Incomplete Multi-View Clustering With Reconstructed Views

Jun Yin and Shiliang Sun

Abstract—As one category of important incomplete multi-view clustering methods, subspace based methods seek the common latent representation of incomplete multi-view data by matrix factorization and then partition the latent representation to get clustering results. However, these methods ignore missing views in the process of matrix factorization, which makes the connection of different views be exploited inadequately. This paper proposes Incomplete Multi-view Clustering with Reconstructed Views (IMCRV), which utilizes the incomplete examples sufficiently. In IMCRV, the missing views of incomplete examples are reconstructed and the reconstructed views are also used to seek the common latent representation. IMCRV also involves the Laplacian regularization to preserve the global property of the latent representation. The gradient descent method with the multiplicative update rule is employed to solve the objective function of IMCRV. The corresponding iterative algorithm is developed and the convergence of the algorithm is proved. IMCRV is compared with many state-of-the-art incomplete multi-view clustering methods under different Incomplete Example Rates (IER) on public multi-view datasets. The experimental results demonstrate the superior effectiveness of IMCRV.

Index Terms—Multi-view clustering, incomplete view, reconstructed view, gradient descent, nonnegative matrix factorization

1 Introduction

In THE real world, many kinds of data appear in multiple views. For instance, news can be reported by different media, a web page can be described by both the image and text, and an image can be represented by various visual features. Different views contain complementary information, and these information cannot be utilized sufficiently by concatenating multiple views directly. Multi-view learning [1], [2], [3], [4] was developed to exploit complementary information of different views and became a hot research topic. Numerous multi-view learning methods such as multi-view dimension reduction [5], [6], [7], multi-view discriminant analysis [8], [9], [10], multi-view support vector machine [11], [12], [13] and multi-view clustering [14], [15], [16] were proposed to solve specific machine learning problems.

As one of important research directions of multi-view learning, multi-view clustering tries to partition multi-view data into clusters in an unsupervised situation. Bickel and Scheffer [17] developed multi-view versions of EM and K-means, which are much more effective than their single-view versions. Chaudhuri *et al.* [18] used Canonical Correlation Analysis (CCA) to project multi-view data into a low dimensional space and then performed clustering in the low dimensional space. Liu *et al.* [19] proposed Nonnegative Matrix Factorization based Multi-view Clustering (MultiNMF), which obtains compatible clustering solutions across multiple views by searching the consensus coefficient matrix. Kumar *et al.* [20] applied co-regularization to spectral clustering to make the clusterings in different views agree with each other and solved the corresponding optimization problem by alternative maximization. Nie *et al.*[21] presented Self-weighted Multi-view Clustering, where the weight of each view is self-learned and the clustering result is obtained directly by the Laplacian rank constrained graph. Kang *et al.* [22] integrated graph fusion and spectral clustering into a framework, which makes the fusion graph approximate the original graph of each view but maintain an explicit cluster structure. Tao *et al.* [23] jointly performed marginalized denoising, consensus representation learning and spectral graph partitioning in a unified optimization framework.

The above multi-view clustering methods assume the completeness of the multi-view data, that is, all the views of the examples exist. However, in real applications, it is common to see that some views of data are missing. For example, for a piece of news, maybe some media report it and others not. In a web page, only the text information is contained while the image not. In these situations, the traditional multi-view clustering methods cannot be conducted directly. Recently, a number of incomplete multi-view clustering methods [24], [25], [26] were proposed to address the incomplete-view problem. It is a simple way to fill the missing views directly. Shao *et al.* [27] filled the missing views with the average value of the existing features and introduced the

- *Jun Yin is with the College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China, and also with the School of Computer Science and Technology, East China Normal University, Shanghai 200241, China. E-mail: junyin@shmtu.edu.cn.*
- *Shiliang Sun is with the School of Computer Science and Technology, East China Normal University, Shanghai 200241, China. E-mail: slsun@cs.ecnu.edu.cn.*

Manuscript received 5 November 2020; revised 29 July 2021; accepted 9 September 2021. Date of publication 13 September 2021; date of current version 3 February 2023.
(Corresponding author: Shiliang Sun.)
Recommended for acceptance by Y. Zhang.
Digital Object Identifier no. 10.1109/TKDE.2021.3112114

dynamic weight for the missing data to minimize the influence of incomplete views. Yu *et al.* [28] proposed Complex Mapping Multi-View Clustering (CMMVC), which fills the missing views with 0, and explores the complex mapping relationship between views by using the nearest neighbor relationship between cross-view examples. Another category of methods is based on kernel technique which fills the kernel matrices of incomplete views. Trivedi *et al.* [29] constructed the kernel matrix corresponding to the incomplete view using Laplacian regularization and then performed kernel CCA based multi-view clustering algorithm with the obtained kernel matrix. Shao *et al.* [30] computed the kernel matrices of incomplete views by optimizing the alignment of shared examples of the views. Liu *et al.* [31] integrated imputation of incomplete kernels and clustering into an unified objective, which makes the imputed kernels more adaptive to clustering. The third category of methods is based on subspace, which transforms the incomplete multi-view data into a new common space. Li *et al.* [32] projected the incomplete-view data into a latent subspace by Nonnegative Matrix Factorization (NMF) and ℓ_1 norm regularization, where the data belonging to the same example in different views have the same representation. Zhao *et al.* [33] also transformed the incomplete-view data into a latent subspace using matrix factorization, and in their method, the Frobenius norm regularization was employed and a graph Laplacian term was involved to couple the incomplete multi-view data. Hu and Chen [34] learned a common latent feature subspace by semi-NMF and they established a consensus basis matrix for all the views to reduce the influence of the missing data. Zong *et al.* [35] proposed Multi-View clustering for Partially mapped cluster and instance (MVP), which separately handles complete examples and incomplete examples, and connects multiple views to find the latent representation by complete examples. Wen *et al.* [36] developed Adaptive Graph Completion based Incomplete Multi-view Clustering (AGC-IMC), which borrows the idea of multi-view spectral clustering and jointly performs the graph completion and consensus representation learning in a unified framework. Tao *et al.* [37] proposed the joint embedding learning and low-rank approximation framework, which approximates the incomplete multi-view data by low-rank matrices and learns a common embedding by linear transformation. Besides, Zhuge *et al.* [38], [39] seeked low dimensional representations of existing data for each view and a common probability label matrix simultaneously. The clustering results can be obtained from the label matrix directly.

The previous incomplete multi-view clustering methods have the following limitations. Unsatisfactory results are usually obtained by filling the missing views directly. These methods often use the average value of the existing features of the same view to fill the missing data, which ignore the connection of different views and also the characteristic of the data distributions. Kernel based methods use the filled kernel matrices to conduct multi-view kernel methods such as kernel CCA and multi-view kernel K-means. However, it is not easy to select an appropriate kernel for each view and set the parameters for each kernel. Subspace based methods have ideal clustering performances by partitioning the common latent representation of different views obtained by matrix factorization. Nevertheless, they ignore the missing views in the process of matrix factorization, which makes that the examples with incomplete views cannot be exploited sufficiently.

In this paper, we propose a novel method Incomplete Multi-view Clustering with Reconstructed Views (IMCRV). IMCRV is designed for partitioning incomplete multi-view data with any number of views. To make use of the connection of different views of all the examples sufficiently, IMCRV reconstructs their missing views for the incomplete examples. The reconstructed views are adopted to seek the latent representation, and conversely, the latent representation is used to reconstruct the missing views. In IMCRV, the Laplacian regularization is also employed to bridge the latent representations of all the examples. An iterative algorithm is developed to solve the objective of IMCRV. In the iteration algorithm, the gradient descent method with the multiplicative update rule is adopted. The convergence proof of the algorithm is presented. Clustering experiments are conducted on public multi-view datasets and the experimental results show that IMCRV performs much better than state-of-the-art multi-view clustering methods.

The main contributions of this paper are highlighted as follows:

1) A novel clustering method IMCRV is proposed to partition incomplete multi-view data with any number of views. Based on the results of NMF, IMCRV reconstructs missing views and the reconstructed views are also used to seek the latent representation.
2) The Laplacian regularization term is integrated with NMF to constitute the objective of IMCRV. This objective is solved by an iteration algorithm, which adopts the gradient descent method with the multiplicative update rule.
3) The convergence of the iteration algorithm of IMCRV is proved.

2 Related Works

2.1 Partial Multi-View Clustering (PVC)

Given a set of examples $X = [x_1, x_2, \ldots, x_N]$, where N is the total number of examples. Each example $x_i = (x_i^{(1)}, x_i^{(2)})$ has two views. For some examples, view 1 or view 2 is missing. Then the incomplete two-view dataset can be represented as $\hat{X} = \{\hat{X}^{(1,2)}, \hat{X}^{(1)}, \hat{X}^{(2)}\}$, where $\hat{X}^{(1,2)}$, $\hat{X}^{(1)}$ and $\hat{X}^{(2)}$ denote the examples presented in both views, view 1 and view 2, respectively. Suppose that $\hat{X}^{(1,2)} \in \mathbb{R}^{(d_1+d_2)\times c}$, $\hat{X}^{(1)} \in \mathbb{R}^{d_1\times m}$ and $\hat{X}^{(2)} \in \mathbb{R}^{d_2\times n}$, where c, m and n are the number of examples presented in $\hat{X}^{(1,2)}$, $\hat{X}^{(1)}$ and $\hat{X}^{(2)}$, respectively, $N = c + m + n$, and d_1 and d_2 are the dimensions of view 1 and view 2.

PVC [32] tries to find a latent subspace using NMF and ℓ_1 norm regularization. Let $\hat{X}^{(1,2)} = [X_c^{(1)}; X_c^{(2)}]$ be composed of $X_c^{(1)} \in \mathbb{R}^{d_1\times c}$ and $X_c^{(2)} \in \mathbb{R}^{d_2\times c}$ which come from two views and the dimension of the latent subspace be k. PVC makes $X_c^{(1)}$ and $X_c^{(2)}$ corresponding to the examples with complete views have the same representation, and $\hat{X}^{(1)}$ and $\hat{X}^{(2)}$ corresponding to the examples with incomplete views have

their individual representations. Its objective function is defined as

$$\begin{aligned}\min_{\substack{P_c,\hat{P}^{(1)},\hat{P}^{(2)}\\U^{(1)},U^{(2)}}} & \left\{\begin{array}{c}\left\|[X_c^{(1)},\hat{X}^{(1)}]-U^{(1)}[P_c,\hat{P}^{(1)}]\right\|_F^2+\lambda\left\|[P_c,\hat{P}^{(1)}]^T\right\|_1\\+\left\|[X_c^{(2)},\hat{X}^{(2)}]-U^{(2)}[P_c,\hat{P}^{(2)}]\right\|_F^2+\lambda\left\|[P_c,\hat{P}^{(2)}]^T\right\|_1\end{array}\right\}\\ s.t.\quad & P_c\geq 0,\hat{P}^{(1)}\geq 0,\hat{P}^{(2)}\geq 0,\\ & U^{(1)}\geq 0,U^{(2)}\geq 0,\end{aligned}\tag{1}$$

where $P_c\in\mathbb{R}^{k\times c}$ is the uniform latent representation of $X_c^{(1)}$ and $X_c^{(2)}$, $\hat{P}^{(1)}\in\mathbb{R}^{k\times m}$ and $\hat{P}^{(2)}\in\mathbb{R}^{k\times n}$ are the latent representations of $\hat{X}^{(1)}$ and $\hat{X}^{(2)}$, respectively, and $U^{(1)}\in\mathbb{R}^{d_1\times k}$ and $U^{(2)}\in\mathbb{R}^{d_2\times k}$ are the basis matrices of the latent subspaces for two views. Eq. (1) is solved by an iterative algorithm. The latent representation for all the examples is $P=[P_c,\hat{P}^{(1)},\hat{P}^{(2)}]\in\mathbb{R}^{k\times(c+m+n)}$, which is concatenated by P_c, $\hat{P}^{(1)}$ and $\hat{P}^{(2)}$.

2.2 Incomplete Multi-Modal Visual Data Grouping (IMVDG)

IMVDG [33] pursuits a latent subspace by matrix factorization and Frobenius norm regularization. It removes the nonnegative constraint to make the optimization easier. To get the global property which is important in subspace clustering, IMVDG incorporates the graph Laplacian into its objective. The objective function of IMVDG is formulated as

$$\begin{aligned}\min_{\substack{P_c,\hat{P}^{(1)},\hat{P}^{(2)}\\U^{(1)},U^{(2)},W}} & \left\|[X_c^{(1)},\hat{X}^{(1)}]-U^{(1)}[P_c,\hat{P}^{(1)}]\right\|_F^2+\lambda\left\|U^{(1)}\right\|_F^2\\ & +\left\|[X_c^{(2)},\hat{X}^{(2)}]-U^{(2)}[P_c,\hat{P}^{(2)}]\right\|_F^2+\lambda\left\|U^{(2)}\right\|_F^2\\ & +\beta\mathrm{tr}\left(PL_WP^T\right)+\gamma\|W\|_F^2\\ s.t.\quad & \forall i\quad W_i^T\mathbf{1}=1,W_i\geq 0,\end{aligned}\tag{2}$$

where $W\in\mathbb{R}^{N\times N}$ is the similarity matrix with each element denoting the similarity of two examples in the latent subspace, $L_W=D-W$ is the Laplacian matrix of W, in which D is the diagonal matrix with $D_{ii}=\sum_{j=1}^{N}W_{ij}$. The graph Laplacian term $\beta\mathrm{tr}(PL_WP^T)$ bridges the connection of all the examples. Eq. (2) is solved by the augmented Lagrange multiplier with alternating direction minimizing strategy [40].

2.3 Doubly Aligned Incomplete Multi-View Clustering (DAIMC)

DAIMC [34] factorizes the data matrix of each view into the basis matrix and the latent representation by weighted semi-NMF. The weight is employed to eliminate missing views in the process of matrix factorization. To reduce the influence of missing views, it also aligns the basis matrices of all the views by ℓ_{21} norm regularization regression. The objective function of DAIMC is defined as

$$\begin{aligned}\min_{U^{(v)},P,B^{(v)}} & \sum_{v=1}^{V}\left(\begin{array}{c}\left\|(X^{(v)}-U^{(v)}P)A^{(v)}\right\|_F^2\\+\alpha\left(\left\|B^{(v)^T}U^{(v)}-I_k\right\|_F^2+\beta\left\|B^{(v)}\right\|_{2,1}\right)\end{array}\right)\\ s.t.\quad & P\geq 0,\end{aligned}\tag{3}$$

where V is the number of views, $X^{(v)}\in\mathbb{R}^{d_v\times N}$ is the data of view v, $U^{(v)}\in\mathbb{R}^{d_v\times k}$ is the basis matrix of the latent subspace for view v, $P\in\mathbb{R}^{k\times N}$ is the common latent representation, $A^{(v)}\in\mathbb{R}^{N\times N}$ is the weight matrix of view v, $B^{(v)}\in\mathbb{R}^{d_v\times k}$ is the regression coefficient matrix of view v and $I_k\in\mathbb{R}^{k\times k}$ is an identity matrix. $A^{(v)}$ is a diagonal matrix defined as

$$A_{ii}^{(v)}=\begin{cases}1, & \text{if view } v \text{ of the example } x_i \text{ exists}\\ 0, & \text{otherwise.}\end{cases}\tag{4}$$

3 Incomplete Multi-View Clustering With Reconstructed Views (IMCRV)

3.1 Formulation

To handle missing views, the examples with complete views and incomplete views are processed in different ways in both of PVC and IMVDG. For the examples with complete views, the latent representations of all views are enforced to be identical. For the examples with incomplete views, only the latent representations of existing views are considered while missing views are neglected. In DAIMC, a weight matrix is introduced to make missing views not be involved in matrix factorization. These three methods all ignore the missing views in the process of seeking the latent representation. The examples with incomplete views are not exploited sufficiently.

Based on the assumption that multi-view data are derived from a common subspace, all the views of any example have the same representation in the latent subspace, which establishes the relationship of different views. Although missing views of an example do not exist, they share the common latent representation with existing views of this example and can be reconstructed by the latent representation. The reconstructed views can also be used to seek the latent representation. This objective is formulated as

$$\begin{aligned}\min_{X^{(v)},U^{(v)},P} & \sum_{v=1}^{V}\left(\left\|X^{(v)}-U^{(v)}P\right\|_F^2+\lambda\left\|U^{(v)}\right\|_F^2\right)\\ s.t.\quad & X_E^{(v)}=X^{(v)}S_E^{(v)},\ X^{(v)}\geq 0,\ U^{(v)}\geq 0,\ P\geq 0,\end{aligned}\tag{5}$$

where $X_E^{(v)}\in\mathbb{R}^{d_v\times N_E^{(v)}}$ is the existing data of view v, in which $N_E^{(v)}$ is the number of existing data for view v, and $S_E^{(v)}\in\mathbb{R}^{N\times N_E^{(v)}}$ is the selection matrix of existing data for view v, which is formed by removing the columns of identity matrix $I\in\mathbb{R}^{N\times N}$ corresponding to the missing data. In Eq. (5), we employ NMF and Frobenius norm regularization. The uniform latent representation P is computed by matrix factorization of $X^{(v)}$, and simultaneously, the complete data $X^{(v)}$ is calculated by reconstruction with $U^{(v)}$ and P. In this process, the relationship of different views of all the examples is utilized sufficiently. The constraint $X_E^{(v)}=X^{(v)}S_E^{(v)}$ makes the reconstructed data only effective for the missing data and the original existing data be preserved.

Global property bridges the connection of all the examples, which is crucial to get a good clustering result. To preserve global property of the latent representation P, we also incorporate the graph Laplacian term $\mathrm{tr}\left(PL_{\sim W}P^T\right)=$

$\sum_{i=1}^{N}\sum_{j=1}^{N}\|P_i - P_j\|_F^2 W_{ij}$ into the objective function of IMCRV. $L_{\widetilde{W}} = \widetilde{D} - \widetilde{W}$ is the Laplacian matrix of $\widetilde{W}$, in which $\widetilde{W}_{ij} = W_{ij} + W_{ji}$ and $\widetilde{D}$ is the diagonal matrix with $\widetilde{D}_{ii} = \sum_{j=1}^{N}(W_{ij} + W_{ji})$. The graph Laplacian term makes any two examples with high similarity close to each other in the latent subspace, which is consistent with the criterion of clustering. The similarity matrix W appearing in graph Laplacian should be regularized to avoid degenerate solutions, where only one element of each column of W equals 1 and all the other elements equal 0. Here, we employ the Frobenius norm and the Shannon entropy for regularization, respectively. Through these two regularizations, all the elements of W will approach $1/N$ and the degenerate solutions are avoided. With the Frobenius norm regularization of W, the final objective function of IMCRV is formulated as

$$\begin{aligned}\min_{X^{(v)},U^{(v)},P,W} & L_1(X^{(v)},U^{(v)},P,W)\\ =&\sum_{v=1}^{V}\Big(\big\|X^{(v)}-U^{(v)}P\big\|_F^2+\lambda\big\|U^{(v)}\big\|_F^2\Big)\\ &+\ \beta\mathrm{tr}\Big(PL_{\widetilde{W}}P^T\Big)+\gamma\|W\|_F^2\\ s.t.\quad & X_E^{(v)}=X^{(v)}S_E^{(v)},\ X^{(v)}\geq 0,\ U^{(v)}\geq 0,\ P\geq 0,\\ & \forall i\quad W_i^T\mathbf{1}=1, W_i\geq 0.\end{aligned} \tag{6}$$

With the Shannon entropy regularization of W, we have the formulation of the final objective function of IMCRV as

$$\begin{aligned}\min_{X^{(v)},U^{(v)},P,W} & L_2(X^{(v)},U^{(v)},P,W)\\ =&\sum_{v=1}^{V}\Big(\big\|X^{(v)}-U^{(v)}P\big\|_F^2+\lambda\big\|U^{(v)}\big\|_F^2\Big)\\ &+\ \beta\mathrm{tr}\Big(PL_{\widetilde{W}}P^T\Big)+\gamma\sum_{i=1}^{N}\sum_{j=1}^{N}W_{ij}\ln W_{ij}\\ s.t.\quad & X_E^{(v)}=X^{(v)}S_E^{(v)},\ X^{(v)}\geq 0,\ U^{(v)}\geq 0,\ P\geq 0,\\ & \forall i\quad W_i^T\mathbf{1}=1, W_i\geq 0.\end{aligned} \tag{7}$$

After obtaining the latent representation P for all the examples by Eqs. (6) or (7), we can conduct standard clustering methods to get the clustering result of IMCRV.

3.2 Optimization

There are four variables in Eqs. (6) and (7), which are not jointly convex optimization problems. However, they are convex in any one variable with the other three given. Therefore, these two optimization problems can be solved by iterative update. In the $t+1$ iteration, the iterative update is conducted as follows.

1) Update $X^{(v)}$ with fixed $U^{(v)}$, P and W

According to Eqs. (6) and (7), it is easy to know $X_{(t+1)}^{(v)} = U_{(t)}^{(v)}P_{(t)}$ without the constraint $X_E^{(v)} = X^{(v)}S_E^{(v)}$. Considering the constraint, we can only apply the reconstruction $U_{(t)}^{(v)}P_{(t)}$ to the missing data and preserve the original existing data. The reconstructed view of the missing data is

$$X_{M(t+1)}^{(v)} = U_{(t)}^{(v)}P_{(t)}S_M^{(v)}, \tag{8}$$

where $S_M^{(v)} \in \mathbb{R}^{N\times N_M^{(v)}}$ is the selection matrix of the missing data for view v, in which $N_M^{(v)}$ is the number of the missing data for view v and $N = N_E^{(v)} + N_M^{(v)}$. $S_M^{(v)}$ is formed by removing the columns of identity matrix $I \in \mathbb{R}^{N\times N}$ corresponding to the indexes of existing data in view v. Combining $X_E^{(v)}$ and $X_{M(t+1)}^{(v)}$, we can obtain $X_{(t+1)}^{(v)}$ as

$$\begin{aligned}X_{(t+1)}^{(v)} &= X_{M(t+1)}^{(v)}F_M^{(v)} + X_E^{(v)}F_E^{(v)}\\ &= U_{(t)}^{(v)}P_{(t)}S_M^{(v)}F_M^{(v)} + X_E^{(v)}F_E^{(v)},\end{aligned} \tag{9}$$

where $F_M^{(v)}$ is the filling matrix of the missing data for view v which is formed by successively replacing the columns of zero matrix $O_M^{(v)} \in \mathbb{R}^{N_M^{(v)}\times N}$ corresponding to the indexes of the missing data in view v with the columns of identity matrix $I_M^{(v)} \in \mathbb{R}^{N_M^{(v)}\times N_M^{(v)}}$, and $F_E^{(v)}$ is the filling matrix of existing data for view v which is formed by successively replacing the columns of zero matrix $O_E^{(v)} \in \mathbb{R}^{N_E^{(v)}\times N}$ corresponding to the indexes of existing data in view v with the columns of identity matrix $I_E^{(v)} \in \mathbb{R}^{N_E^{(v)}\times N_E^{(v)}}$.

In the update of $X^{(v)}$, the missing data $X_{M(t+1)}^{(v)}$ is reconstructed by Eq. (8) and the complete data $X_{(t+1)}^{(v)}$ is obtained by Eq. (9), which combines the existing data $X_E^{(v)}$ and the missing data $X_{M(t+1)}^{(v)}$. The existing data $X_E^{(v)}$ remains unchanged in the iteration.

2) Update $U^{(v)}$ with fixed $X^{(v)}$, P and W

We use gradient descent to seek $U^{(v)}$ minimizing the objective function in Eqs. (6) and (7). With fixed P and W, the terms $\beta\mathrm{tr}(PL_{\widetilde{W}}P^T)$, $\gamma\|W\|_F^2$ and $\gamma\sum_{i=1}^{N}\sum_{j=1}^{N}W_{ij}\ln W_{ij}$ are fixed. Now we need to minimize

$$L(U^{(v)}) = \sum_{v=1}^{V}\Big(\big\|X^{(v)}-U^{(v)}P\big\|_F^2+\lambda\big\|U^{(v)}\big\|_F^2\Big). \tag{10}$$

The derivative of $L(U^{(v)})$ with respect to $U^{(v)}$ can be easily represented as

$$\bigtriangledown L(U^{(v)}) = 2\Big(U^{(v)}PP^T - X^{(v)}P^T + \lambda U^{(v)}\Big). \tag{11}$$

According to gradient descent, the update rule of $U^{(v)}$ is

$$\begin{aligned}&U_{(t+1)ir}^{(v)}\\ =&U_{(t)ir}^{(v)} - 2\mu_{ir}\Big(U_{(t)}^{(v)}P_{(t)}P_{(t)}^T - X_{(t+1)}^{(v)}P_{(t)}^T + \lambda U_{(t)}^{(v)}\Big)_{ir}.\end{aligned} \tag{12}$$

Using the multiplicative update rule [41], we set the learning rate μ_{ir} as

$$\mu_{ir} = \frac{U_{(t)ir}^{(v)}}{2\Big(U_{(t)}^{(v)}P_{(t)}P_{(t)}^T + \lambda U_{(t)}^{(v)}\Big)_{ir}}. \tag{13}$$

Substituting Eq. (13) into Eq. (12), we have

$$U_{(t+1)ir}^{(v)} = \frac{U_{(t)ir}^{(v)}\Big(X_{(t+1)}^{(v)}P_{(t)}^T\Big)_{ir}}{\Big(U_{(t)}^{(v)}P_{(t)}P_{(t)}^T + \lambda U_{(t)}^{(v)}\Big)_{ir}}. \tag{14}$$

3) Update P with fixed $X^{(v)}$, $U^{(v)}$ and W

Like the update of $U^{(v)}$, P is also updated with gradient descent. Now the terms $\sum_{v=1}^{V}\lambda\big\|U^{(v)}\big\|_F^2$, $\gamma\|W\|_F^2$ and $\gamma\sum_{i=1}^{N}\sum_{j=1}^{N}W_{ij}\ln W_{ij}$ are fixed. We should minimize

$$L(P)=\sum_{v=1}^{V}\left\|X^{(v)}-U^{(v)}P\right\|_F^2+\beta\mathrm{tr}\left(PL_{\widetilde{W}}P^T\right). \tag{15}$$

We have the derivative of $L(P)$ with respect to P as

$$\begin{aligned}\bigtriangledown L(P)&=2\left(\sum_{v=1}^{V}\left({U^{(v)}}^TU^{(v)}P-{U^{(v)}}^TX^{(v)}\right)+\beta PL_{\widetilde{W}}\right)\\&=2\left(\begin{array}{l}\sum_{v=1}^{V}\left({U^{(v)}}^TU^{(v)}P-{U^{(v)}}^TX^{(v)}\right)\\+\beta P\widetilde{D}-\beta P\widetilde{W}\end{array}\right).\end{aligned} \tag{16}$$

P is updated as

$$\begin{aligned}&P_{rj}^{(t+1)}=P_{rj}^{(t)}\\&-2\eta_{rj}\left(\begin{array}{l}\sum_{v=1}^{V}\left({U_{(t+1)}^{(v)}}^TU_{(t+1)}^{(v)}P_{(t)}-{U_{(t+1)}^{(v)}}^TX_{(t+1)}^{(v)}\right)\\+\beta P_{(t)}\widetilde{D}_{(t)}-\beta P_{(t)}\widetilde{W}_{(t)}\end{array}\right)_{rj}.\end{aligned} \tag{17}$$

According to the multiplicative update rule and keeping η_{rj} nonnegative, we set the learning rate

$$\eta_{rj}=\frac{P_{rj}^{(t)}}{2\left(\sum_{v=1}^{V}{U_{(t+1)}^{(v)}}^TU_{(t+1)}^{(v)}P_{(t)}+\beta P_{(t)}\widetilde{D}_{(t)}\right)_{rj}}. \tag{18}$$

Combining Eqs. (17) and (18), we have

$$P_{rj}^{(t+1)}=\frac{P_{rj}^{(t)}\left(\sum_{v=1}^{V}{U_{(t+1)}^{(v)}}^TX_{(t+1)}^{(v)}+\beta P_{(t)}\widetilde{W}_{(t)}\right)_{rj}}{\left(\sum_{v=1}^{V}{U_{(t+1)}^{(v)}}^TU_{(t+1)}^{(v)}P_{(t)}+\beta P_{(t)}\widetilde{D}_{(t)}\right)_{rj}}. \tag{19}$$

4) Update W with fixed $X^{(v)}$, $U^{(v)}$ and P

Since the term $\sum_{v=1}^{V}(\|X^{(v)}-U^{(v)}P\|_F^2+\lambda\|U^{(v)}\|_F^2)$ is fixed, we consider the following two optimization problems as

$$\begin{aligned}\min_{W}\quad&\beta\mathrm{tr}\left(PL_{\widetilde{W}}P^T\right)+\gamma\|W\|_F^2\\s.t.\quad&\forall i\quad W_i^T\mathbf{1}=1,W_i\geq 0,\end{aligned} \tag{20}$$

and

$$\begin{aligned}\min_{W}\quad&\beta\mathrm{tr}\left(PL_{\widetilde{W}}P^T\right)+\gamma\sum_{i=1}^{N}\sum_{j=1}^{N}W_{ij}\ln W_{ij}\\s.t.\quad&\forall i\quad W_i^T\mathbf{1}=1,W_i\geq 0.\end{aligned} \tag{21}$$

According to $\mathrm{tr}\left(PL_{\widetilde{W}}P^T\right)=\sum_{i=1}^{N}\sum_{j=1}^{N}\left\|P_i-P_j\right\|_F^2W_{ij}$, for Eq. (20), we have

$$\begin{aligned}&\min_{W}\quad\beta\mathrm{tr}\left(PL_{\widetilde{W}}P^T\right)+\gamma\|W\|_F^2\\\Leftrightarrow&\min_{W}\quad\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\beta\left\|P_i-P_j\right\|_F^2W_{ij}+\gamma W_{ij}^2\right)\\\Leftrightarrow&\forall i\quad\min_{W_i}\quad\sum_{j=1}^{N}\left(W_{ij}+\frac{\beta\left\|P_i-P_j\right\|_F^2}{2\gamma}\right)^2\\\Leftrightarrow&\forall i\quad\min_{W_i}\quad\|W_i+Q_i\|_F^2,\end{aligned} \tag{22}$$

where Q_i is a column vector with the jth element defined as $Q_{ij}=\frac{\beta\left\|P_i-P_j\right\|_F^2}{2\gamma}$. Now Eq. (20) can be transformed to

$$\begin{aligned}\forall i\quad\min_{W_i}\quad&\|W_i+Q_i\|_F^2\\s.t.\quad&W_i^T\mathbf{1}=1,W_i\geq 0.\end{aligned} \tag{23}$$

The closed form solution of Eq. (23) is

$$W_i=\left(\frac{1+\sum_{j=1}^{l}\widetilde{Q}_{ij}\mathbf{1}}{l}-Q_i\right)_+, \tag{24}$$

where the operator $(R)_+=max(R,O)$, in which O is a zero matrix with the same size as R, l is the number of nearest neighbors of P_i which have chance to connect to P_i, $\widetilde{Q}_i$ is a vector whose elements are the elements of Q_i with the ascending order. The detailed proof of Eq. (24) is presented in [42]. Thus, the update rule of W for Eq. (6) is

$$W_i^{(t+1)}=\left(\frac{1+\sum_{j=1}^{l}\widetilde{Q}_{ij}^{(t+1)}\mathbf{1}}{l}-Q_i^{(t+1)}\right)_+. \tag{25}$$

For Eq. (21), we have

$$\begin{aligned}&\min_{W}\quad\beta\mathrm{tr}\left(PL_{\widetilde{W}}P^T\right)+\gamma\sum_{i=1}^{N}\sum_{j=1}^{N}W_{ij}\ln W_{ij}\\&s.t.\quad\forall i\quad W_i^T\mathbf{1}=1,W_i\geq 0\\\Leftrightarrow&\min_{W}\quad\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\beta\left\|P_i-P_j\right\|_F^2W_{ij}+\gamma W_{ij}\ln W_{ij}\right)\\&s.t.\quad\forall i,j\quad\sum_{j=1}^{N}W_{ij}=1,W_{ij}\geq 0.\end{aligned} \tag{26}$$

Eq. (26) can be solved by the Lagrangian multiplier method and its solution is

$$W_{ij}=\frac{exp\left(-\frac{\beta\left\|P_i-P_j\right\|_F^2}{\gamma}\right)}{\sum_{j=1}^{N}exp\left(-\frac{\beta\left\|P_i-P_j\right\|_F^2}{\gamma}\right)}. \tag{27}$$

The detailed deduction can be found in [43]. Then, for Eq. (7), W is updated as

$$W_{ij}^{(t+1)}=\frac{exp\left(-\frac{\beta\left\|P_i^{(t+1)}-P_j^{(t+1)}\right\|_F^2}{\gamma}\right)}{\sum_{j=1}^{N}exp\left(-\frac{\beta\left\|P_i^{(t+1)}-P_j^{(t+1)}\right\|_F^2}{\gamma}\right)}. \tag{28}$$

From Eqs. (9), (14) and (19), it can be seen that, to make $X^{(v)}\geq 0$, $U^{(v)}\geq 0$ and $P\geq 0$, we should guarantee the input data $X_E^{(v)}\geq 0$. In Section 3.5, we will prove that the iteration can converge when $X^{(v)}\geq 0$, $U^{(v)}\geq 0$ and $P\geq 0$. We summarize the iterative algorithm of IMCRV in Algorithm 1. The convergence criterion of Algorithm 1 is set as

$$|(J_1(t+1)-J_1(t))/J_1(t)|<\epsilon, \tag{29}$$

where $J_1(t)$ is the value of the objective (6) or (7).

Algorithm 1. IMCRV Algorithm

Input:
Incomplete multi-view data $X_E^{(v)}|_{v=1}^V$ from V views; parameters λ, β and γ.

Output:
Complete multi-view data $X^{(v)}|_{v=1}^V$; basis matrices $U^{(v)}|_{v=1}^V$; uniform latent representation P; similarity matrix W.

1: Initialize $U^{(v)}|_{v=1}^V$, P and W.
2: **repeat**
3: With fixed $U^{(v)}|_{v=1}^V$ and P, update $X^{(v)}|_{v=1}^V$ using Eq. (9).
4: With fixed $X^{(v)}|_{v=1}^V$ and P, update $U^{(v)}|_{v=1}^V$ using Eq. (14).
5: With fixed $X^{(v)}|_{v=1}^V$, $U^{(v)}|_{v=1}^V$ and W, update P using Eq. (19).
6: With fixed P, update W using Eqs. (25) or (28).
7: **until** Eqs. (6) or (7) converges.

3.3 Initialization

The variables $U^{(v)}$, P and W need to be initialized to perform the optimization of IMCRV. The objective of IMCRV is a nonconvex optimization problem and Algorithm 1 will converge to a local mimimum. As the solution of IMCRV is local optimum, different initializations may result in different performances. It is a simple way to use the random matrix for initialization. In this paper, to improve the effectiveness of IMCRV, we present a novel initialization method. The initializations of $U^{(v)}$, P and W for Eqs. (6) and (7) are obtained by the following two target functions respectively as

$$\begin{aligned}\min_{U^{(v)},P,W} &\sum_{v=1}^{V}\left(\left\|X_E^{(v)}-U^{(v)}PS_E^{(v)}\right\|_F^2+\lambda\left\|U^{(v)}\right\|_F^2\right)\\ &+\beta\mathrm{tr}\left(PL_{\widetilde{W}}P^T\right)+\gamma\|W\|_F^2 \qquad (30)\\ s.t.\quad & U^{(v)}\geq 0,\ P\geq 0,\\ \forall i\quad & W_i^T\mathbf{1}=1, W_i\geq 0,\end{aligned}$$

and

$$\begin{aligned}\min_{U^{(v)},P,W} &\sum_{v=1}^{V}\left(\left\|X_E^{(v)}-U^{(v)}PS_E^{(v)}\right\|_F^2+\lambda\left\|U^{(v)}\right\|_F^2\right)\\ &+\beta\mathrm{tr}\left(PL_{\widetilde{W}}P^T\right)+\gamma\sum_{i=1}^{N}\sum_{j=1}^{N}W_{ij}\ln W_{ij} \qquad (31)\\ s.t.\quad & U^{(v)}\geq 0,\ P\geq 0,\\ \forall i\quad & W_i^T\mathbf{1}=1, W_i\geq 0.\end{aligned}$$

It can be seen from Eqs. (30) and (31) that only the existing data $X_E^{(v)}$ is employed to seek the initializations of $U^{(v)}$, P and W, and the reconstructed view of the missing data $X_M^{(v)}$ is not considered. Obviously, these two targets can be solved by iterative updates. Before the iteration, $U^{(v)}$ and P are initialized by random matrices, and W is initialized as an $N\times N$ matrix with all the elements equaling $1/N$. The iterative procedure is presented as follows.

1) Update $U^{(v)}$ with fixed P and W

Like step 2) of the iterative update in Section 3.2, we can easily obtain the update rule of $U^{(v)}$ using the gradient descent with multiplicative update rule. It is updated as

$$U_{(t+1)ir}^{(v)}=\frac{U_{(t)ir}^{(v)}\left(X_E^{(v)}S_E^{(v)^T}P_{(t)}^T\right)_{ir}}{\left(U_{(t)}^{(v)}P_{(t)}S_E^{(v)}S_E^{(v)^T}P_{(t)}^T+\lambda U_{(t)}^{(v)}\right)_{ir}}. \qquad (32)$$

2) Update P with fixed $U^{(v)}$ and W

Similarly, using the gradient descent with the multiplicative update rule, we can update P as

$$\begin{aligned}&P_{rj}^{(t+1)}\\ &=\frac{P_{rj}^{(t)}\left(\sum_{v=1}^{V}U_{(t+1)}^{(v)^T}X_E^{(v)}S_E^{(v)^T}+\beta P_{(t)}\widetilde{W}_{(t)}\right)_{rj}}{\left(\sum_{v=1}^{V}U_{(t+1)}^{(v)^T}U_{(t+1)}^{(v)}P_{(t)}S_E^{(v)}S_E^{(v)^T}+\beta P_{(t)}\widetilde{D}_{(t)}\right)_{rj}}.\end{aligned} \qquad (33)$$

3) Update W with fixed $U^{(v)}$ and P

It is easy to find that the update of W is the same as step 4) in Section 3.2. For the targets in Eqs. (30) and (31), W can be updated by Eqs. (25) and (28), respectively.

The initialization algorithm is summarized in Algorithm 2. We set the convergence criterion of Algorithm 2 as

$$|(J_2(t+1)-J_2(t))/J_2(t)|<\epsilon, \qquad (34)$$

where $J_2(t)$ is the value of the objective (30) or (31).

Algorithm 2. Initialization of IMCRV

Input:
Incomplete multi-view data $X_E^{(v)}|_{v=1}^V$ from V views; parameters λ, β and γ.

Output:
Initializations of basis matrices $U^{(v)}|_{v=1}^V$, uniform latent representation P and similarity matrix W.

1: Initialize $U^{(v)}|_{v=1}^V$ and P with random matrices and initialize W using an $N\times N$ matrix with all the elements equaling $1/N$.
2: **repeat**
3: With fixed P, update $U^{(v)}|_{v=1}^V$ using Eq. (32).
4: With fixed $U^{(v)}|_{v=1}^V$ and W, update P using Eq. (33).
5: With fixed P, update W using Eqs. (25) or (28).
6: **until** Eqs. (30) or (31) converges.

3.4 Complexity Analysis

The computational complexity of Algorithm 1 is determined by updating $X^{(v)}$, $U^{(v)}$, P and W in Eqs. (9), (14), (19) and (25) (or Eq. (28)). The most time consuming parts are matrix multiplication for Eqs. (9), (14), (19) and (28), and data sorting for Eq. (25). With $k<N$ and $k<d_v|_{v=1}^V$, which usually hold, the computational complexities of Eqs. (9), (14), (19) and (25) (or Eq. (28)) are $\mathcal{O}(\sum_{v=1}^V d_vN^2)$, $\mathcal{O}(\sum_{v=1}^V d_vkN)$, $\mathcal{O}(kN^2+\sum_{v=1}^V d_vkN)$, $\mathcal{O}(N^2logN)$ (or $\mathcal{O}(kN^2)$), respectively. Suppose T_1 is the iteration times of Algorithm 1. In summary, the computational complexity of Algorithm 1 is $\mathcal{O}(T_1(N^2logN+\sum_{v=1}^V d_vN^2))$ (or $\mathcal{O}(T_1\sum_{v=1}^V d_vN^2)$). Analyzing Algorithm 2 similarly and considering $N\leq\sum_{v=1}^V N_E^{(v)}$, we can obtain that, the computational complexity of Algorithm 2 is $\mathcal{O}(T_2(N^2logN+\sum_{v=1}^V d_vN_E^{(v)}N))$ (or $\mathcal{O}(T_2\sum_{v=1}^V d_vN_E^{(v)}N)$), where T_2 is the iteration times of Algorithm 2.

3.5 Convergence

It is easy to know that the objective of Eq. (6) $L_1(X^{(v)}, U^{(v)}, P, W) \geq 0$ and the objective of Eq. (7) $L_2(X^{(v)}, U^{(v)}, P, W) \geq -N \ln N$. If the objectives of Eqs. (6) and (7) are nonincreasing in the iterative process, the IMCRV algorithm will converge to a local minimum. The objectives are obviously nonincreasing with the update of $X^{(v)}$ in Eq. (9). Since the updates of W in Eqs. (25) and (28) are obtained by the Lagrangian multiplier method, the objectives must be nonincreasing. We get the updates of $U^{(v)}$ and P in Eqs. (14) and (19) by gradient descent. We know that the objectives must be nonincreasing if the learning rates are sufficiently small. Now we prove that we can also guarantee the nonincrease of the objectives using the learning rates in Eqs. (13) and (18). First, we give the definition of the auxiliary function, which is used in the NMF algorithm [41].

Definition 1. *$G(Z, Z^{(t)})$ is an auxiliary function of $L(Z)$, if the conditions $G(Z, Z^{(t)}) \geq L(Z)$ and $G(Z, Z) = L(Z)$ hold.*

With the definition of the auxiliary function, we have the following lemmas.

Lemma 1. *If G is an auxiliary function of L, then L is nonincreasing under the update $Z^{(t+1)} = \arg\min_Z G(Z, Z^{(t)})$.*

Proof. It is easy to know that $L(Z^{(t+1)}) \leq G(Z^{(t+1)}, Z^{(t)}) \leq G(Z^{(t)}, Z^{(t)}) = L(Z^{(t)})$. □

Lemma 2. *If A is a matrix with the element*

$$A_{ir} = \frac{\left(U_{(t)}^{(v)} P P^T + \lambda U_{(t)}^{(v)}\right)_{ir}}{U_{(t)ir}^{(v)}}, \tag{35}$$

then

$$\begin{aligned} & G\left(U^{(v)}, U_{(t)}^{(v)}\right) \\ &= L\left(U_{(t)}^{(v)}\right) + \mathrm{tr}\left(\left(U^{(v)} - U_{(t)}^{(v)}\right)^T \nabla L\left(U_{(t)}^{(v)}\right)\right) \\ &\quad + \mathrm{tr}\left(\left(U^{(v)} - U_{(t)}^{(v)}\right)^T \left(A * \left(U^{(v)} - U_{(t)}^{(v)}\right)\right)\right), \end{aligned} \tag{36}$$

is an auxiliary function of $L\left(U^{(v)}\right)$, where $*$ is the Hadamard product.

The proof of Lemma 2 is presented in Appendix A, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TKDE.2021.3112114.

Lemma 3. *If B is a matrix with the element*

$$B_{rj} = \frac{\left(\sum_{v=1}^{V} U^{(v)T} U^{(v)} P_{(t)} + \beta P_{(t)} \widetilde{D}\right)_{rj}}{P_{rj}^{(t)}}, \tag{37}$$

then

$$\begin{aligned} G\left(P, P_{(t)}\right) = L\left(P_{(t)}\right) &+ \mathrm{tr}\left(\left(P - P_{(t)}\right)^T \nabla L\left(P_{(t)}\right)\right) \\ &+ \mathrm{tr}\left(\left(P - P_{(t)}\right)^T \left(B * \left(P - P_{(t)}\right)\right)\right), \end{aligned} \tag{38}$$

is an auxiliary function of $L(P)$.

The proof of Lemma 3 is presented in Appendix B, available in the online supplemental material. With Lemmas 1, 2 and 3, we can obtain the following two theorems.

Theorem 1. *$L(U^{(v)})$ is nonincreasing under the update rule presented in Eq. (14).*

Proof. The derivative of $G(U^{(v)}, U_{(t)}^{(v)})$ with respect to $U^{(v)}$ is

$$\nabla G\left(U^{(v)}, U_{(t)}^{(v)}\right) = \nabla L\left(U_{(t)}^{(v)}\right) + 2A * \left(U^{(v)} - U_{(t)}^{(v)}\right). \tag{39}$$

Let $\nabla G\left(U^{(v)}, U_{(t)}^{(v)}\right) = 0$, we have

$$U^{(v)} = U_{(t)}^{(v)} - \frac{1}{2} \nabla L\left(U_{(t)}^{(v)}\right) \oslash A, \tag{40}$$

where $\oslash$ is the element-wise division. $G(U^{(v)}, U_{(t)}^{(v)})$ is minimized with Eq. (40). According to Lemmas 1 and 2, $L(U^{(v)})$ is nonincreasing when $U^{(v)}$ is updated with Eq. (40). Writing the elements of Eq. (40) explicitly, we have

$$\begin{aligned} U_{(t+1)ir}^{(v)} &= U_{(t)ir}^{(v)} - \frac{U_{(t)ir}^{(v)} \left(U_{(t)}^{(v)} P P^T - X^{(v)} P^T + \lambda U_{(t)}^{(v)}\right)_{ir}}{\left(U_{(t)}^{(v)} P P^T + \lambda U_{(t)}^{(v)}\right)_{ir}} \\ &= \frac{U_{(t)ir}^{(v)} \left(X^{(v)} P^T\right)_{ir}}{\left(U_{(t)}^{(v)} P P^T + \lambda U_{(t)}^{(v)}\right)_{ir}}. \end{aligned} \tag{41}$$

□

Theorem 2. *$L(P)$ is nonincreasing under the update rule presented in Eq. (19).*

Proof. Let the derivative of $G\left(P, P_{(t)}\right)$ with respect to P be zero, we obtain

$$\nabla G\left(P, P_{(t)}\right) = \nabla L\left(P_{(t)}\right) + 2B * \left(P - P_{(t)}\right) = 0. \tag{42}$$

Then, we have

$$P = P_{(t)} - \frac{1}{2} \nabla L\left(P_{(t)}\right) \oslash B. \tag{43}$$

We obtain the minimal $G\left(P, P_{(t)}\right)$ with P presented in Eq. (43). According to Lemmas 1 and 3, $L(P)$ is nonincreasing with the update rule in Eq. (43). Writing the elements of Eq. (43) explicitly, we have

$$\begin{aligned} P_{rj}^{(t+1)} &= P_{rj}^{(t)} - \frac{P_{rj}^{(t)} \left(\begin{array}{c} \sum_{v=1}^{V} \left(U^{(v)T} U^{(v)} P_{(t)} - U^{(v)T} X^{(v)}\right) \\ + \beta P_{(t)} \widetilde{D} - \beta P_{(t)} \widetilde{W} \end{array}\right)_{rj}}{\left(\sum_{v=1}^{V} U^{(v)T} U^{(v)} P_{(t)} + \beta P_{(t)} \widetilde{D}\right)_{rj}} \\ &= \frac{P_{rj}^{(t)} \left(\sum_{v=1}^{V} U^{(v)T} X^{(v)} + \beta P_{(t)} \widetilde{W}\right)_{rj}}{\left(\sum_{v=1}^{V} U^{(v)T} U^{(v)} P_{(t)} + \beta P_{(t)} \widetilde{D}\right)_{rj}}. \end{aligned} \tag{44}$$

□

According to the previous analysis and these two theorems, we can conclude that Algorithm 1 converges to a local minimum. Similar to the proof of the convergence of Algorithm 1, the convergence of Algorithm 2 can also be proved.

3.6 The Relationship to Other Methods

PVC, IMVDG, DAIMC, MVP, CMMVC and our IMCRV are all based on matrix factorization. PVC, IMVDG and MVP mainly aim at processing incomplete two-view data. It will increase the computational difficulty to generalize PVC,

IMVDG and MVP to more views. DAIMC introduces the weight matrix to handle missing views. In PVC, IMVDG, MVP and DAIMC, the missing views are ignored and only the existing views are involved in matrix factorization. CMMVC simply fills the missing views with 0 before matrix factorization, which does not utilize the existing views and ignores the distribution character of different multi-view data.

Considering multi-view data are derived from a common latent subspace, as presented in Eq. (8), IMCRV reconstructs the missing views with basis matrices and the common latent representation, which can strengthen the connection of different views. Then the reconstructed views are also used to seek the latent representation. In this procedure, the examples with incomplete views are exploited sufficiently.

4 Experiments

In this section, we evaluate the effectiveness of IMCRV by clustering experiments. The experiments are conducted on four public multi-view datasets. IMCRV is compared with the following multi-view clustering methods.

BSV. Best Single View. Clustering is conducted for each single view and the one with the best result is employed.

MultiNMF [19]. Nonnegative Matrix Factorization based Multi-view Clustering.

Spec-Pair [20]. Pairwise Co-regularization Multi-view Spectral Clustering. Spec-Pair makes the eigenvectors of different views have pairwise similarity.

Spec-Cent [20]. Centroid based Co-regularization Multi-view Spectral Clustering, Spec-Cent regularizes the eigenvectors of different views towards a common consensus.

PVC [32]. Partial Multi-View Clustering.

IMVDG [33]. Incomplete Multi-Modal Visual Data Grouping.

DAIMC [34]. Doubly Aligned Incomplete Multi-view Clustering algorithm.

SRLC [38]. Simultaneous Representation Learning and Clustering.

AGC-IMC [36]. Adaptive Graph Completion based Incomplete Multi-view Clustering.

IMCNMF. Incomplete Multi-view Clustering based on Nonnegative Matrix Factorization. IMCNMF removes the Laplacian regularization term from the objective of IMCRV. The objective of IMCNMF is defined as

$$\begin{aligned}\min_{X^{(v)},U^{(v)},P}&\sum_{v=1}^{V}\left(\left\|X^{(v)}-U^{(v)}P\right\|_F^2+\lambda_1\left\|U^{(v)}\right\|_F^2\right)\\&+\lambda_2\|P\|_F^2\\ s.t.\quad & X_E^{(v)}=X^{(v)}S_E^{(v)},\ X^{(v)}\geq 0,\ U^{(v)}\geq 0,\ P\geq 0,\end{aligned}\tag{45}$$

In the experiments, MultiNMF, Spec-Pair, Spec-Cent, PVC, IMVDG, DAIMC, AGC-IMC, IMCNMF and our IMCRV are first performed to obtain new features, and then K-means is used to clustering these new features. For BSV, K-means is performed on each view and the one with the best clustering result is reported. For SRLC, the clustering result is obtained by the probability label matrix directly. All of the clustering results are evaluated by the cluster validity index: Normalized Mutual Information (NMI) and Accuracy. In the following, IMCRV with the Frobenius norm regularization and the Shannon entropy regularization are denoted as IMCRV-1 and IMCRV-2, respectively. The code of IMCRV is released at https://github.com/yinjun8429shmtu/yinjun_shmtu.

4.1 Datasets

3Sources. 3Sources dataset contains 948 news articles covering 416 distinct news stories. Each source can be seen as one view of a story. Each story was manually annotated with one or more of the six topical labels. In our experiment, we use 169 stories which were reported in all the three sources. The dimensions of three views are 3068, 3631 and 3560, respectively.

Multiple Feature. Multiple Feature dataset [44] consists of features of handwritten numerals. Each numeral contains 200 examples. These examples are represented in terms of the six feature sets and each feature set can be seen as one view of the examples. In our experiments, we use three feature sets: 76 dimensional fourier coefficients of the character shapes, 47 dimensional zernike moments and 6 dimensional morphological features.

Wikipedia. Wikipedia dataset [45] contains 2866 documents belonging to 10 categories: art, biology, geography, history, literature, media, music, royalty, sport and warfare. These documents are text-image pairs, which form a two-view dataset. The text feature is represented by 10D vector, and the image is represented by 128D bag of words features.

SensIT-Vehicle. SensIT-Vehicle dataset contains three different types of vehicles. Each vehicle is described by acoustic and seismic signal features. The signals are collected by sensors when the vehicle is running across the testing field. In our experiments, 12000 vehicles from three types are employed. For each vehicle, 50D acoustic features and 50D seismic features form its two views.

The datasets are summarized in Table 1.

4.2 Experimental Results and Analysis

We randomly select 10% to 90% examples, with the interval of 20%, as incomplete examples with missing views. The incomplete examples evenly distribute in all views and there is no overlap, i.e., only one view of the incomplete examples is missing. The random selection of incomplete examples is repeated 10 times and the average results are reported. BSV, MultiNMF, Spec-Pair and Spec-Cent cannot be directly applied to the incomplete examples. Before performing these methods, we use the average value of existing data to fill the missing views. For MultiNMF, Spec-Pair, Spec-Cent, PVC, IMVDG, DAIMC, AGC-IMC, IMCNMF and our IMCRV, the dimensions of new features are set as the number of clusters. The parameters of MultiNMF, Spec-Pair, Spec-Cent, PVC, IMVDG, DAIMC, AGC-IMC and SRLC are set as suggested in the original papers. The parameters λ, β and γ of IMCRV are empirically set and the parameter study will be discussed in Section 4.3. As PVC and IMVDG can not process three-view data directly, on 3sources and Multiple Feature datasets, we obtain three two-view datasets with different combinations of the original three-view data first, and then perform PVC and IMVDG using three two-view datasets, respectively, and report the average results. Since the clustering results of

TABLE 1
Dataset Summary

Dataset	Size	Class	View	Dimensionality
3Sources	169	6	3	3068/3631/3560
Multiple Feature	2000	10	3	76/47/6
Wikipedia	2866	10	2	10/128
SensIT-Vehicle	12000	3	2	50/50

K-means vary with different initializations, we run K-means 20 times and adopt the mean results.

Tables 2, 3, 4 and 5 list NMI and Accuracy under different Incomplete Example Ratio (IER) settings on 3Sources, Multiple Feature, Wikipedia and SensIT-Vehicle datasets, respectively. From Table 2, we can see that, on 3Sources dataset, IMCRV-1 or IMCRV-2 performs best, no matter which IER is set. BSV has a relatively good performance under small IER. However, its performance declines quickly along with the increase of IER. Incomplete multi-view clustering methods PVC, IMVDG, SRLC, AGC-IMC, IMCNMF and IMCRV perform better than other methods under big IER. It is probably caused by that the influences of incomplete views are reduced in these methods. It can be seen from Table 3 that, on Multiple Feature dataset, IMCRV-1 or IMCRV-2 performs best under IER=10%, IER=30% and IER=50%. IMCRV-2 obtains the second highest NMI and IMCRV-1 obtains the second highest accuracy under IER=70%. Under IER=90%, IMCRV-1 and IMCRV-2 obtain the second and third highest accuracies respectively, which are only a little lower than the accuracy of SRLC. We can get from Table 4 that, in most cases, IMCRV-1 or IMCRV-2 obtains the best performances on Wikipedia dataset. Under IER=70% and IER=90%, SRLC obtains the highest accuracies and IMCRV-2 obtains the highest NMI. From Table 5, it can be found that, on SensIT-Vehicle dataset, IMCRV-1 or IMCRV-2 outperforms all the other methods, except for accuracy with IER=70%. IMVDG obtains the highest accuracy under IER=70%, and IMCRV-2 has the second highest accuracy.

To evaluate IMCRV further, we test it using multi-view data with more missing views. Experiments are conducted on 3sources dataset. We also randomly select 10% to 90% examples, with the interval of 20%, as incomplete examples with missing views. However, different from the previous experiments, two views of some examples are missing. PVC and IMVDG cannot run in this situation. Table 6 lists NMI and Accuracy of other nine methods under different IER settings. It can be seen from Table 6 that, under all IER settings, IMCRV-1 or IMCRV-2 has the best clustering performance.

On the whole, with missing views, incomplete multi-view clustering methods perform better than other multi-

TABLE 2
NMI/Accuracy Under Different IER Settings on 3Sources Dataset

Method\IER	10%	30%	50%	70%	90%
BSV	0.5562/0.6023	0.5116/0.5763	0.4438/0.5341	0.4034/0.5043	0.3597/0.4697
MultiNMF	0.4614/0.5053	0.4541/0.4977	0.4353/0.5007	0.4106/0.4701	0.3922/0.4751
Spec-Pair	0.4476/0.5161	0.4233/0.4880	0.4041/0.4833	0.3947/0.4722	0.3633/0.4538
Spec-Cent	0.4937/0.5511	0.4772/0.5342	0.4594/0.5272	0.4583/0.5283	0.4172/0.4930
PVC	0.5497/0.5890	0.5660/0.6197	0.5382/0.5986	0.5303/0.6100	0.4928/0.5792
IMVDG	0.5550/0.5950	0.5386/0.5862	0.5440/0.6009	0.5245/0.5901	0.4952/0.5783
DAIMC	0.5307/0.5674	0.4197/0.4983	0.3813/0.4707	0.3490/0.4583	0.3123/0.4369
SRLC	0.5162/0.6201	0.4822/0.6201	0.4688/0.5911	0.4302/0.5615	0.4226/0.5604
AGC-IMC	0.5484/0.6498	0.5194/0.6280	0.5468/0.6426	0.5449/0.6265	0.5378/0.6359
IMCNMF	0.5664/0.6245	0.5434/0.6024	0.5473/0.6032	0.5485/0.5939	0.5303/0.6070
IMCRV-1	**0.6163/0.6638**	**0.6019/0.6518**	**0.6023/0.6475**	**0.5761/0.6487**	0.5439/0.6228
IMCRV-2	0.5967/0.6408	0.5869/0.6378	0.6015/0.6370	0.5590/0.6038	**0.5751/0.6564**

TABLE 3
NMI/Accuracy Under Different IER Settings on Multiple Feature Dataset

Method\IER	10%	30%	50%	70%	90%
BSV	0.5675/0.5796	0.5142/0.5332	0.4691/0.4919	0.4286/0.4416	0.3802/0.3874
MultiNMF	0.4979/0.5117	0.4774/0.5138	0.4311/0.4570	0.3993/0.4278	0.3663/0.3870
Spec-Pair	0.5890/0.5985	0.5407/0.5619	0.5003/0.5220	0.4728/0.4941	0.4420/0.4640
Spec-Cent	0.5847/0.5895	0.5373/0.5554	0.5039/0.5279	0.4674/0.4818	0.4392/0.4587
PVC	0.5869/0.5685	0.5590/0.5485	0.5196/0.5276	0.4878/0.5063	0.4542/0.4755
IMVDG	0.5379/0.5676	0.4882/0.5432	0.4711/0.5177	0.4289/0.4765	0.3885/0.4385
DAIMC	0.4773/0.5373	0.3952/0.4691	0.3097/0.3753	0.2837/0.3410	0.2909/0.3356
SRLC	0.5595/0.5677	0.5316/0.5581	0.5162/0.5411	0.5040/**0.5443**	0.4782/**0.5096**
AGC-IMC	0.5430/0.4348	0.5406/0.4387	0.5325/0.4319	**0.5199**/0.4308	**0.5087**/0.4226
IMCNMF	0.5958/0.5846	0.5564/0.5530	0.4835/0.4843	0.4522/0.4959	0.4117/0.4227
IMCRV-1	0.6222/0.5918	**0.5732/0.5739**	**0.5621/0.5844**	0.5054/0.5341	0.4411/0.4924
IMCRV-2	**0.6223/0.5992**	0.5726/0.5719	0.5580/0.5826	0.5072/0.5330	0.4421/0.4937

TABLE 4
NMI/Accuracy Under Different IER Settings on Wikipedia Dataset

Method\IER	10%	30%	50%	70%	90%
BSV	**0.5201**/0.5536	0.4502/0.5076	0.3865/0.4530	0.3335/0.3927	0.2731/0.3367
MultiNMF	0.3621/0.4773	0.3262/0.4398	0.2845/0.4122	0.2502/0.3883	0.2042/0.3473
Spec-Pair	0.3926/0.4941	0.3534/0.4644	0.3064/0.4260	0.2427/0.3531	0.1652/0.2767
Spec-Cent	0.4435/0.5121	0.3837/0.4735	0.3381/0.4350	0.3022/0.4038	0.2690/0.3661
PVC	0.5069/0.5593	0.4089/0.4692	0.3411/0.4155	0.2773/0.3684	0.2276/0.3321
IMVDG	0.4993/0.5299	0.4313/0.4939	0.3683/0.4547	0.3125/0.4139	0.2555/0.3638
DAIMC	0.4588/0.5495	0.3176/0.4536	0.1821/0.3339	0.0808/0.2310	0.0254/0.1669
SRLC	0.3630/0.4813	0.3472/0.4752	0.3021/0.4447	0.2723/**0.4152**	0.2319/**0.3775**
AGC-IMC	0.4798/0.5046	0.4254/0.4560	0.3577/0.4044	0.3102/0.3722	0.2811/0.3296
IMCNMF	0.5115/0.5422	0.4574/0.4915	0.3995/0.4296	0.3337/0.3709	0.2438/0.2882
IMCRV-1	0.5158/**0.5623**	**0.4626/0.5174**	0.4044/**0.4597**	0.3336/0.4027	0.2996/0.3140
IMCRV-2	0.5042/0.5388	0.4620/0.5033	**0.4075**/0.4547	**0.3497**/0.4046	**0.3177**/0.3492

TABLE 5
NMI/Accuracy Under Different IER Settings on SensIT-Vehicle Dataset

Method\IER	10%	30%	50%	70%	90%
BSV	0.1469/0.5407	0.1184/0.5075	0.0937/0.4760	0.0696/0.4448	0.0566/0.4235
MultiNMF	0.2904/0.6897	0.2585/0.6681	0.2316/0.6402	0.2104/0.6138	0.1793/0.5385
Spec-Pair	0.2196/0.6354	0.2100/0.6195	0.2014/0.6039	0.1649/0.5632	0.0709/0.4633
Spec-Cent	0.2815/0.6862	0.2588/0.6696	0.2336/0.6506	0.1982/0.6252	0.1539/0.5889
PVC	0.3156/0.6786	0.2899/0.6788	0.2374/0.6239	0.2043/0.5915	0.1945/0.6060
IMVDG	0.3254/0.6940	0.2900/0.6776	0.2327/0.6236	0.2214/**0.6335**	0.1968/0.6082
DAIMC	0.1971/0.6241	0.1264/0.5310	0.0828/0.4654	0.0503/0.4188	0.0265/0.3822
SRLC	0.2385/0.6284	0.2265/0.6254	0.1960/0.5867	0.1540/0.5384	0.1159/0.5010
AGC-IMC	0.2784/0.6287	0.2677/0.6516	0.2396/0.6218	0.2123/0.5884	0.1692/0.5597
IMCNMF	0.3138/0.6726	0.2845/0.6519	0.2537/0.6219	0.1983/0.5989	0.1616/0.5606
IMCRV-1	**0.3298/0.7057**	**0.2919/0.6846**	**0.2579/0.6627**	0.2195/0.6297	**0.2007/0.6199**
IMCRV-2	0.3278/0.7024	0.2897/0.6803	0.2540/0.6551	**0.2251**/0.6325	0.1910/0.6175

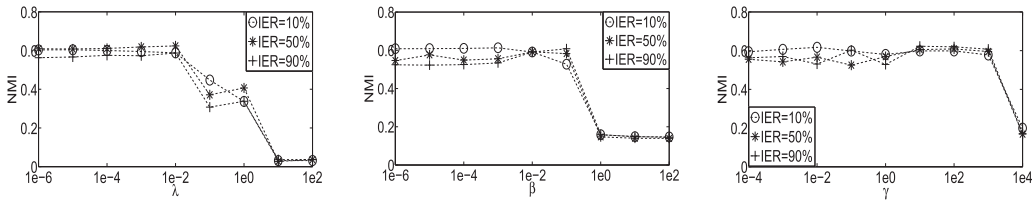
TABLE 6
NMI/Accuracy With More Missing Views on 3Sources Dataset

Method\IER	10%	30%	50%	70%	90%
BSV	0.5373/0.5845	0.4496/0.5395	0.3891/0.4887	0.3190/0.4427	0.2511/0.4030
MultiNMF	0.4529/0.5045	0.4275/0.4873	0.4050/0.4820	0.3247/0.4170	0.3088/0.4252
Spec-Pair	0.4493/0.5233	0.4170/0.5077	0.3910/0.4654	0.3208/0.4362	0.2793/0.4095
Spec-Cent	0.4778/0.5434	0.4619/0.5236	0.4178/0.4896	0.3552/0.4459	0.3028/0.4129
DAIMC	0.5205/0.5611	0.4213/0.5016	0.3254/0.4350	0.2851/0.4236	0.1966/0.3762
SRLC	0.4766/0.5953	0.4770/0.6195	0.4190/0.5615	0.3631/0.5030	0.3033/0.4391
AGC-IMC	0.5477/0.6545	0.5418/0.6306	0.4806/0.5956	0.4786/0.5910	0.4219/0.5146
IMCNMF	0.5831/0.6088	0.5715/0.6342	0.5585/0.6366	0.5263/0.6229	0.4238/0.5444
IMCRV-1	**0.6242/0.6742**	0.5810/0.6575	**0.5944/0.6683**	0.5433/0.6229	0.4849/**0.6124**
IMCRV-2	0.5891/0.6302	**0.6038/0.6677**	0.5791/0.6346	**0.5471/0.6291**	**0.4938**/0.5999

view clustering methods, especially for big IER. With the increase of IER, the performance decline almost appears in all the methods. In most cases, IMCRV obtains the best performances. To evaluate the effect of graph Laplacian, in the experiments, IMCRV is compared with IMCNMF, which ignores the graph Laplacian term and only employs NMF to seek the latent representation. In all the experiments, IMCRV outperforms IMCNMF. It indicates that graph Laplacian is useful to improve the clustering performance.

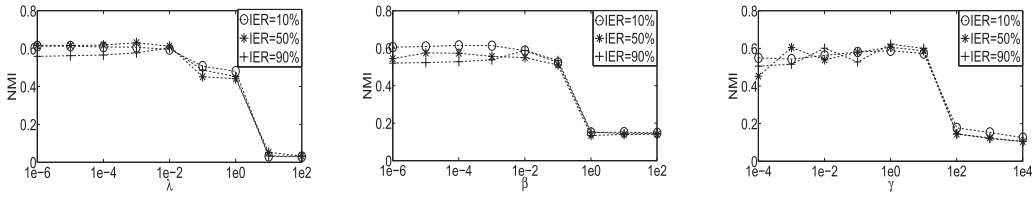
4.3 Parameter Study

There are three parameters λ, β and γ in IMCRV. The experiments are conducted on 3Sources dataset with IER=10%, IER=50% and IER=90%. We fix two parameters and traverse the third parameter. λ and β are selected from the set $\{10^{-6}, 10^{-5}, 10^{-4}, \ldots, 10, 100\}$. γ is selected from the set $\{10^{-4}, 10^{-3}, 10^{-2}, \ldots, 10^{3}, 10^{4}\}$. Figs. 1 and 2 show NMI of IMCRV-1 and IMCRV-2 versus the variations of three parameters, respectively. From these two figures, it can be found that IMCRV-1 and IMCRV-2 have relatively good performances with $\lambda = 1e-2$, which is consistent with the conclusion in [32]. IMCRV-1 performs well when $\beta = \{1e-3, 1e-2, 1e-1\}$ and $\gamma = \{1e1, 1e2\}$. IMCRV-2 performs well when $\beta = \{1e-4, 1e-3, 1e-2\}$ and $\gamma = \{1e0, 1e1\}$. From Figs. 1 and 2, we can also see that, for IMCRV-1 and IMCRV-2, a relatively big β is needed to obtain the best performance when IER is big. Taking IMCRV-1 as an example, we can find that, it has the best



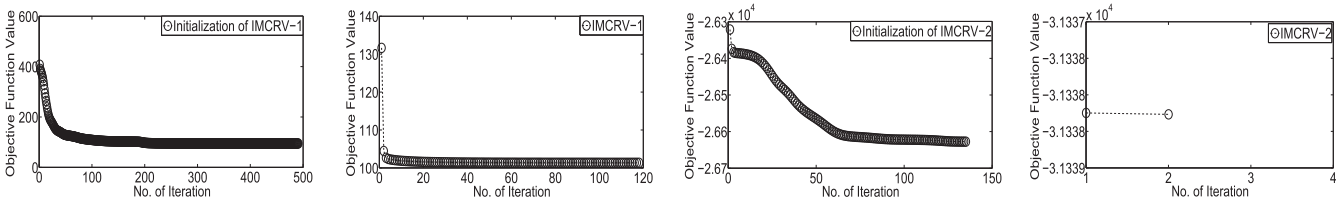
(a) λ (b) β (c) γ

Fig. 1. NMI of IMCRV-1 versus the variations of parameters λ, β and γ.



(a) λ (b) β (c) γ

Fig. 2. NMI of IMCRV-2 versus the variations of parameters λ, β and γ.



(a) Initialization of IMCRV-1 (b) IMCRV-1 (c) Initialization of IMCRV-2 (d) IMCRV-2

Fig. 3. The values of the objective functions versus iterations on Wikipedia dataset.

performance with $\beta = 1e-3$ under IER=10% and with $\beta = 1e-1$ under IER=90%. When IER increases, the reliability of the NMF term in the objective function of IMCRV decreases. Correspondingly, the importance of the graph Laplacian term increases. Therefore, we should set a relative big β under big IER.

4.4 Convergence Study

The theoretical analysis of convergence of IMCRV has been presented in Section 4. Now we evaluate the convergence by experiments. Fig. 3 shows the convergence curves of IMCRV-1, IMCRV-2 and the initialization algorithms of them on Wikipedia datasets. Here, the incomplete multi-view data with IER=30% are used. It can be seen from the figure that the values of the objective functions of all the algorithms monotonically decrease along with the increase of iteration. As a whole, IMCRV-2 converges faster than IMCRV-1. With the basis matrices $U^{(v)}|_{v=1}^{V}$ and the latent representation P obtained by their initialization algorithms, IMCRV-1 and IMCRV-2 can converge very fast. Especially for IMCRV-2, it converges after only two iterations.

5 Conclusion

In this paper, we proposed a new incomplete multi-view clustering method IMCRV. IMCRV seeks the common latent representation of all the views through matrix factorization. Different from previous methods based on matrix factorization, in IMCRV, the missing views are not ignored. These views are reconstructed and the reconstructed views are also used to seek the latent representation. In order to improve the clustering performance, IMCRV also employs Laplacian regularization to bridge the latent representations of all the examples. The objective of IMCRV is solved by an iterative algorithm. Gradient descent with the multiplicative update rule is adopted in the iteration and the convergence of the iteration algorithm is proved. Experiments on public multi-view datasets show that IMCRV outperforms other state-of-the-art incomplete multi-view clustering methods.

Acknowledgments

We would like to acknowledge support from National Natural Science Foundation of China (Grants No. 61603243 and No. 62076096), China Postdoctoral Science Foundation (Grant No. 2017M611503), Shanghai Municipal Project (Grant No. 20511100900), Open Research Fund of KLATASDS-MOE, and Fundamental Research Funds for the Central Universities.

References

[1] C. Xu, D. Tao, and C. Xu, "A survey on multi-view learning," *arXiv*, pp. 1–59, Apr. 20, 2013. [Online]. Available: https://arxiv.org/pdf/1304.5634.pdf

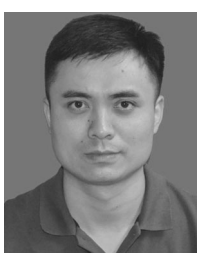
[2] S. Sun, J. Shawe-Taylor , and L. Mao, "PAC-Bayes analysis of multi-view learning," *Inf. Fusion*, vol. 35, pp. 117–131, 2017.

[3] 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, 2017.

[4] S. Li, H. Liu, Z. Tao, and Y. Fu, "Multi-view graph learning with adaptive label propagation," in *Proc. IEEE Int. Conf. Big Data*, 2017, pp. 110–115.

[5] J. Rupnik and J. Shawe-Taylor , "Multi-view canonical correlation analysis," in *Proc. Conf. Data Mining Data Warehouses*, 2010, pp. 1–4.

[6] Y. Luo, D. Tao, K. Ramamohanarao, C. Xu, and Y. Wen, "Tensor canonical correlation analysis for multi-view dimension reduction," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 11, pp. 3111–3124, Nov. 2015.
[7] J. Yin and S. Sun, "Multiview uncorrelated locality preserving projection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 9, pp. 3442–3455, Sep. 2020.
[8] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 188–194, Jan. 2016.
[9] S. Sun, X. Xie, and M. Yang, "Multiview uncorrelated discriminant analysis," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 3272–3284, Dec. 2016.
[10] P. Hu, D. Peng, Y. Sang, and Y. Xiang, "Multi-view linear discriminant analysis network," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5352–5365, Nov. 2019.
[11] J. Xu, J. Han, F. Nie, and X. Li, "Multi-view scaling support vector machines for classification and feature selection," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 7, pp. 1419–1430, Jul. 2020.
[12] X. Xie and S. Sun, "Multi-view support vector machines with the consensus and complementarity information," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 12, pp. 2401–2413, Dec. 2020.
[13] S. Sun, X. Xie, and C. Dong, "Multiview learning with generalized eigenvalue proximal support vector machines," *IEEE Trans. Cybern.*, vol. 49, no. 2, pp. 688–697, Feb. 2019.
[14] C. Zhang *et al.*, "Generalized latent multi-view subspace clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 1, pp. 86–99, Jan. 2020.
[15] 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.
[16] G. Chao, S. Sun, and J. Bi, "A survey on multi-view clustering," *IEEE Trans. Artif. Intell.*, vol. 2, no. 2, pp. 146–168, Apr. 2021.
[17] S. Bickel and T. Scheffer, "Multi-view clustering," in *Proc. IEEE Int. Conf. Data Mining*, 2004, pp. 19–26.
[18] K. Chaudhuri, S. M. Kakade, K. Livescu, and K. Sridharan, "Multi-view clustering via canonical correlation analysis," in *Proc. Int. Conf. Mach. Learn.*, 2009, pp. 129–136.
[19] J. Liu, C. Wang, J. Gao, and J. Han, "Multi-view clustering via joint nonnegative matrix factorization," in *Proc. SIAM Int. Conf. Data Mining*, 2013, pp. 252–260.
[20] A. Kumar, P. Rai, and H. Daume, "Co-regularized multi-view spectral clustering," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2011, pp. 1413–1421.
[21] F. Nie, J. Li, and X. Li, "Self-weighted multiview clustering with multiple graphs," in *Proc. Int. Joint Conf. Artif. Intell.*, 2017, pp. 2564–2570.
[22] Z. Kang *et al.*, , "Multi-graph fusion for multi-view spectral clustering," *Knowl.-Based Syst.*, vol. 189, 2020, Art. no. 105102.
[23] Z. Tao, H. Liu, S. Li, Z. Ding, and Y. Fu, "Marginalized multiview ensemble clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 2, pp. 600–611, Feb. 2020.
[24] J. Wen, Z. Zhang, Y. Xu, and Z. Zhong, "Incomplete multi-view clustering via graph regularized matrix factorization," in *Proc. Eur. Conf. Comput. Vis. Workshops*, 2018, pp. 1–16.
[25] G. Chao *et al.*, "Multi-view cluster analysis with incomplete data to understand treatment effects," *Inf. Sci.*, vol. 494, pp. 278–293, 2019.
[26] X. Liu *et al.*, , "Efficient and effective incomplete multi-view clustering," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 4392–4399.
[27] W. Shao, L. He, C. Lu, and P. S. Yu, "Online multi-view clustering with incomplete views," in *Proc. IEEE Int. Conf. Big Data*, 2016, pp. 1012–1017.
[28] H. Yu, J. Xiong, and X. Zhang, "Multi-view clustering by exploring complex mapping relationship between views," *Pattern Recognit. Lett.*, vol. 138, pp. 230–236, 2020.
[29] A. Trivedi, P. Rai, H. Daumé III, and S. L. DuVall , "Multiview clustering with incomplete views," in *Proc. Int. Conf. Neural Inf. Process. Syst. Workshop*, 2010, pp. 1–7.
[30] W. Shao, X. Shi, and P. S. Yu, "Clustering on multiple incomplete datasets via collective kernel learning," in *Proc. IEEE Int. Conf. Data Mining*, 2013, pp. 1181–1186.
[31] X. Liu, M. Li, L. Wang, Y. Dou, J. Yin, and E. Zhu, "Multiple kernel k-means with incomplete kernels," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 2259–2265.
[32] S.-Y. Li, Y. Jiang, and Z.-H. Zhou, "Partial multi-view clustering," in *Proc. AAAI Conf. Artif. Intell.*, 2014, pp. 1968–1974.
[33] H. Zhao, H. Liu, and Y. Fu, "Incomplete multi-modal visual data grouping," in *Proc. Int. Joint Conf. Artif. Intell.*, 2016, pp. 2392–2398.
[34] M. Hu and S. Chen, "Doubly aligned incomplete multi-view clustering," in *Proc. Int. Joint Conf. Artif. Intell.*, 2018, pp. 2262–2268.
[35] L. Zong, F. Miao, X. Zhang, X. Liu, and H. Yu, "Incomplete multi-view clustering with partially mapped instances and clusters," *Knowl.-Based Syst.*, vol. 212, 2021, Art. no. 106615.
[36] J. Wen *et al.*, "Adaptive graph completion based incomplete multi-view clustering," *IEEE Trans. Multimedia*, vol. 23, pp. 2493–2504, Jul. 2021.
[37] H. Tao, C. Hou, D. Yi, J. Zhu, and D. Hu, "Joint embedding learning and low-rank approximation: A framework for incomplete multiview learning," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1690–1703, Mar. 2021.
[38] W. Zhuge, C. Hou, X. Liu, H. Tao, and D. Yi, "Simultaneous representation learning and clustering for incomplete multi-view data," in *Proc. Int. Joint Conf. Artif. Intell.*, 2019, pp. 4482–4488.
[39] W. Zhuge, H. Tao, T. Luo, L.-L. Zeng, C. Hou, and D. Yi, "Joint representation learning and clustering: A framework for grouping partial multiview data," *IEEE Trans. Knowl. Data Eng.*, early access, Oct. 02, pp. 1–14, 2020, doi: 10.1109/TKDE.2020.3028422.
[40] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2011, pp. 612–620.
[41] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2001, pp. 556–562.
[42] 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.
[43] L. Zhang, L. Qiao, and S. Chen, "Graph-optimized locality preserving projections," *Pattern Recognit.*, vol. 43, no. 6, pp. 1993–2002, 2010.
[44] M. van Breukelen and R. P. W. Duin, "Neural network initialization by combined classifiers," in *Proc. Int. Conf. Pattern Recognit.*, 1998, pp. 215–218.
[45] J. C. Pereira *et al.*, "On the role of correlation and abstraction in cross-modal multimedia retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 521–535, Mar. 2014.



Jun Yin received the BS degree in mathematics and the PhD degree in pattern recognition and intelligence system from the Nanjing University of Science and Technology, Nanjing, China, in 2006 and 2011, respectively. He is an associate professor with the College of Information Engineering, Shanghai Maritime University, and a postdoctoral fellow with the School of Computer Science and Technology, East China Normal University. From 2019 to 2020, he is a visiting scholar with the Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania. His research interests include multi-view learning, dimension reduction, manifold learning, etc. He is on the editorial board of the journal of *Neural Processing Letters*.

Shiliang Sun received the PhD degree in pattern recognition and intelligent systems from Tsinghua University, Beijing, China, in 2007. He is a professor with the School of Computer Science and Technology and the head of the Pattern Recognition and Machine Learning Research Group, East China Normal University, Shanghai, China. From 2009 to 2010, he was a visiting researcher with the Department of Computer Science, Centre for Computational Statistics and Machine Learning, University College London, London, U.K. In 2014, he was a visiting researcher with the Department of Electrical Engineering, Columbia University, New York, NY, USA. His current research interests include kernel methods, multi-view learning, learning theory, approximate inference, sequential modeling, deep learning, and their applications. His research results have expounded in more than 100 publications at peer-reviewed journals and conferences. He is on the editorial board of multiple international journals, including *Neurocomputing* and the *IEEE Transactions on Neural Networks and Learning Systems*.