | 38.62 | 26.08 | 49.11 | 47.69 | 42.49 | 57.28 | 61.30 | 52.31 |
| DAIMC [21] | 52.98 | 49.07 | 41.64 | 56.33 | 52.43 | 50.73 | 68.99 | 67.21 | 63.56 |
| OMVC [21] | 36.48 | 28.42 | 24.34 | 43.95 | 41.11 | 39.53 | 59.37 | 48.76 | 45.44 |
| UEAF | 58.45 | 55.18 | 49.89 | 59.29 | 57.75 | 54.08 | **77.99** | 72.90 | 70.65 |
| AWIMVC(ours) | **59.78** | **56.68** | **53.49** | **68.28** | **59.05** | **56.80** | 75.27 | **73.85** | **73.49** |



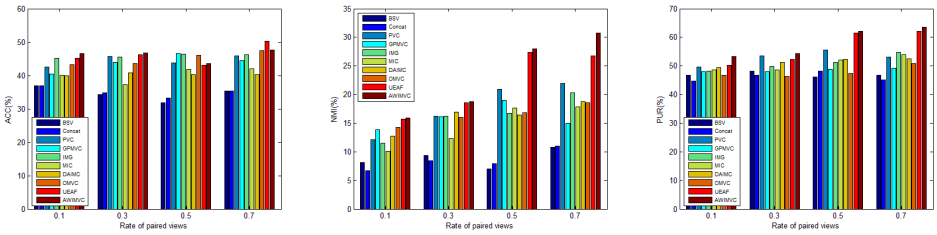

(a) (b) (c)

**FIGURE 2.** ACC, NMI, and purity of different methods on handwritten dataset with different rates of paired samples.




(a) (b) (c)

**FIGURE 3.** ACC, NMI, and purity of different methods on BUAA dataset with different rates of paired samples.




(a) (b) (c)

**FIGURE 4.** ACC, NMI, and purity of different methods on CORNELL dataset with different rates of paired samples.

learning a weight vector for the views to measure the importance of the different views and taking into account the underlying information of the missing views are beneficial to the incomplete multi-view clustering.

6) From Tables 2-6 and Figs. 2-6, AWIMVC outperforms UEAF in most cases, which demonstrates that learning the common representation and affinity matrix of the learned representation simultaneously, which guides to learn a more discriminative and compact consensus representation, is a good choice for incomplete multi-view clustering.

### C. PARAMETER SENSITIVITY ANALYSIS

In this section, we focus on analyzing the sensitivity of the tunable parameters. We first fix parameters $k = 9$ and $s = 3$,

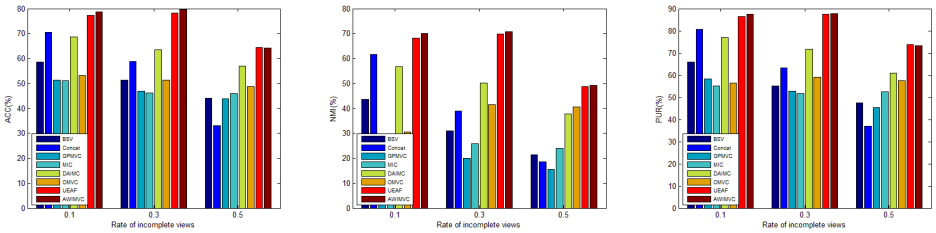



**FIGURE 5.** ACC, NMI, and purity of different methods on BBCSport dataset with different rates of incomplete samples.

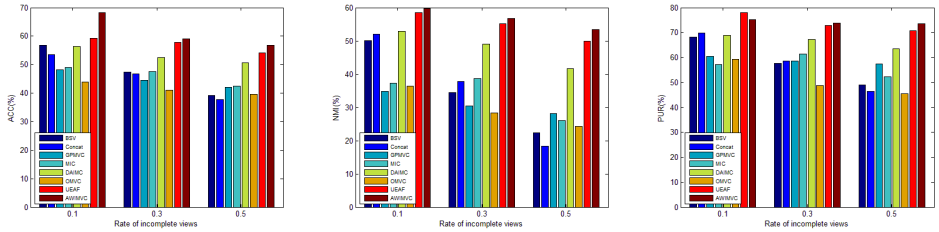



**FIGURE 6.** ACC, NMI, and purity of different methods on 3sources dataset with different rates of incomplete samples.

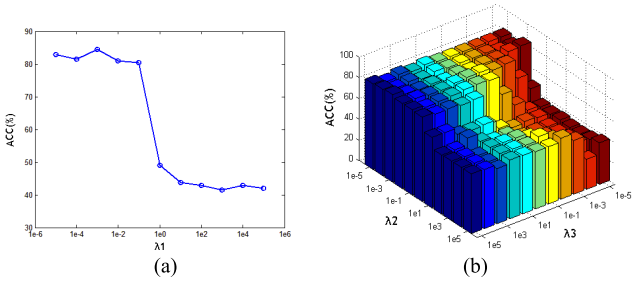



**FIGURE 7.** ACC(%) v.s. (a) parameter $\lambda_1$ by fixing parameters $\lambda_2$ and $\lambda_3$ and (b) parameters $\lambda_2$ and $\lambda_3$ by fixing parameter $\lambda_1$ on the BUAA dataset with 30% paired samples.

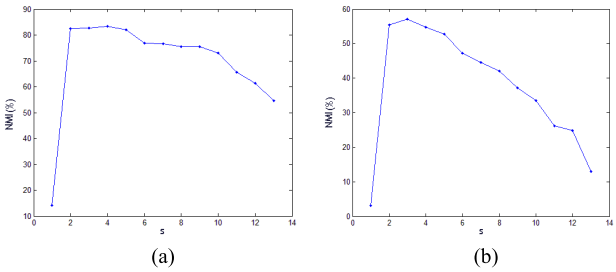



**FIGURE 9.** NMI(%) v.s. parameter $s$ of the proposed approach on (a) BUAA dataset with 30% paired samples and (b) 3sources dataset with 30% missing instances of each view.

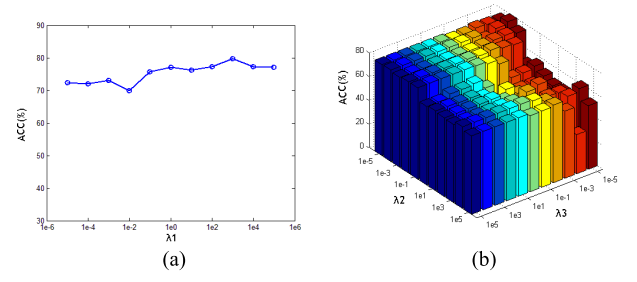



**FIGURE 8.** ACC(%) v.s. (a) parameter $\lambda_1$ by fixing $\lambda_2$ and $\lambda_3$ and (b) parameters $\lambda_2$ and $\lambda_3$ by fixing parameter $\lambda_1$ on the BBCSport dataset with 30% missing instances of each view.

and conduct some experiments on the BUAA and BBCSport datasets to analyze the sensitivity of ACC *w.r.t.* $\lambda_1$, $\lambda_2$ and $\lambda_3$. From Fig. 7, we can see that AWIMVC can obtain encouraging results when $\lambda_1$, $\lambda_2$, and $\lambda_3$ are located in the ranges of $[10^{-5}, 10^{-1}]$, $[10^{-5}, 10^{-3}]$, and $[10^{1}, 10^{5}]$, respectively. From Fig. 8, when $\lambda_1$, $\lambda_2$, and $\lambda_3$ are located in the ranges of $[10^{2}, 10^{5}]$, $[10^{-5}, 10^{-2}]$, and $[10^{1}, 10^{3}]$, respectively, AWIMVC achieves a satisfactory clustering performance. In the experiments, the grid search strategy is exploited to find the three optimal parameters [32]. Moreover, we show the NMI(%) *w.r.t.* $s$ on the BUAA and 3 Sources datasets in Fig. 9. From Fig.9, it can be found that the proposed method obtains a satisfactory clustering result with a small parameter $s$ and we set $s = 3$ in all experiments.

For different datasets, it is still an open problem to adaptively select the optimal values for these parameters. In this paper, we use the following strategy to find the optimal parameters for the proposed method. We fix the parameters $k$ and $s$. And then we determine the parameters $\lambda_1$, $\lambda_2$, and $\lambda_3$. When determining parameters $\lambda_1$, $\lambda_2$, and $\lambda_3$, we define a candidate set $\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}, 10^{4}, 10^{5}\}$ for the three parameters. From Figs. 7 and 8, we can find that the proposed method is relatively insensitive

to the selection of parameter $\lambda_1$ to some extent. Thus we can fix parameter $\lambda_1$ at first, and define a candidate parameter range $\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}, 10^{4}, 10^{5}\}$ for parameters $\lambda_2$ and $\lambda_3$. By performing the proposed method with different combinations of parameters $\lambda_2$ and $\lambda_3$ selected from the candidate range, we can obtain a best combination of these two parameters. Then we fix the two parameters with the obtained best values and use the similar approach to find the optimal value from the candidate parameter range for parameter $\lambda_1$. Finally, we can find the optimal combinations of these three parameters from the 3D space formed by the candidate parameters of $\lambda_1, \lambda_2$, and $\lambda_3$. When determining parameter $s$, we fix the parameters $\lambda_1, \lambda_2$, and $\lambda_3$ with the obtained best values and parameter $k$. We define a candidate set {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 } for parameter $s$. Then perform the proposed method with different values of parameter $s$. In this way, we can find the best value of parameter $s$. When determining parameter $k$, we fix the parameters $s, \lambda_1, \lambda_2$, and $\lambda_3$ with the obtained best values. We define a candidate set {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 } and use the similar approach to find the optimal value from the candidate set for parameter $k$.

### D. EXPERIMENTAL CONVERGENCE STUDY

In this subsection, we conduct the experiments on the BUAA and 3 Sources datasets to show the convergence of the proposed algorithm [33]. ACC and the objective function value *v.s.* iterations on the BUAA and 3Sources datasets are shown in Fig. 10. In Fig. 10, the blue curve shows the objective function value and the red curve indicates the ACC of our method. It can be observed that the objective function value is not increasing after 20 iterations and efficiently converges to the stationary point.

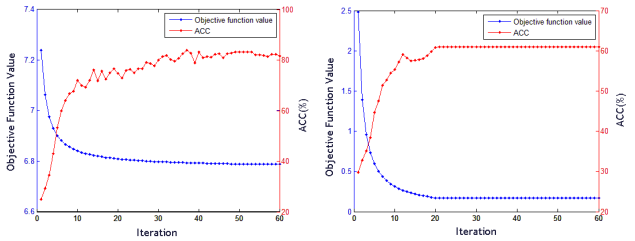



**FIGURE 10.** Objective function value and ACC(%) v.s. iteration on (a) BUAA dataset with 30% paired samples and (b) 3sources dataset with 30% missing instances of each view.

### E. RUNTIME ANALYSIS

In this section, we conduct experiments on Handwritten, BUAA, BBCSport, 3Sources, and Cornell datasets to report the runtime of different methods on these five datasets. From Table 7, we can see that BSV and Concat are much faster than the other methods since the two methods fill the missing instances in every view with the average of instances in the corresponding view and perform k-means on the views.

**TABLE 7.** Runtime(seconds) of different methods.

| | Handwritten | BUAA | BBCSport | 3Sources | Cornell |
|---|---|---|---|---|---|
| BSV | 15.21 | 1.35 | 2.15 | 1.70 | 1.24 |
| Concat | 15.54 | 1.15 | 1.78 | 1.60 | 1.62 |
| PVC | 37.47 | 4.29 | ---- | ---- | 12.00 |
| GPMVC | 72.86 | 4.36 | 156.67 | 193.73 | 17.46 |
| IMG | 15796.98 | 22.25 | ---- | ---- | 37.30 |
| MIC | 1584.25 | 22.53 | 142.05 | 290.81 | 44.18 |
| DAIMC | 511.19 | 9.11 | 2725.32 | 13246.26 | 381.44 |
| OMVC | 977.26 | 8.23 | 85.03 | 248.22 | 383.24 |
| UEAF | 1348.18 | 1.47 | 14.72 | 114.76 | 9.02 |
| AWIMVC | 2203.97 | 3.12 | 15.55 | 117.86 | 25.39 |

However, PVC, GPMVC, IMG, MIC, DAIMC, OMVC, UEAF, and AWIMVC learn a common representation for all views and perform k-means on the learned representation. Although AWIMVC does not cost the least time, the proposed method achieves much better performance than the other methods in most cases.

## VI. CONCLUSION

We developed a method for incomplete multi-view clustering. The proposed method simultaneously learns a common latent representation for all the views, an affinity matrix of the learned representation, and an adaptive weight vector for different views in a unified framework. Furthermore, learning the affinity matrix of the learned representation guides to learn a more discriminative and compact consensus representation, which is beneficial to clustering. Extensive experimental results demonstrate the effectiveness of our method in dealing with the incomplete multi-view clustering tasks.

## REFERENCES

[1] C. Xu, D. Tao, and C. Xu, "A survey on multi-view learning," 2013, pp. 1–59, *arXiv:1304.5634*. [Online]. Available: http://arxiv.org/abs/1304.5634

[2] S. Sun, "A survey of multi-view machine learning," *Neural Comput. Appl.*, vol. 23, nos. 7–8, pp. 2031–2038, Dec. 2013.

[3] S. Bickel and T. Scheffer, "Multi-view clustering," in *Proc. 4th IEEE Int. Conf. Data Mining (ICDM)*, vol. 4, Jul. 2004, pp. 19–26.

[4] U. von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.

[5] V. R. de Sa, "Spectral clustering with two views," in *Proc. ICML*, 2005, pp. 20–27.

[6] X. Cao, C. Zhang, H. Fu, S. Liu, and H. Zhang, "Diversity-induced multi-view subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 586–594.

[7] Y. Wang, L. Wu, X. Lin, and J. Gao, "Multiview spectral clustering via structured low-rank matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4833–4843, Oct. 2018.

[8] L. Sun, S. Ji, and J. Ye, "Canonical correlation analysis for multilabel classification: A least-squares formulation, extensions, and analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 1, pp. 194–200, Jan. 2011.

[9] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Comput.*, vol. 16, no. 12, pp. 2639–2664, Dec. 2004.

[10] K. Chaudhuri, S. M. Kakade, K. Livescu, and K. Sridharan, "Multi-view clustering via canonical correlation analysis," in *Proc. 26th Annu. Int. Conf. Mach. Learn. (ICML)*, 2009, pp. 129–136.

[11] X. Shen, Q. Sun, and Y. Yuan, "A unified multiset canonical correlation analysis framework based on graph embedding for multiple feature extraction," *Neurocomputing*, vol. 148, pp. 397–408, Jan. 2015.

[12] H. Gao, F. Nie, X. Li, and H. Huang, "Multi-view subspace clustering," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4238–4246.

[13] J. Liu, C. Wang, J. Gao, and J. Han, "Multi-view clustering via joint nonnegative matrix factorization," in *Proc. SIAM Int. Conf. Data Mining*, May 2013, pp. 252–260.
[14] 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.
[15] S. Y. Li, Y. Jiang, and Z. H. Zhou, "Partial multi-view clustering," in *Proc. AAAI*, 2014, pp. 1968–1974.
[16] W. Shao, L. He, and P. S. Yu, "Multiple incomplete views clustering via weighted nonnegative matrix factorization with $L_{2,1}$ regularization," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, 2015, pp. 318–334.
[17] H. Zhao, H. Liu, and Y. Fu, "Incomplete multi-modal visual data grouping," in *Proc. Int. Joint Conf. Artif. Intell.*, 2016, pp. 2392–2398.
[18] D. Kong, C. Ding, and H. Huang, "Robust nonnegative matrix factorization using $L_{2,1}$-norm," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, 2011, pp. 673–682.
[19] B. Wu, E. Wang, Z. Zhu, W. Chen, and P. Xiao, "Manifold NMF with $L_{2,1}$ norm for clustering," *Neurocomputing*, vol. 273, pp. 78–88, Jan. 2018.
[20] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.
[21] J. Wen, Z. Zhang, Y. Xu, B. Zhang, L. Fei, and H. Liu, "Unified embedding alignment with missing views inferring for incomplete multi-view clustering," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 5393–5400.
[22] F. Nie, J. Li, and X. Li, "Self-weighted multiview clustering with multiple graphs," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 2564–2570.
[23] Y. Li, F. Nie, H. Huang, and J. Huang, "Large-scale multi-view spectral clustering via bipartite graph," in *Proc. AAAI*, 2015, pp. 2750–2756.
[24] Y. Shen, Z. Wen, and Y. Zhang, "Augmented lagrangian alternating direction method for matrix separation based on low-rank factorization," *Optim. Methods Softw.*, vol. 29, no. 2, pp. 239–263, Mar. 2014.
[25] X. Guo, "Robust subspace segmentation by simultaneously learning data representations and their affinity matrix," in *Proc. Int. Joint Conf. Artif. Intell.*, 2015, pp. 3547–3553.
[26] Z. Zhang, L. Liu, F. Shen, H. T. Shen, and L. Shao, "Binary multi-view clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1774–1782, Jul. 2019.
[27] X. Cai, F. Nie, and H. Huang, "Multi-view k-means clustering on big data," in *Proc. Int. Joint Conf. Artif. Intell.*, 2013, pp. 2598–2604.
[28] D. Huang, J. Sun, and Y. Wang, "The BUAA-VisNir face database instructions," School Comput. Sci. Eng., Beihang Univ., Beijing, China, Tech. Rep. IRIP-TR-12-FR-001, 2012.
[29] N. Rai, S. Negi, S. Chaudhury, and O. Deshmukh, "Partial multi-view clustering using graph regularized NMF," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 2192–2197.
[30] M. Hu and S. Chen, "Doubly aligned incomplete multi-view clustering," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 2262–2268.
[31] W. Shao, L. He, C.-T. Lu, and P. S. Yu, "Online multi-view clustering with incomplete views," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2016, pp. 1012–1017.
[32] J. Wen, B. Zhang, Y. Xu, J. Yang, and N. Han, "Adaptive weighted nonnegative low-rank representation," *Pattern Recognit.*, vol. 81, pp. 326–340, Sep. 2018.
[33] G.-S. Xie, X.-Y. Zhang, S. Yan, and C.-L. Liu, "SDE: A novel selective, discriminative and equalizing feature representation for visual recognition," *Int. J. Comput. Vis.*, vol. 124, no. 2, pp. 145–168, Sep. 2017.

**WANYU DENG** (Associate Member, IEEE) received the B.S. degree in computer science and technology and the M.S. degree in software technology and theory from Northwest Polytechnical University, China, in 2001 and 2004, respectively, and the Ph.D. degree from the Department of Computer Science and Technology from Xian Jiaotong University, China, in 2010.

He has participated in several national research projects and got FUJIXEROX scholarships of Xian Jiaotong University. He is the author or coauthor of more than 30 refereed international journal and conference papers covering topics of collaborative filtering, personalized service, and artificial intelligence. His research interests include machine learning, collaborative filtering, and personalized service. He won the Science Star Scholars of Shaanxi province, in 2013. Till now, he has held four fund projects, including National Natural Science Foundation of China, and Science Plan Project of Shaanxi province.

**LIXIA LIU** received the B.S. degree in information and computing science from the Xi'an University of Posts and Telecommunications, China, in 2018, where she is currently pursuing the M.S. degree with the Technology of Computer Application. Her research interests include artificial intelligence, machine learning, and data mining.

**JIANQIANG LI** received the B.S. degree in network engineering from Zhengzhou Technology and Business University, China, in 2018. He is currently pursuing the M.S. degree with the Computer Technology, Xi'an University of Posts and Telecommunications, China. His research interests include deep learning and computer vision.

**YIJUN LIN** is currently pursuing the B.S. degree with the Data Science and Big Data Technology, Xi'an University of Posts and Telecommunications, China.