CCIM-SLR: Incomplete Multiview Co-Clustering by Sparse Low-Rank Representation

Zhenjiao Liu[1,2], Zhikui Chen[1*], Kai Lou[1], Praboda Rajapaksha[2], Liang Zhao[1], Noel Crespi[2], Xiaodi Huang[3]

[1]The School of Software Technology, Dalian University of Technology, Dalian, 116620, China.
[2]Samovar,Telecom SudParis, Institut Polytechnique de Paris, Palaiseau, 91120, France.
[3]School of Computing, Mathematics, and Engineering, Charles Sturt University, Albury, NSW 2640,Australia.

*Corresponding author(s). E-mail(s): zkchen@dlut.edu.cn;
Contributing authors: liuzhenjiao@mail.dlut.edu.cn;
loukai@mail.dlut.edu.cn; praboda_rajapaksha@telecom-sudparis.eu;
liangzhao@dlut.edu.cn; noel.crespi@mines-telecom.fr;
xhuang@csu.edu.au;

Abstract

Clustering incomplete multiview data in real-world applications has become a topic of recent interest. However, producing clustering results from multiview data with missing views and different degrees of missing data points is a challenging task. To address this issue, we propose a co-clustering method for incomplete multiview data by sparse low-rank representation (CCIM-SLR). The proposed method integrates the global and local structures of incomplete multiview data and effectively captures the correlations between samples in a view, as well as between different views by using sparse low-rank learning. CCIM-SLR can alternate between learning the shared hidden view, visible view, and cluster partitions within a co-learning framework. An iterative algorithm with guaranteed convergence is used to optimize the proposed objective function. Compared with other baseline models, CCIM-SLR achieved the best performance in the comprehensive experiments on the five benchmark datasets, particularly on those with varying degrees of incompleteness.

Keywords: Incomplete multiview, Co-clustering, Sparse low-rank representation, Shared hidden view

1 Introduction

In recent years, the availability of multiview data has increased significantly. However, the success of the current multiview learning approaches heavily relies on complete and consistent data from various sources [1–6]. Unfortunately, in real-world scenarios, multi-view data often suffer from incompleteness due to various factors, including missing views and data points. Moreover, the degree of such data incompleteness varies depending on the context. For instance, during land multi-view data acquisition, sensor acquisition failure can easily lead to incomplete optical and SAR image view data. In the medical field, collecting multi-view medical data with missing records of patients with different degrees is often necessary due to privacy concerns. To tackle these issues, incomplete multiview clustering (IMVC) fuses information between multiple views of complex missing data. However, IMVC is challenging in real-world scenarios. Many approaches have been proposed to address IMVC to date, which can be grouped into shallow and deep multiview clustering models.

Shallow IMVC models are limited to low-level features, which can result in samples representing the same object being restricted to the same potential representation in the potential subspace [7–9]. In contrast, Wen et al.[10] introduced a structured deep incomplete multi-view clustering network that combines the spatial relationships between data points with a deep model. Moreover, this method can directly obtain the optimal clustering indicator matrix in one stage.

While both shallow and deep approaches to IMVC have been successful in real-world applications, there are still several limitations that need to be addressed. First, many existing approaches fail to consider both global and local incomplete multi-view information. Second, although data recovery and cluster partitions are essential for IMVC, many existing methods do not combine them effectively. Third, existing deep learning-based IMVC methods have issues with training stability and clustering performance. Finally, some existing approaches to IMVC do not demonstrate their robustness on datasets with missing views and varying degrees of missing data points.

To address the limitations discussed above, we propose a novel framework called Incomplete Multiview Co-Clustering by Sparse Low-Rank Representation (CCIM-SLR). As illustrated in Fig. 1, sample 2 lacks the video modality, while sample 3 lacks the text modality. Using the low-rank sparse representation imputation method we introduced, we effectively restored the missing video modality for sample 2 and performed text modality imputation for sample 3. Specifically, our approach learns a low-rank sparse representation matrix for each view, which is then used to fill in missing samples within each view. This process is achieved through the use of association information between missing and observed samples within a particular view, as well as by considering the association relationships between samples in different views through learning common subspace representations. Furthermore, by introducing the filled view data and the learned implicit view data in a clustering process, the data recovery process and the clustering process can complement each other and lead to improved clustering results. To ensure even more accurate data recovery, we use the adjustable low-rank approximation representation model Γ-norm, which replaces the traditional kernel norm that can only produce low-rank feature representations under certain conditions.

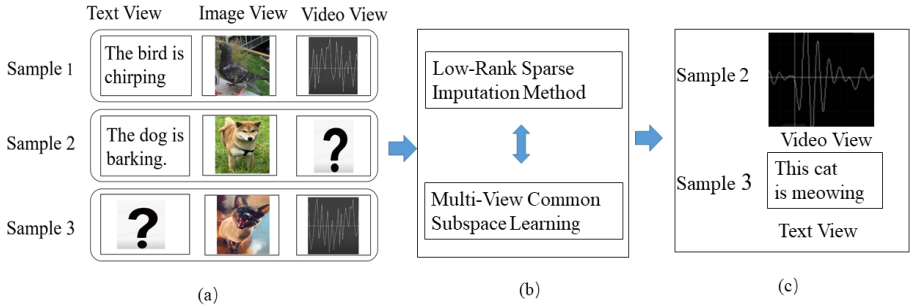


Fig. 1 Specific examples of case problems and proposed solutions. (a) An example of missing data; (b) Imputation mechanism; (c) Data recovery.

In summary, the contributions of this paper are as follows.

- This paper presents CCIM-SLR, a novel approach to incomplete multi-view clustering that leverages a low-rank sparse representation matrix to recover data from missing samples. CCIM-SLR utilizes association information between missing samples and observed samples within views, as well as their association information between views.
- To improve the stability of clustering results for multi-view data with different missing rates, CCIM-SLR uses the Γ-norm model, which is an adjustable low-rank representation method. Γ-norm shows the accuracy of achieving a low-rank representation and the stability of data recovery.
- CCIM-SLR learns both a visible view and a hidden view within a co-learning framework in an end-to-end manner, using a mutual interplay between the view data recovery and a clustering process. This approach avoids the need for post-processing steps such as k-means for final clustering assignment results.
- CCIM-SLR has been validated through both theoretical proofs and experiments. Based on the experimental results, CCIM-SLR outperformed state-of-the-art approaches on the five incomplete multiview datasets. The robustness of CCIM-SLR has been demonstrated through experiments on incomplete multi-view datasets with different missing rates of data points.

The remainder of this paper is organized as follows: Section 2 provides a review of related work, while Section 3 outlines the proposed method and methodologies used in this work. Section 4 reports the experimental results of our CCIM-SLR, together with comparisons of other methods. This paper concludes in Section 5.

2 Related Work and Background

In this section, we review related work and describe some background on IMVC.

2.1 Related work

As mentioned before, the current approaches to IMVC can be grouped into shallow and deep ones. As a shallow approach, Guo J et al. [11] proposed a simple and easy-to-implement method that can reconstruct samples and intersample relations through anchors and fully integrates intraview and interview similarities. In the literature, a weighted semi-nonnegative matrix factorization-based method was proposed to reduce the influence of view incompleteness in clustering [12]. Wen et al. [13] developed a new graph regularization matrix decomposition model to consider the local geometric information and the unbalanced resolution of incomplete multiview observations. Wen et al. [14] proposed introducing a local retention reconstruction term to infer missing views so that all views can be naturally aligned and adding an adaptive weighting strategy for capturing the importance of different views. In [15], feature space-based missing-view inference and manifold space-based similarity graph learning were proposed to better explore the potential information of missing views. Liang et al. [16] developed a reproduced representation; on this basis, a set of incomplete graphs was used to make full use of the geometric structure of the data. Yin et al. [17] introduced a cosine similarity metric to further enhance the preservation of the flow structure of the original multiview, called incomplete multiview clustering with cosine similarity (IMCCS).

As for deep approaches to IMVC, Xu et al. [18] designed an adversarial incomplete multiview clustering (AIMC) method that captures the overall structure and obtains a deeper semantic understanding by seeking the common potential space of multi-view data and inferring incomplete data at the same time. Xu et al. [19] proposed to establish a new multi-view clustering complementarity mechanism that can obtain supplementary information and be regarded as supervisory information with high confidence. Therefore, this method achieves the consistency information of multi-view clustering. To solve the incomplete multiview problem by explicitly generating the data of missing views, Wang et al. [20] applied adaptive fusion and a cycle consistency generation model for incomplete multiview clustering. The deep IMVC model can take into account high representation ability and save time and space. For example, Zheng et al. [21] proposed a method that harnesses the complementary information concealed within view-specific partial graphs obtained from incomplete views. Furthermore, a rank constraint is applied to the Laplacian matrix of the fusion graph to enhance the recovery of the optimal clustering structure in the original data. A generative adversarial network-based model was proposed by Wang et al. [22], which can effectively generate incomplete view data and capture better common structures in IMVC.

Except for the shallow and deep approaches to IMVC, the incomplete multiview clustering algorithm with low-rank sparsity [23, 24] and the multiview algorithm with one-step clustering [25–27] have shown good advantages in the field of the multiview study. However, these approaches can further be improved by overcoming some limitations as mentioned before. Before presenting our proposed approach of CCIM-SLR, we need to provide two background works: 1) sparse low-rank representation through multiview subspace (SRRS) learning; and 2) multiview clustering with the cooperation of visible and hidden views.

2.2 Multiview subspace learning for sparse low-rank representation

SRRS has a significant effect on incomplete multiview data recovery [23]. For a dataset of incomplete multiview $\{\mathbf{X}_v\}_{v=1}^{s}$ with s views, SRRS imputes missing values of all views by the following expression:

$$\begin{aligned} &\min_{\{R^{(v)},E^{(v)}\},H} \sum_{v=1}^{s} \mathcal{B}(R^{(v)}, E^{(v)}) \\ &s.t. \forall v, \mathcal{P}_v(\widehat{R}^{(v)} {X_o}^{(v)}) = H + E^{(v)}, H^\top H = I \end{aligned} \tag{1}$$

where $R^{(v)} \in \mathbb{R}^{\bar{m}^{(v)} \times m^{(v)}}$ is the sparse low-rank representation matrix, $\bar{m}^{(v)}$ is the count of missing samples, and $m^{(v)}$ is the observed samples. $E^{(v)} \in \mathbb{R}^{m \times t}$ is the noise matrix, m is the number of all samples ($m{=}\bar{m}^{(v)}{+}m^{(v)}$), t is the unified dimension of the subspace, $X_o^{(v)}$ is the matrix constructed from the observed samples, and $\widehat{R}^{(v)} \in \mathbb{R}^{m \times m^{(v)}}$ is composed of the matrix $R^{(v)}$ and the identity matrix constructed from the indices of the observed samples. $\mathcal{P}_v(\widehat{R}^{(v)} X_o^{(v)})$: $\mathbb{R}^{m \times d^{(v)}} \to \mathbb{R}^{m \times t}$ represents an operator that projects the samples of all views into the corresponding subspace. SRRS learns that the common representation of all views is $H \in \mathbb{R}^{m \times t}$, and adding constraints to H can effectively avoid trivial solutions. $\mathcal{B}(R^{(v)}, E^{(v)})$ is defined as follows:

$$\mathcal{B}(R^{(v)}, E^{(v)}) = \lambda_1 \|R^{(v)}\|_1 + \lambda_2 \|R^{(v)}\|_* + \lambda_3 \|E^{(v)}\|_1 \tag{2}$$

where λ_1, λ_2 and λ_3 are the parameters and $\|R^{(v)}\|_1$ and $\|R^{(v)}\|_*$ represent the sparse and low-rank constraints on $R^{(v)}$, respectively. To reduce the influence of the noise matrix and make the subspace representations robust, we add a sparse constraint to E. SRRS is a technique that can impute missing values by taking into account both intraview and inter-view relations.

2.3 Multiview clustering with the cooperation of visible and hidden views

Another important work is the multiview clustering model, called multiview clustering with the cooperation of visible and hidden views (MV-Co-VH) [25]. MV-Co-VH is a clustering method that integrates and optimizes both visible and hidden views:

$$\begin{aligned} &\min J(U, Z, \widetilde{Z}, w) = \lambda \sum_{i=1}^{c} \sum_{j=1}^{m} u_{ij} \|h_j - \widetilde{z}_j\|^2 \\ &+ (1-\lambda) \sum_{v=1}^{s} w_{(v)} \sum_{i=1}^{c} \sum_{j=1}^{m} u_{ij} \|{x_j}^{(v)} - {z_j}^{(v)}\|^2 + \eta \sum_{v=1}^{s} w_{(v)} \ln w_{(v)} \\ &s.t. \sum_{i=1}^{c} u_{ij} = 1, u_{ij} \in (0,1), 1 \leqslant j \leqslant m \\ &\sum_{v=1}^{s} w_{(v)} = 1, 0 \leqslant w_{(v)} \leqslant 1, H \geq 0 \end{aligned} \tag{3}$$

where λ is a parameter, matrix $U \in \mathbb{R}^{c\times m}$ is the cluster indicator matrix, c denotes the number of categories, and m represents the number of samples. If u_{ij}=1, sample j belongs to cluster i; otherwise, u_{ij} =0. Z=$\{Z^{(1)}, Z^{(2)}, Z^{(3)}.....Z^{(v)}\}$ is the cluster center matrix for each view. w=$[\omega_1, \omega_2, ...\omega_s]$ contains the weight of each view. H $\in \mathbb{R}^{m\times t}$ denotes the shared hidden view of all views. t is the sample dimension of the hidden view. $\widetilde{Z}$ represents the corresponding clustering center matrix from the hidden view. From Equation (3), MV-Co-VH is a method for extracting hidden views from multiview data through nonnegative matrix factorization. It is also a multiview clustering framework that combines explicit and implicit views to obtain clustering results in one step.

3 The Proposed CCIM-SLR

In this section, we describe our CCIM-SLR method, which learns recovery data from the global and local structures of their original data with incomplete multiview. Apart from that, a one-step clustering strategy is also adopted to produce clustering results that combine shared hidden space and visible view effectively. CCIM-SLR comprises two parts: 1) shared hidden subspace learning based on SRRS; and 2) incomplete multiview co-clustering by SRRS.

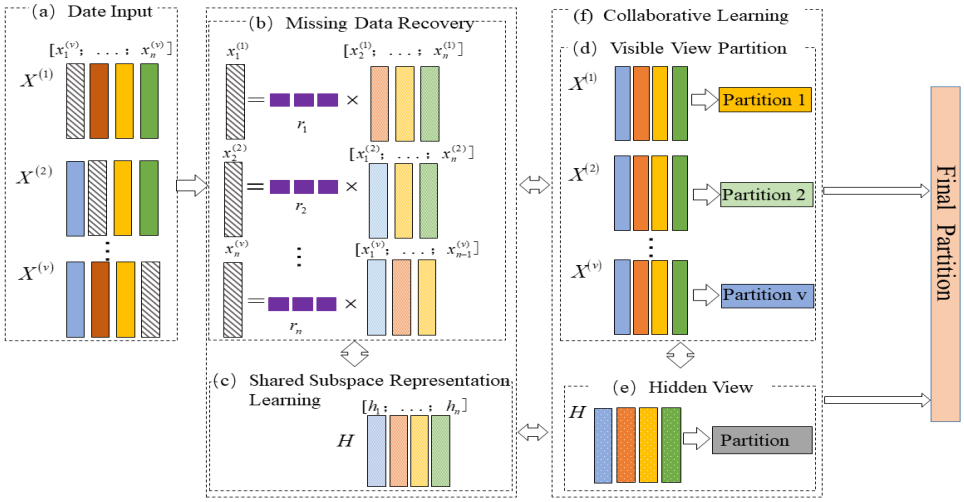


Fig. 2 Overview of CCIM-SLR. CCIM-SLR consists of six major components. (1) dotted module (a) represents the data input module; (2) dotted module (b) corresponds to the recovery of missing data through sparse low-rank techniques; (3) dotted module (c) focuses on shared subspace representation learning; (4) dotted module (d) stands for the partitioning module for the visible view; (5) dotted module (e) represents the partitioning module for the hidden view; (6) dotted module (f) signifies a collaborative learning partition module.

The components and the pipeline of CCIM-SLR are illustrated in Fig. 2. As shown, the proposed framework mainly includes three key modules, i.e., the data input module, the missing data recovery module, and the collaborative learning module. These modules are described in detail in the following sections.

3.1 Shared hidden subspace learning based on SRRS

The low-rank sparse representation-based methods have widely been used to recover missing data [28–30]. We introduce the γ-norm [31], with the primary goal of obtaining estimations that closely approximate the true rank. Specifically, the γ-norm for matrix R is defined as follows:

$$\|R\|_\gamma = \sum_{i=1} \frac{(1+\gamma)\epsilon_{R_i}}{\gamma + \epsilon_{R_i}}, \quad \gamma > 0 \tag{4}$$

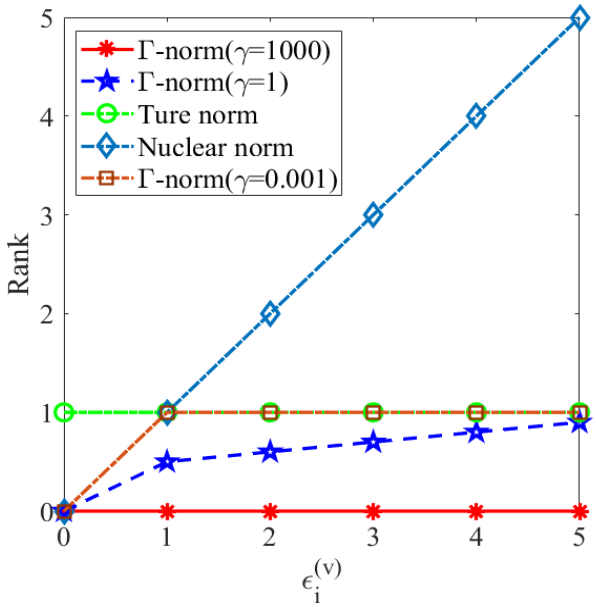


Fig. 3 The performance of different functions on rank estimation changes with the change in positive singular value $\epsilon_i^{(v)}$ (true rank is 1).

In Equation (4), γ is a penalty parameter, and the i-th singular value of the SVD decomposition of the matrix R is denoted as ϵ_{R_i}. However, the γ-norm applies only to the processing of single-view data. To extend to multiview data, we introduce the Γ-norm [24] to implement the low-rank constraint on matrix $R^{(v)}$, with the following formula:

$$\|R^{(v)}\|_G = \sum_{v=1}^{s} \frac{\epsilon_i^{(v)}}{\epsilon_i^{(v)} + \gamma} \tag{5}$$

where $\epsilon_i^{(v)}$ is the i-th singular value of matrix $R^{(v)}$, and if $\gamma \to 0$, we have $\|R^{(v)}\|_G \to$ rank $(R^{(v)})$. An appropriate value of γ (e.g., $\gamma = 0.001$) is chosen in such as way that a value that is closer to the genuine rank is obtained.

Fig. 3 is to demonstrate the influence of various functions on the matrix rank as the singular values vary. From Fig. 3, it becomes evident that the Γ-norm outperforms others in terms of convergence speed and its ability to approach the true rank. The experimental results show that when γ=0.001, the obtained results are closest to the true rank. To obtain the incomplete multiview shared hidden subspace of data, we combine the above terms into one model as follows:

$$\begin{aligned} &\min_{\{R^{(v)},E^{(v)},A^{(v)},H\}} \sum_{v=1}^{s} \lambda_1\|R^{(v)}\|_1 + \lambda_2\|R^{(v)}\|_G + \lambda_3\|E^{(v)}\|_{2,1} + \lambda_4\|A^{(v)}\|_{2,1} \\ &s.t. \forall v, \mathcal{P}_v(\widehat{R}^{(v)}X_o^{(v)}) = H + E^{(v)}, H^\top H = I, X^{(v)} = \widehat{R}^{(v)}X_o^{(v)} \\ &\quad H \geq 0 \end{aligned} \tag{6}$$

where λ_1, λ_2, λ_3 and λ_4 are the weight parameters for the data $X = \{X^{(1)}, ...X^{(v)}\}$, $X^{(v)} = [X_{cp}^{(v)}; X_o^{(v)}]$ denotes multiview data, $X_{cp}^{(v)} \in \mathbb{R}^{\bar{m}^{(v)} \times d^{(v)}}$ represents the matrix consisting of missing samples, $X_o^{(v)} \in \mathbb{R}^{m^{(v)} \times d^{(v)}}$ represents the matrix consisting of observed samples, $d^{(v)}$ is the sample dimension of the v-th view, $\bar{m}^{(v)}$ is missing samples, and $m^{(v)}$ is the observed samples. The total number of samples is m (m=$\bar{m}^{(v)}$+$m^{(v)}$). $H \in \mathbb{R}^{m \times t}$ stands for the obtained subspace representation of a hidden view, and t is the unified dimension of the subspace. The orthogonal constraint $H^\top H = I$ makes the bases independent of each other.

To fully utilize the data observed in other views to recover missing data, we formulate the missing sample linear reconstruction as Equation (7):

$$\forall v, X^{(v)} = \widehat{R}^{(v)}X_o^{(v)} \tag{7}$$

where $\widehat{R}^{(v)} \in \mathbb{R}^{m \times m^{(v)}}$ is composed of the matrix $R^{(v)}$, and the identity matrix constructed from the indices of the observed samples. To acquire a better representation of subspace H, we utilize the following operation:

$$\begin{aligned} \mathcal{P}_v(\widehat{R}^{(v)}X_o^{(v)}) &= \widehat{R}^{v}X_o^{(v)}A^{(v)} \\ &= H + E^{(v)} \end{aligned} \tag{8}$$

where $E^{(v)} \in \mathbb{R}^{m \times t}$ is the noise matrix from the original space to subspace $H^{(v)}$, and $A^{(v)} \in \mathbb{R}^{d^{(v)} \times t}$ is the linear transformation matrix that converts the original spatial data to subspace $H^{(v)}$. To guarantee the sparsity of the data, we add L_1 constraints

[32] to the matrix R. In addition, the $L_{2,1}$ norm third and fourth terms of Equation (6) are added to discard irrelevant features.

Similar to the method proposed by [23], our proposed model (Equation (6)) recovers data using constraints on linear representations. Most importantly, Equation (6) introduces the Γ-norm to impose low-rank constraints on the linear representation matrix. Such a cutting-edge nonconvex low-rank representation method can better capture the correlation between samples than the kernel norm.

3.2 Incomplete multiview co-clustering by sparse low-rank representation

Most of the existing incomplete multiview learning methods based on subspaces are two-step multiview clustering methods (Step 1: subspace acquisition, and Step 2: clustering). These methods do not effectively combine the clustering process with the filling process. Therefore, we integrate clustering results and shared hidden subspace learning into the same objective function as expressed in Equation (9).

$$\begin{aligned}
&\min_{\{R^{(v)},E^{(v)},A^{(v)},H,Z^{(v)},\widetilde{Z}\}} \lambda\sum_{i=1}^{c}\sum_{j=1}^{m}u_{ij}\|h_j-\widetilde{z}_i\|^2\\
&+(1-\lambda)\sum_{v=1}^{s}\sum_{i=1}^{c}\sum_{j=1}^{m}u_{ij}\|{x_j}^{(v)}-{z_i}^{(v)}\|^2\\
&+\sum_{v=1}^{s}(\lambda_1\|R^{(v)}\|_1+\lambda_2\|R^{(v)}\|_G+\lambda_3\|E^{(v)}\|_{2,1}+\lambda_4\|A^{(v)}\|_{2,1})\\
&s.t.\forall v,\mathcal{P}_v(\widehat{R}^{(v)}X_o^{(v)})=H+E^{(v)},H^\top H=I,X^{(v)}=\widehat{R}^{(v)}X_o^{(v)}\\
&H\geq 0,\sum_{i=1}^{c}u_{ij}=1,u_{ij}\in(0,1),1\leqslant j\leqslant m
\end{aligned} \tag{9}$$

where λ is the clustering model weight parameter, $U\in\{0,1\}^{c*m}$ is the partition matrix, c is the number of clusters, and m is the number of complete samples. When sample j belongs to class i, u_{ij}=1, and 0 otherwise. $\widetilde{Z}=[\widetilde{z}_1,\widetilde{z}_2,....\widetilde{z}_c]$ and $Z^{(v)}=\left[z_1^{(v)},z_2^{(v)},....z_c^{(v)}\right]$ are the cluster center of the hidden view and cluster centers of each view, respectively.

In Equation (9), the first two terms are about the partition matrix with a clustering algorithm, the goal of which is to obtain the global and local partitions of all incomplete views. With the joint optimization model, we can capture the correlations between and within the intraview and interview samples through their sparse low-rank representation and the hidden view, respectively. By alternating the process of data recovery and clustering, we can obtain a high-performance padding matrix.

3.3 Optimization procedure

This section presents the Alternating Direction Method of Multipliers (ADMMs) as a solution for the problem stated in Equation (9). For that, we introduce several auxiliary variables to transform Equation (9) into the following expression:

$$\begin{aligned}
&\min_{\{R^{(v)},E^{(v)},A^{(v)},H,Z^{(v)},\widetilde{Z}\}} \lambda\sum_{i=1}^{c}\sum_{j=1}^{m}u_{ij}\|h_j-\widetilde{z}_i\|^2 \\
&+(1-\lambda)\sum_{v=1}^{s}\sum_{i=1}^{c}\sum_{j=1}^{m}u_{ij}\|{x_j}^{(v)}-{z_i}^{(v)}\|^2 \\
&+\sum_{v=1}^{s}(\lambda_1\|Q^{(v)}\|_1+\lambda_2\|M^{(v)}\|_G+\lambda_3\|E^{(v)}\|_{2,1}+\lambda_4\|A^{(v)}\|_{2,1}) \\
s.t.&\forall v,\mathcal{P}_v(\widehat{R}^{(v)}X_o^{(v)})=H+E^{(v)},H^\top H=I,X^{(v)}=\widehat{R}^{(v)}X_o^{(v)} \\
&H\geq 0,\sum_{i=1}^{c}u_{ij}=1,u_{ij}\in(0,1),1\leqslant j\leqslant m \\
&R^{(v)}=Q^{(v)},R^{(v)}=M^{(v)}
\end{aligned} \tag{10}$$

Therefore, the augmented Lagrangian function of Equation (10) is equivalent to the following function:

$$\begin{aligned}
L=&\lambda\sum_{i=1}^{c}\sum_{j=1}^{m}u_{ij}\|h_j-\widetilde{z}_i\|^2+(1-\lambda)\sum_{v=1}^{s}\sum_{i=1}^{c}\sum_{j=1}^{m}u_{ij}\|{x_j}^{(v)}-{z_i}^{(v)}\|^2 \\
&+\sum_{v=1}^{s}(\lambda_1\|Q^{(v)}\|_1+\lambda_2\|M^{(v)}\|_G+\lambda_3\|E^{(v)}\|_{2,1}+\lambda_4\|A^{(v)}\|_{2,1}) \\
&+\frac{\mu}{2}\sum_{v=1}^{s}(\left\|R^{(v)}-Q^{(v)}+\frac{C_2^{(v)}}{\mu}\right\|_F^2+\left\|R^{(v)}-M^{(v)}+\frac{C_3^{(v)}}{\mu}\right\|_F^2 \\
&+\left\|\widehat{R}^{(v)}X_o^{(v)}A^{(v)}-H-E^{(v)}+\frac{C_1^{(v)}}{\mu}\right\|_F^2)
\end{aligned} \tag{11}$$

where $C_1^{(v)}$, $C_2^{(v)}$, $C_3^{(v)}$ are Lagrange multipliers and μ is a penalty parameter. Then, we can solve all unknown variables in the objective function (11) by alternative optimization as follows:

Update variable $R^{(v)}$: By removing irrelevant terms and fixing the other variables, the function becomes:

$$\begin{aligned}
L(R^{(v)})=&\left\|R^{(v)}-Q^{(v)}+\frac{C_2^{(v)}}{\mu}\right\|_F^2+\left\|R^{(v)}-M^{(v)}+\frac{C_3^{(v)}}{\mu}\right\|_F^2 \\
&+\left\|\widehat{R}^{(v)}X_o^{(v)}A^{(v)}-H-E^{(v)}+\frac{C_1^{(v)}}{\mu}\right\|_F^2
\end{aligned} \tag{12}$$

The partial derivative of $L(R^{(v)})$ with respect to $R^{(v)}$ is given by:

$$\begin{aligned}\frac{\partial L(R^{(v)})}{\partial(R^{(v)})} &= \frac{\partial(\left\|R^{(v)} - S_1^{(v)}\right\|_F^2)}{\partial(R^{(v)})} + \frac{\partial\left\|R^{(v)} - S_2^{(v)}\right\|_F^2}{\partial(R^{(v)})} \\ &\quad + \frac{\partial\left\|R^{(v)}X_o^{(v)}A^{(v)} + \iota(S_3^{(v)})\right\|_F^2}{\partial(R^{(v)})} \\ &= 2(R^{(v)} - S_1^{(v)}) + 2(R^{(v)} - S_2^{(v)}) + 2R^{(v)}X_o^{(v)}A^{(v)}A^{(v)^T}{X_o}^{(v)^T} \\ &\quad + 2\iota(S_3^{(v)})A^{(v)^T}{X_o}^{(v)^T}\end{aligned} \tag{13}$$

where $S_1^{(v)} = Q^{(v)} - \frac{C_2^{(v)}}{\mu}$, $S_2^{(v)} = M^{(v)} - \frac{C_3^{(v)}}{\mu}$, $S_3^{(v)} = -H - E^{(v)} + \frac{C_1^{(v)}}{\mu}$, and the ι operation refers to selecting the row to represent the missing sample from the matrix according to the index of the missing sample.

By setting $\partial L(R^{(v)})/\partial(R^{(v)}) = 0$, the optimal $R^{(v)}$ can be obtained as follows:

$$R^{(v)} = \frac{-\iota(S_3^{(v)})A^{(v)^T}X_o^{(v)^T} + S_1^{(v)} + S_2^{(v)}}{2I + {X_o}^{(v)}A^{(v)}A^{(v)^T}{X_o}^{(v)^T}} \tag{14}$$

Update variable $Q^{(v)}$: By eliminating irrelevant terms and holding the other variables constant, $Q^{(v)}$ can be computed as follows:

$$\min_{\{Q^{(v)}\}} \lambda_1 \left\|Q^{(v)}\right\|_1 + \frac{\mu}{2}\left\|R^{(v)} - Q^{(v)} + \frac{C_2^{(v)}}{\mu}\right\|_F^2 \tag{15}$$

Equation (15) can be computed as[33]:

$$Q^{(v)} = \vartheta_{\frac{\lambda_1}{\mu}}(R^{(v)} + \frac{C_2^{(v)}}{\mu}) \tag{16}$$

Update variable $M^{(v)}$: By removing irrelevant terms and fixing the other variables in Equation (11), we can calculate $M^{(v)}$ as follows:

$$M^{(v)} = \underset{\{M^{(v)}\}}{\operatorname{argmin}} \lambda_2\|M^{(v)}\|_G + \frac{\mu}{2}\left\|M^{(v)} - P^{(v)}\right\|_F^2 \tag{17}$$

where $P^{(v)} = R^{(v)} + \frac{C_3^{(v)}}{\mu}$. We set the nonconvex surrogate of $\operatorname{rank}(M^{(v)})$ as $\|M^{(v)}\|_G$. It is difficult to obtain the solution from Equation (17) because it is a nonconvex function. It can be solved for a nonconvex function through regularization techniques and the difference of convex (DC) programming proposed by Moreau-Yosida [34]. Hence, the subproblem becomes:

$$M^{(v)^{t+1}} = \underset{\{M^{(v)^t}\}}{\operatorname{argmin}} \lambda_2\|M^{(v)^t}\|_G + \frac{\mu^t}{2}\left\|M^{(v)^t} - P^{(v)^t}\right\|_F^2 \tag{18}$$

To solve Equation (18), we develop Theorem 1 and provide the proof as below.

Theorem 1. Let $P = U\Sigma_P V^T$ be the singular value decomposition(SVD) of P, where $\Sigma_P = diag(\sigma_P)$. Set $F(M^{(v)}) = \|M^{(v)^t}\|_G = f \circ \sigma_M$.

$$\min_{\{M^{(v)}\}} F(M^{(v)}) + \frac{\mu}{2}\left\|M^{(v)} - P\right\|_F^2 \tag{19}$$

Therefore, the problem of the next optimal solution is transformed into $M^* = U\Sigma_M^* V^T$, where $\Sigma_M^* = diag(\sigma^*)$ and $\sigma^* = prox_{f,\mu}(\sigma_P)$, and $prox_{f,\mu}(\sigma_P)$ is the MoreauYosida operator, as follows:

$$prox_{f,\mu}(\sigma_P) = \underset{\{\sigma\}}{\operatorname{argmin}} f(\sigma) + \frac{\mu}{2}\|\sigma - \sigma_P\|_2^2 \tag{20}$$

Proof. Given $P = U\Sigma_P V^T$, $\Sigma_P = U^T P V$, and recording $D^{(v)} = U^{(v)^T} M^{(v)} V^{(v)}$. Since it has the same singular value as $M^{(v)}$, the formula is converted as follows:

$$F(M^{(v)}) + \frac{\mu}{2}\left\|M^{(v)} - P\right\|_F^2, \tag{21}$$

$$= F(D^{(v)}) + \frac{\mu}{2}\left\|D^{(v)} - \Sigma_P\right\|_F^2, \tag{22}$$

$$\geq F(\Sigma_D^{(v)}) + \frac{\mu}{2}\left\|\Sigma_D^{(v)} - \Sigma_P\right\|_F^2, \tag{23}$$

$$= F(\Sigma_M^{(v)}) + \frac{\mu}{2}\left\|\Sigma_M^{(v)} - \Sigma_P\right\|_F^2, \tag{24}$$

$$= f(\sigma) + \frac{\mu}{2}\|\sigma - \sigma_P\|_2^2 \quad , \tag{25}$$

$$\geq f(\sigma^*) + \frac{\mu}{2}\|\sigma^* - \sigma_P\|_2^2 \quad , \tag{26}$$

It should be noted that Equation (22) is valid because the Frobenius norm is unitarily invariant. Equation (23) is based on the Hoffman-Wielandt inequality and Equation (24) holds as we have $\Sigma_M^{(v)} = \Sigma_D$. Thus, Equation (24) is the lower bound of Equation (21) as $\Sigma_D^{(v)} = \Sigma_M^{(v)} = M^{(v)} = U^{(v)^T} D^{(v)} V^{(v)}$ holds, and the SVD of $D^{(v)}$ is $D^{(v)} = U^{(v)^T} \Sigma_D^{(v)} V^{(v)}$. When we perform a minimization operation on Equation (25), we obtain σ^*. Therefore, we have $D^* = U diag(\sigma^*) V^T$, the optimal solution to Equation (19). We have completed the proof of Theorem 1.

Through the inspiration from the Moreau-Yosida regularization technique and the difference of convex (DC) programming, we transform Equation (18) to address the difference between two convex functions. The concave term is iterated for optimization in each iteration. Then, the optimization formula is as follows:

$$\sigma^{t+1} = \operatorname{argmin} f(\sigma^t) + \frac{\mu^t}{2}\left\|\sigma^t - \sigma_P^t\right\|_2^2 \tag{27}$$

which admits a closed-form solution [35], as shown in Equation (28).

$$\sigma^{t+1} = (\sigma_T^t - \frac{\varphi_t}{\mu_t})_+ \tag{28}$$

at point σ^t, the gradient representation of $f(\cdot)$ is denoted as $\varphi_t = \partial f(\sigma^t)$ and $U^{(v)}diag(\sigma_P^t)^{(v)}V^{(v)^T}$ is the singular value decomposition of $(R^{(v)} + \frac{C_3^{(v)}}{\mu})$. Through optimization iterations, the final convergence obtains the best advantage σ^*. The solution can be derived as follows:

$$M^{(v)^{t+1}} = U^{(v)}diag(\sigma^{(v)^*})V^{(v)^T}. \tag{29}$$

Update variable $E^{(v)}$: after removing the irrelevant terms and fixing the other variables, the subproblem becomes:

$$L(E^{(v)}) = \lambda_3 \left\| E^{(v)} \right\|_{2,1} + \frac{\mu}{2} \left\| E^{(v)} - S_4^{(v)} \right\|_F^2 \tag{30}$$

where $S_4^{(v)} = \widehat{R}^v X_o^{(v)} A^{(v)} - H + \frac{C_1^{(v)}}{\mu}$, by setting $\partial L(E^{(v)})/\partial(E^{(v)}) = 0$, the solution can be derived as follows:

$$E^{(v)} = \frac{\mu S_4^{(v)}}{\lambda_3 Y^{(v)} + \mu I} \tag{31}$$

where $Y^{(v)} = diag\left\{ \frac{1}{\left\| e_1^{(v)} \right\|_2}, \frac{1}{\left\| e_2^{(v)} \right\|_2}, \cdots \frac{1}{\left\| e_1^{(v)} \right\|_n} \right\}$, $e_i^{(v)}$ represents the i-th row vector of $E^{(v)}$.

Update variable $A^{(v)}$: by removing the irrelevant terms and fixing the other variables, we can calculate $A^{(v)}$ as follows:

$$L(A^{(v)}) = \lambda_4 \left\| A^{(v)} \right\|_{2,1} + \frac{\mu}{2} \left\| \widehat{R}^{(v)} X_o^{(v)} A^{(v)} + S_3^{(v)} \right\|_F^2 \tag{32}$$

where $S_3^{(v)} = -H - E + \frac{C_1^{(v)}}{\mu}$. By setting $\partial L(A^{(v)})/\partial(A^{(v)}) = 0$, we can obtain the optimal $A^{(v)}$ as follows:

$$A^{(v)} = \frac{-\mu X_o^{(v)^T} \widehat{R}^{(v)^T} S_3^{(v)}}{\lambda_4 G^{(v)} + \mu X_o^{(v)^T} \widehat{R}^{(v)^T} \widehat{R}^{(v)} X_o^{(v)}} \tag{33}$$

where $G^{(v)} = diag\left\{ \frac{1}{\left\| a_1^{(v)} \right\|_2}, \frac{1}{\left\| a_2^{(v)} \right\|_2}, \cdots \frac{1}{\left\| a_1^{(v)} \right\|_n} \right\}$, and $a_i^{(v)}$ is the i-th row vector of $A^{(v)}$.

Update variable H: by removing the irrelevant terms and fixing the other variables, we can calculate H as follows:

$$L(H) = \lambda \sum_{i=1}^{c} \sum_{j=1}^{m} u_{ij} \left\| h_j - \widetilde{z}_i \right\|^2 + \frac{\mu}{2} \sum_{v=1}^{s} \left(\left\| B^{(v)} - H \right\|_F^2 \right) \tag{34}$$

where $B^{(v)} = \widehat{R}^v X_o^{(v)} A^{(v)} - E^{(v)} + \frac{C_1^{(v)}}{\mu}$. By setting $\partial L(H)/\partial(H) = 0$, we can obtain the optimal H as follows:

$$H = \frac{2\lambda \sum_{i=1}^{c} U_{in} Z_{it} + \mu \sum_{v=1}^{s} B_{nt}^{(v)}}{2\lambda \sum_{i=1}^{c} U_{in} + \mu s} \tag{35}$$

Update variable U: We also use the K-means algorithm and Euclidean distance to measure the similarity between samples. If the distance from the i-th sample to the j-th cluster center is smaller than the distance to other cluster centers, the element in the matrix u_{ij} is 1, and 0 otherwise. According to our proposed model, the distance D_{ij} can be expressed as:

$$D_{ij} = \lambda \sum_{i=1}^{c} \sum_{j=1}^{m} u_{ij} \left\| h_j - \widetilde{z}_i \right\|^2 + (1-\lambda) \sum_{i=1}^{s} \sum_{i=1}^{c} \sum_{j=1}^{m} u_{ij} \left\| x_j^{(v)} - z_i^{(v)} \right\|^2 \tag{36}$$

According to the K-means algorithm, the specific update method of matrix U is as follows:

$$u_{ij} = \begin{cases} 1, \forall k \in [1,m] \; and \; k \neq j, & D_{ij} \leq D_{ik} \\ 0, \qquad\qquad \exists k \in [1,m], & D_{ij} \geq D_{ik} \end{cases} \tag{37}$$

Update variable $Z^{(v)}$: To update the variable $Z^{(v)}$, we can calculate it by removing irrelevant terms and holding other variables constant, as shown below:

$$L(Z^{(v)}) = (1-\lambda) \sum_{i=1}^{c} \sum_{j=1}^{m} u_{ij} \left\| x_j^{(v)} - z_i^{(v)} \right\|^2 \tag{38}$$

$Z^{(v)}$ can be solved as follows:

$$z_i^{(v)} = \frac{\sum_j^N u_{ij} x_j^{(v)}}{\sum_{j=1}^N u_{ij}} \tag{39}$$

Update variable $\widetilde{Z}$: by removing the irrelevant terms and fixing the other variables, we can calculate $\widetilde{Z}$ as follows:

$$L(\widetilde{Z}) = \lambda \sum_{i=1}^{c} \sum_{j=1}^{m} u_{ij} \left\| h_j - \widetilde{z}_i \right\|^2 \tag{40}$$

$\widetilde{Z}$ can be solved as follows:

$$\widetilde{z}_i = \frac{\sum_j^N u_{ij} h_j}{\sum_{j=1}^N u_{ij}} \tag{41}$$

Update Variables $C_1^{(v)}$, $C_2^{(v)}$, $C_3^{(v)}$: We update $C_1^{(v)}$, $C_2^{(v)}$, $C_3^{(v)}$ as follows:

$$C_1^{(v)} = C_1^{(v)} + \mu(\widehat{R}^{(v)} X_o^{(v)} A^{(v)} - H - E^{(v)}) \tag{42}$$

$$C_2^{(v)} = C_2^{(v)} + \mu(R^{(v)} - Q^{(v)}) \tag{43}$$

$$C_3^{(v)} = C_3^{(v)} + \mu(R^{(v)} - M^{(v)}) \tag{44}$$

Update Variables μ: we update μ by:

$$\mu = min(\rho\mu, \mu_0) \tag{45}$$

where ρ and μ_0 represent preset parameters.

3.4 Complexity analysis of CCIM-SLR

As discussed in Section 3.3, the computational expense of our algorithm is mainly due to operations such as matrix inversion and singular value decomposition of self-matrices. The algorithm of CCIM-SLR is summarized in Algorithm 1. The computational complexities of Steps 8, 10, and 11 in Algorithm 1 are approximately $\mathrm{O}(m^{(v)^3})$, $\mathrm{O}(\bar{m}^{(v)}m^{(v)^2})$, and $\mathrm{O}(m^{(v)^3})$, respectively. Therefore, the complexity of the entire optimization of the algorithm is approximately $\mathrm{O}(\tau(2m^{(v)^3} + \bar{m}^v m^{(v)^2}))$, where τ is the number of iterations, $\bar{m}^{(v)}$ is the number of missing samples, and $m^{(v)}$ is the number of observed samples of the view.

Algorithm 1 CCIM-SLR

Require: Incomplete multi-view dataset $X^{(v)}$, parameters λ, λ_1, λ_2, λ_3, λ_4, ρ, μ_0.
Ensure: The resulting clusters.

```
Initialize H;
Initialize U;
Initialize E^(v);
Initialize V^(v);
Initialize Ṽ;
while not converge do
    for v = 1 to V do
        Update R^(v) via Eq. (14);
        Update Q^(v) via Eq. (16);
        Update M^(v) via Eq. (29);
        Update E^(v) via Eq. (31);
        Update A^(v) via Eq. (33);
        Update Z^(v) via Eq. (39);
        Update C_1^(v) via Eq. (42);
        Update C_2^(v) via Eq. (43);
        Update C_3^(v) via Eq. (44);
    end for
    Update H via Eq. (35);
    Update U via Eq. (37);
    Update Z̃ by solving Eq. (41);
    Update μ by μ = min(ρμ; max_μ ) ;
end while
```

3.5 Convergence analysis of CCIM-SLR

In addition to the update step of $M^{(v)}$, the other steps are easily found to be bounded. Therefore, we analyze the effect of the update step of $M^{(v)}$ on our objective function.

We write $\|M^{(v)}\|_G$ as $K(M^{(v)})$ in Eq. (9):

$$\begin{aligned}\min J(M^{(v)}, R^{(v)}, C_3^{(v)}, \mu) = \sum_{v=1}^{s}(K(M^{(v)}) + \frac{\mu}{2}\left\|M^{(v)} - R^{(v)}\right\|_F^2 \\ + < \frac{C_3^{(v)}}{\mu}, M^{(v)} - R^{(v)} >)\end{aligned} \tag{46}$$

where $< .,. >$ represents the sum of the products of corresponding components between two matrices.

Lemma 1. $M^{(v)^t}$ and $R^{(v)^t}$ are bounded if $\sum_{t=1}^{\infty} \frac{(\mu^t - \mu^{t-1})}{2(\mu^{t-1})^2} < \infty$.

Proof. With some algebra, we can obtain:

$$\begin{aligned}&J(M^{(v)^t}, R^{(v)^t}, C_3^{(v)^t}, \mu^t) \\ =&J(M^{(v)^t}, R^{(v)^t}, C_3^{(v)^{t-1}}, \mu^{t-1}) \\ &+ \frac{(\mu^t - \mu^{t-1})}{2}\left\|M^{(v)} - R^{(v)}\right\|_F^2 \\ &+ Tr[(C_3^{(v)^t} - C_3^{(v)^{t-1}})(M^{(v)} - R^{(v)})] \\ =&J(M^{(v)^t}, R^{(v)^t}, C_3^{(v)^{t-1}}, \mu^{t-1}) \\ &+ \frac{(\mu^t - \mu^{t-1})}{2(\mu^{t-1})^2}\left\|(C_3^{(v)^t} - C_3^{(v)^{t-1}})\right\|_F^2\end{aligned} \tag{47}$$

Then,

$$\begin{aligned}&J(M^{(v)^{t+1}}, R^{(v)^{t+1}}, C_3^{(v)^t}, \mu^t) \\ \leq&J(M^{(v)^{t+1}}, R^{(v)^t}, C_3^{(v)^t}, \mu^t) \\ \leq&J(M^{(v)^t}, R^{(v)^t}, C_3^{(v)^t}, \mu^t) \\ \leq&J(M^{(v)^t}, R^{(v)^t}, C_3^{(v)^{t-1}}, \mu^{t-1}) \\ &+ \frac{(\mu^t - \mu^{t-1})}{2(\mu^{t-1})^2}\left\|(C_3^{(v)^t} - C_3^{(v)^{t-1}})\right\|_F^2\end{aligned} \tag{48}$$

By iterating the above inequality (48) t times, we obtain:

$$\begin{aligned}&J(M^{(v)^{t+1}}, R^{(v)^{t+1}}, C_3^{(v)^t}, \mu^t) \\ \leq&J(M^{(v)^1}, R^{(v)^1}, C_3^{(v)^0}, \mu^0) \\ &+ \sum_{i=1}^{t}\frac{(\mu^i - \mu^{i-1})}{2(\mu^{i-1})^2}\left\|(C_3^{(v)^t} - C_3^{(v)^{t-1}})\right\|_F^2\end{aligned} \tag{49}$$

As $\left\|(C_3^{(v)^t} - C_3^{(v)^{t-1}})\right\|_F^2$ is bounded, the other terms included in the right-hand side of the inequality are also bounded. Therefore, $J(M^{(v)^{t+1}}, R^{(v)^{t+1}}, C_3^{(v)^t}, \mu^t)$ is upper bounded.

In addition, we have

$$\begin{aligned} & J(M^{(v)^{t+1}}, R^{(v)^{t+1}}, C_3^{(v)^t}, \mu^t) + \frac{1}{2\mu^t}\left\|(C_3^{(v)^t})\right\|_F^2 \\ =& K(M^{(v)^{t+1}}) + \frac{\mu^t}{2}\left\|M^{(v)^{t+1}} - R^{(v)^{t+1}} + \frac{C_3^{(v)^t}}{\mu^t}\right\|_F^2 \end{aligned} \tag{50}$$

By observing several terms on the right side of Equation (50), we find that each of them is finite, so $M^{(v)^{t+1}}$ and $R^{(v)^{t+1}}$ are also finite respectively. Therefore, $\left\{M^{(v)^t}\right\}$ and $\left\{R^{(v)^t}\right\}$ are also bounded.

Lemma 2. Let $\left\{M^{(v)^t}, R^{(v)^t}, C_3^{(v)^t}\right\}$ be the sequence and $\left\{M^{(v)^*}, R^{(v)^*}, C_3^{(v)^*}\right\}$ be an accumulation point. Then $\left\{M^{(v)^*}, R^{(v)^*}\right\}$ is a stationary point if we have $\lim_{t\to\infty} \mu^t \left(\boldsymbol{R}^{(v)^{t+1}} - \boldsymbol{R}^{(v)^t}\right) \to 0$.

Proof. The sequence $\left\{M^{(v)^t}, R^{(v)^t}, C_3^{(v)^t}\right\}$ is bounded as shown in Lemma 2. By the Bolzano-Weierstrass theorem, at least one accumulation point must exist in this sequence, e.g., $\left\{M^{(v)^*}, R^{(v)^*}, C_3^{(v)^*}\right\}$. Therefore, we presume that $\left\{M^{(v)^t}, R^{(v)^t}, C_3^{(v)^t}\right\}$ itself converges to $\left\{M^{(v)^*}, R^{(v)^*}, C_3^{(v)^*}\right\}$.

Since $R^{(v)^t} - M^{(v)^t} = (C_3^{(v)^t} - C_3^{(v)^{t-1}})/\mu^{t-1}$ holds, we have $\lim_{t\to\infty} R^{(v)^t} - M^{(v)^t} = 0$. Therefore, the primal feasibility condition is fulfilled.

For $M^{(v)^{t+1}}$, it holds that

$$\begin{aligned} & \partial_M \left(M^{(v)^{t+1}}, R^{(v)^t}, C_3^{(v)^t}, \mu^t\right) |_{M(v)^{t+1}} \\ =& \partial_M K\left(M^{(v)^{t+1}}\right) + {C_3}^{(v)^t} + \mu^t \left(R^{(v)^t} - M^{(v)^t}\right) \\ =& \partial_M K\left(M^{(v)^{t+1}}\right) + {C_3}^{(v)^{t+1}} + \mu^t \left(R^{(v)^{t+1}} - R^{(v)^t}\right) = 0 \end{aligned} \tag{51}$$

If the singular value decomposition of $M^{(v)}$ is $\boldsymbol{U}^{(v)} \operatorname{diag}\left(\sigma_i^{(v)}\right) V^{(v)^T}$ according to Theorem 1,

$$\partial_M K\left(M^{(v)^{t+1}}\right) |_{M(v)^{t+1}} = U \operatorname{diag}\left(\tau^{(v)}\right) V^{(v)^T}, \tag{52}$$

where $\tau_i = \gamma/(\gamma + \sigma_i)^2$ when $\sigma_i \neq 0$; else, it acts as $1/\gamma$. Since $\sigma_i \in (0, 1/\gamma]$ is finite, $\partial_M K\left(M^{(v)^{t+1}}\right) |_{M(v)^{t+1}}$ is bounded. $C_3^{(v)^t}$ is bounded as a Lagrange multiplier. $\mu^t \left(R^{(v)^{t+1}} - R^{(v)^t}\right)$ is bounded. Under the assumption that $\lim_{t\to\infty} \mu^t \left(R^{(v)^{t+1}} - R^{(v)^t}\right) \to 0$,

$$\partial_M K\left(M^{(v)^*}\right) + {C_3}^{(v)^*} = 0 \tag{53}$$

Hence, $\left\{M^{(v)^*}, R^{(v)^*}, C_3^{(v)^*}\right\}$ satisfies the Karush–Kuhn–Tucker conditions of $J(M^{(v)^{t+1}}, R^{(v)^{t+1}}, C_3^{(v)^t})$. Therefore, $\left\{M^{(v)^*}, R^{(v)^*}\right\}$ is the point satisfying the condition.

4 Experiments

This section reports in detail the performance evaluation of CCIM-SLR by comparing it with the state-of-the-art methods against five real-world datasets. Furthermore, we present experimental results on the proposed optimization approach, and its convergence property to demonstrate the efficiency of CCIM-SLR and the robustness of CCIM-SLR.

4.1 Datasets

To validate the clustering performance of the proposed method under different data dimensions, we used the five representative datasets in our experiments. The statistics of the datasets are listed in Table 1.

Table 1 Statistics of the datasets

Dataset	Clusters	Views	samples	Features
SensIT300	3	2	300	50/50
Statlog	7	2	2310	9/10
Wisconsin	5	2	265	1703/265
WebKB	2	2	1051	1840/3000
Yale	15	3	165	4096/3304/ 6750

- **SensIT300** [1][36]: SensIT300 contains sensory data collected from an intelligent transportation system targeting three vehicle types. This is one of the main datasets used in many research papers to evaluate the performance of clustering algorithms. This dataset consists of 300 samples under three different classes with two views, and the sample data in each view consists of features of 50 dimensions respectively. The three classes are three types of transportation, while the views are split into vibration information and sound view obtained through sensor transmission.
- **Statlog** [2][37]: The Statlog dataset was collected by the Vision Group, University of Massachusetts. The total number of samples in this dataset is 2310. The dataset contains seven kinds of outdoor images that were hand-segmented to create a classification for every pixel. Each sample in the dataset has feature dimensions of either 9 or 10.

[1] https://github.com/Liuzhenjiao123/multiview-data-sets/blob/master/sensIT300.mat
[2] https://github.com/Liuzhenjiao123/multiview-data-sets/tree/master

- **Wisconsin** [3][38]: The Wisconsin dataset is a real-world multiview dataset that contains 256 samples with different descriptions from 5 different categories (Student Pages, Program Pages, Course Pages, Staff Pages, and Faculty and Staff Pages). The content view and reference view are the two types of views identified from each sample, with 1703 and 265 feature dimensions, respectively.
- **WebKB** [4] [39]: The WebKB dataset consists of 1051 samples under 2 classes. Each sample in this dataset corresponds to two types of features: i) those derived from the textual content of the web page, and ii) those derived from the anchor text containing links to other web pages. In this dataset, the dimension of the linked representation is 1840, while the other dimension is 3000.
- **Yale** [5][40]: The Yale dataset is a collection of 165 pictures from 15 people. The pictures are distinguished by different expressions, gestures, and lights. The Yale dataset contains three views. The feature dimensions of each view are 4096, 3304, and 6750, respectively.

4.2 Baseline approaches

To validate the performance of the proposed CCIM-SLR, we compared it with five IMC methods: IMC-GRMF [41], IMSC-AGL [42], UEAF [14], DAIMC [12], and HCP-IMSC [43].

- **IMC-GRMF**: The IMC-GRMF method uses the orthogonal matrix factorization technique to learn the latent subspace. The local information of each view is incorporated to help fuse the complementary information of views, which results in a better-shared representation.
- **IMSC-AGL**: The IMSC-AGL method utilizes low-rank representations of adaptive learning of graphs in a multiview scenario. To obtain more refined low-dimensional representations, this model employs a number of spectral constraints.
- **UEAF**: The UEAF model is designed for incomplete multiview clustering and serves as a unified and robust embedding alignment approach. Differing from other methods, UEAF infers incomplete information by maintaining the consistency of the local structure of the views and learning the local structure shared among multiple views through reversing graph regularization.
- **DAIMC**: The DAIMC method is characterized by weighted semi-NMF: semi-nonnegative matrix factorization, which learns a weight matrix that can be adapted to multiple incomplete cases. This model performs an $L_{2,1}$ regularization to obtain a supplemental cluster-friendly matrix representation that can be shared by views.
- **HCP-IMSC**: The HCP-IMSC method uses higher-order information to improve the clustering performance of incomplete multiview. Tensor decomposition is adopted in the process of capturing higher-order association relations. Then, under hypergraph-induced superLaplace regularization, the missing view samples are restricted to be reconstructed by neighboring samples.

---

[3] https://lig-membres.imag.fr/grimal/data.html
[4] https://github.com/Liuzhenjiao123/dataset4
[5] http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html

4.3 Experimental setups

We removed 10%, 30%, 50%, 70%, and 90% of the sample data in each view from the five datasets with the incomplete view. Specifically, all IMC-GRMF, IMSC-AGL, UEAF, DAIMC, and HCP-IMSC perform post-clustering operations (e.g., K-means) based on obtained latent representations to produce their final clustering results. Considering that the clustering results of K-means are affected by the initialization of seed points, we performed K-means 10 times in the experiment to obtain the average value. For setting the parameters of the compared methods, we choose the values within the parameter ranges specified in the original papers.

4.4 Evaluation metrics

In our experiments, we used four performance metrics to evaluate the clustering performances: NMI - Normalized mutual information [44], ACC - Accuracy [45], ARI - Adjusted Rand index [46], and F-score [47].

Table 2 Mean NMIs(%), ACCs(%), ARIs(%) and F-scores(%) of different methods on SensIT300 , Statlog , Wisconsin , WebKB and Yale datasets

Dataset	Method \ PER	NMI					ACC					ARI					F-score				
		0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
SensIT300	IMSC-AGL	22.50	19.86	16.20	14.40	15.00	65.74	62.73	60.00	58.00	56.33	24.78	21.31	17.24	14.56	15.18	49.85	47.69	44.93	43.37	43.60
SensIT300	DAIMC	20.44	17.29	15.67	10.58	8.89	64.08	60.63	59.18	51.90	50.12	22.54	17.87	16.27	10.43	8.28	48.68	45.73	44.80	41.28	40.59
SensIT300	UEAF	21.23	17.83	15.85	14.09	15.00	65.00	60.67	59.33	58.00	56.00	23.74	17.93	16.05	15.11	12.92	49.52	45.35	44.10	43.32	41.92
SensIT300	IMC-GRMF	15.50	8.36	5.65	4.74	2.19	60.20	51.00	47.33	39.13	35.67	16.51	8.45	5.72	4.18	0.07	44.85	38.98	36.99	36.57	**47.59**
SensIT300	HCP-IMSC	32.17	**30.08**	21.22	18.67	15.32	**72.12**	65.17	57.86	57.33	56.67	**32.17**	**30.16**	21.35	18.67	15.12	**56.12**	50.15	48.05	46.59	43.88
SensIT300	Ours	**32.70**	26.88	**24.65**	**19.93**	**17.37**	68.13	**67.00**	**64.07**	**61.27**	**57.27**	29.27	26.84	**24.51**	**19.86**	**15.21**	53.79	**51.57**	**50.30**	**47.12**	44.39
Statlog	IMSC-AGL	10.81	20.51	13.51	8.76	3.08	27.45	37.69	30.95	24.11	17.92	5.91	15.09	7.63	2.28	0.27	21.71	27.89	21.59	18.96	18.77
Statlog	DAIMC	47.35	39.59	34.15	30.49	26.61	56.66	51.29	44.52	41.19	35.26	36.11	27.27	18.80	18.01	13.56	45.98	38.65	32.63	30.99	27.52
Statlog	UEAF	48.83	37.61	34.98	32.61	29.83	48.09	46.10	43.51	37.40	39.18	25.62	23.60	18.28	13.48	10.42	40.48	35.90	32.40	29.54	27.34
Statlog	IMC-GRMF	43.93	41.76	38.21	31.07	30.31	54.68	53.84	47.45	36.36	40.61	33.85	30.46	28.53	18.14	16.97	43.43	40.69	39.45	30.62	29.41
Statlog	HCP-IMSC	50.43	**46.16**	40.21	36.05	32.51	57.12	53.84	49.45	45.06	39.56	38.12	32.11	26.31	20.17	17.32	48.23	44.19	39.25	34.42	30.25
Statlog	Ours	**50.53**	45.22	**41.43**	**37.39**	**33.84**	**57.52**	**53.97**	**49.76**	**45.36**	**41.40**	**39.02**	**34.09**	**29.98**	**24.93**	**20.09**	**48.52**	**44.45**	**40.66**	**36.14**	**32.15**
Wisconsin	IMSC-AGL	20.98	18.81	14.10	13.64	12.24	42.66	38.87	34.34	34.72	32.57	16.56	11.10	8.29	5.01	5.76	41.04	33.01	32.02	32.11	32.57
Wisconsin	DAIMC	29.87	26.52	24.04	22.10	16.75	51.39	43.77	45.66	48.83	44.00	24.96	16.95	16.52	15.88	12.70	46.01	39.37	39.02	40.87	36.07
Wisconsin	UEAF	35.94	**40.14**	34.25	**33.71**	**29.21**	60.75	57.35	50.56	**55.47**	**45.28**	34.56	34.00	25.34	**26.25**	**16.06**	53.18	50.78	44.07	**47.57**	41.32
Wisconsin	IMC-GRMF	25.69	18.57	10.63	7.9	6.2	43.69	37.21	33.28	39.17	44.10	14.52	9.87	5.22	4.80	0.51	40.04	33.78	30.51	34.16	**45.97**
Wisconsin	HCP-IMSC	27.21	24.13	26.63	25.53	23.63	50.16	40.11	49.32	44.15	38.11	21.12	17.41	20.32	17.01	14.13	42.14	39.18	40.21	14.78	37.42
Wisconsin	Ours	**42.33**	35.35	**36.45**	28.90	24.68	**65.51**	**60.98**	**54.26**	44.45	43.92	**42.32**	**35.57**	**27.56**	17.93	14.24	**59.10**	**54.17**	**47.38**	39.72	40.14
WebKB	IMSC-AGL	65.50	30.11	50.13	35.22	7.1	95.05	82.78	91.34	85.06	62.13	79.36	40.73	67.53	46.84	5.83	92.74	76.54	87.62	79.25	59.30
WebKB	DAIMC	60.57	52.35	41.54	44.96	38.16	93.14	90.67	84.64	90.56	88.79	69.79	63.84	50.72	59.51	53.33	90.89	88.30	83.63	87.80	85.73
WebKB	UEAF	68.41	70.05	64.51	59.63	**61.76**	95.43	95.62	94.86	94.10	**94.57**	80.97	81.77	78.60	75.31	**76.99**	93.23	93.48	92.47	91.55	**92.26**
WebKB	IMC-GRMF	51.70	34.37	2.8	9.1	2.9	92.01	87.82	61.08	78.21	71.36	67.85	52.22	4.36	21.92	10.11	88.59	83.80	59.61	74.87	68.52
WebKB	HCP-IMSC	71.11	68.23	61.12	53.92	45.72	95.12	93.21	92.08	91.72	86.86	82.12	79.21	77.12	67.45	52.75	93.12	92.60	91.13	87.23	81.23
WebKB	Ours	**73.31**	**72.25**	**69.57**	**67.52**	57.99	**96.63**	**96.44**	**95.87**	**95.61**	93.82	**85.53**	**84.77**	**82.61**	**81.33**	73.85	**94.99**	**93.88**	**94.77**	**93.60**	91.30
Yale	IMSC-AGL	**68.49**	65.67	**70.38**	65.71	**68.80**	**67.09**	62.55	**68.97**	61.33	**64.36**	**44.52**	43.71	**47.32**	44.46	**46.97**	**59.75**	56.67	62.06	56.33	**60.12**
Yale	DAIMC	57.16	53.68	54.26	44.24	41.36	53.21	50.18	48.79	37.21	34.42	32.44	26.87	25.21	14.49	11.36	44.83	42.32	41.42	31.36	28.92
Yale	UEAF	61.21	61.87	61.02	60.84	61.97	55.21	55.88	54.85	55.15	55.58	37.81	38.70	37.91	36.84	38.31	50.85	50.93	49.89	50.02	51.14
Yale	IMC-GRMF	63.67	64.12	56.58	46.55	45.27	57.64	58.24	48.30	37.82	36.06	41.75	42.22	32.40	19.04	17.38	53.65	53.71	44.67	33.52	31.67
Yale	HCP-IMSC	64.42	63.77	58.51	53.36	47.92	56.85	60.85	54.79	53.36	47.92	41.60	41.38	35.07	27.81	20.56	52.28	53.12	47.07	40.88	33.90
Yale	Ours	66.58	**72.37**	69.76	**65.74**	64.12	65.55	**63.79**	67.73	**62.41**	60.21	43.12	**54.19**	46.13	**45.02**	43.46	57.79	**63.87**	**62.14**	**56.95**	56.78

Table 3 Ablation study of the CCIM-SLR performance (%) on SensIT300 , Wisconsin, and WebKB datasets

Dataset	Method \ PER	NMI					ACC					F-score				
		0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
SensIT300	Ablation-1	19.84	15.22	12.78	10.11	6.06	61.82	57.13	55.06	51.06	46.13	47.60	44.18	41.81	39.90	37.46
SensIT300	Ablation-2	20.19	15.41	13.14	7.1	4.2	56.82	53.61	47.13	42.41	56.67	47.85	44.24	42.18	37.55	35.78
SensIT300	Ours	**32.70**	**26.88**	**24.65**	**19.93**	**17.37**	**68.13**	**67.00**	**64.07**	**61.27**	**57.27**	**53.79**	**51.57**	**50.30**	**47.12**	**44.39**
Wisconsin	Ablation-1	39.75	36.18	33.71	27.13	24.19	62.64	53.50	51.84	42.86	42.19	53.86	47.25	47.13	39.05	37.98
Wisconsin	Ablation-2	37.20	**38.01**	31.43	27.67	23.25	60.75	55.39	51.92	44.08	41.88	52.39	49.59	45.25	39.70	38.21
Wisconsin	Ours	**42.33**	35.35	**32.45**	**28.90**	**24.68**	**65.51**	**60.98**	**54.26**	**44.45**	**43.92**	**59.10**	**54.17**	**47.38**	**39.72**	**40.14**
WebKB	Ablation-1	68.33	57.15	49.34	41.11	4.1	94.53	88.69	85.11	56.17	50.12	92.09	84.78	80.71	57.82	56.77
WebKB	Ablation-2	63.82	41.49	26.64	14.12	5.23	93.92	83.93	73.26	57.51	57.01	91.32	78.06	66.19	57.84	61.71
WebKB	Ours	**73.31**	**72.25**	**69.57**	**67.52**	**57.99**	**96.63**	**96.44**	**95.87**	**95.61**	**93.82**	**94.99**	**93.88**	**94.77**	**93.60**	**91.30**

Table 4 Two incomplete multi-view clustering methods based on association models exhibit differences in ACCs(%), F-scores(%), running time (seconds), and computational complexity when applied to datasets SensIT300 and Statlog, with 50% incomplete samples in each view. m_o denotes the number of observed samples.

Dataset	Method	ACC	F-score	Running time (seconds)	computational complexity
SensIT300	HCP-IMSC	57.86	48.05	1.4312	$O(sm^3 + s(m-m_o)^3 + cmslog(s) + cm^2s)$
SensIT300	CCIM-SLR	**64.07**	**50.30**	**1.4284**	$O(\tau(2{m^{(v)}}^3 + \bar{m}^v{m^{(v)}}^2))$
Statlog	HCP-IMSC	49.45	39.25	219.9447	$O(sm^3 + s(m-m_o)^3 + cmslog(s) + cm^2s)$
Statlog	CCIM-SLR	**49.76**	**40.66**	**62.7427**	$O(\tau(2{m^{(v)}}^3 + \bar{m}^v{m^{(v)}}^2))$

- **NMI - Normalized mutual information**: The NMI indicator measures the quality of clusters defined as:

$$NMI = \frac{\sum_{i=1}^{C}\sum_{j=1}^{C} N_{i,j} \ln \frac{N_{i,j}}{N_i \hat{N}_j}}{\sqrt{(\sum_{i=1}^{C} N_i \ln \frac{N_i}{N})(\sum_{j=1}^{C} \hat{N}_j \ln \frac{\hat{N}_j}{N})}} \tag{54}$$

  where N is the number of samples in a complete view, N_i and $\hat{N}_j$ are the numbers of samples in the i-th cluster and the number of samples of the j-th label, respectively. The number of samples in the intersection between the i-th cluster and j-th label is represented by $N_{i,j}$.
- **ACC - Accuracy**: ACC measures the cluster quality. ACC is estimated by:

$$ACC = \frac{\sum_{i=1}^{N} \delta(map(r_i), l_i)}{N} \tag{55}$$

  where N is the number of samples, r_i and l_i are a predicted cluster label of x_i and the corresponding ground-true label, respectively. If x=y, then $\delta(x, y) = 1$, and 0 otherwise. map(r_i) represents the function of the optimal permutation mapping.
- **ARI - Adjusted Rand index (ARI)**: ARI is a performance evaluation indicator of the clustering model. A larger value indicates a better clustering result. ARI is calculated using the following four indicators: 1) A true positive (TP) represents true positive, 2) A true negative (TN) means true negative, 3) A false-positive (FP)

is false-positive, and 4) A false-negative (FN) is false-negative. ARI is formulated as:

$$ARI = \frac{2 \times (TP \cdot TN - FN \cdot FP)}{(TP + FN)(TN + FN) + (TP + FP)(FP + TN)} \tag{56}$$

- **F-scores**: The F-score integrates the recall and precision of a classifier into a single metric that compares the performances of two classifiers:

$$F - score = 2 \times \frac{precision \times recall}{precision + recall} \tag{57}$$

where $precision = \frac{TP}{TP + FP}$, and $recall = \frac{TP}{TP + FN}$.

4.5 Comparisons of the performance of clustering and discussion

Table 2 lists the evaluation scores of NMI, ACC, ARI, and F-score and the results of different IMC baseline methods and our proposed method on the five datasets with different missing rates. From this table, we can make the following important observations.

1) Based on the experiments, our proposed CCIM-SLR achieved the best performance compared with all other state-of-the-art methods. In particular, our method achieved 96.63% accuracy on the WebKB dataset with missing 10% samples. Compared to the proposed CCIM-SLR, IMSC-AGL obtained comparable results on the Yale dataset. However, CCIM-SLR still shows its superiority on other datasets with an increased missing rate.

2) Although UEAF exhibits good performance in terms of metric scores, it lacks robustness. The main reason for this is that performing well, UEAF must satisfy the condition that the feature dimensions of all views are larger than the cluster number. From this perspective, our method has stronger robustness in handling complex types of incomplete multiview.

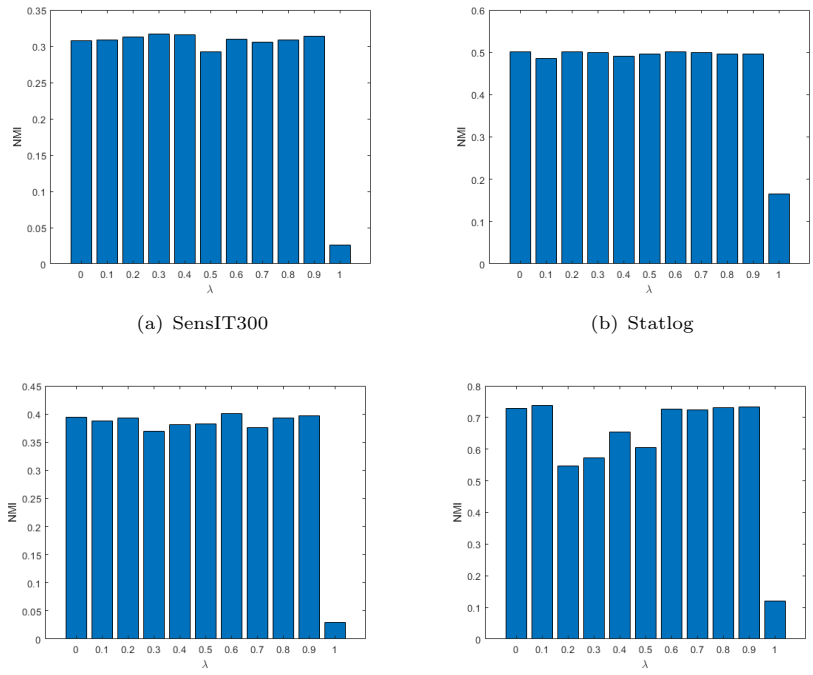
3) In general, DAIMC ignores padding for missing views. As a result, the achieved NMI score of our proposed CCIM-SLR is 28.03% higher than that of DAIMC in a case in which 50% of samples in the Webkb dataset are missing. CCIM-SLR uses an advanced filling mechanism. The experimental results show that this mechanism can result in better clustering performance.

4) Compared with IMC-GRMF, CCIM-SLR can maintain meaningful semantic relationships between the original view by building a consistent structure. In particular, clustering performance on the datasets with large differences in sample dimensions between Wisconsin and WebKB. CCIM-SLR can handle all kinds of incomplete data, which is much better than IMC-GRMF.

5) From Table 2, the performance of HCP-IMSC in terms of several metric scores is superior to other compared methods, which shows the advantage of capturing high-order correlation. In a case of an increasing missing rate, CCIM-SLR, however, can produce excellent incomplete multiview clustering results.

4.6 Parameter sensitivity of CCIM-SLR

Comparative experiments are performed in this section to investigate the sensitivity of each parameter of CCIM-SLR. Because the NMI index can objectively evaluate the accuracy of the comparison between a community division and the standard division, we use NMI to determine the range of parameters under satisfactory clustering results. Our main focus is on the following parameters in Equation (9): clustering model weight parameter λ, sparsity term parameter λ_1, low-rank term parameter λ_2, noise term parameter λ_3, flexible term parameter λ_4, and Lagrange operator control parameters μ and ρ.

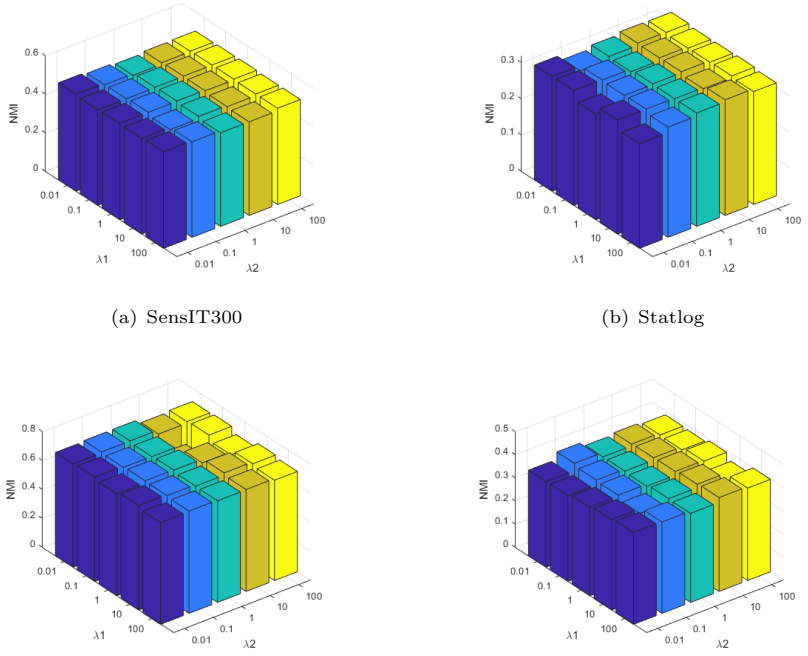


(a) SensIT300 (b) Statlog

(c) Wisconsin (d) WebKB

Fig. 4 NMI (%) versus parameter λ of the proposed CCIM-SLR on the (a) SensIT300 dataset with 10% incomplete samples of each view, (b) Statlog dataset with 10% incomplete samples of each view, (c) Wisconsin dataset with 10% incomplete samples of each view, and (d) WebKB dataset with 10% incomplete samples of each view.

1) Parameter λ: Fig. 4 shows the NMI (%) scores for different scales of λ parameters. Our proposed CCIM-SLR achieved satisfactory performances on the SensIT300, Statlog, Wisconsin, and WebKB datasets when λ was in the range of [0, 0.9], [0, 0.9], [0, 0.9], and [0.6, 0.9], respectively. Based on the results, the best values for the λ parameter should range between [0.6, 0.9]. These values are used in further experiments.



(a) SensIT300 (b) Statlog

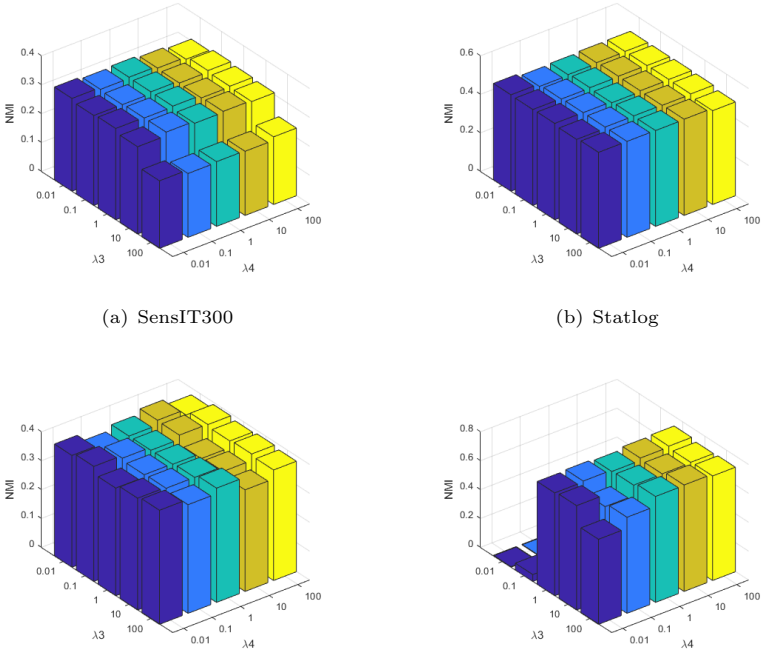
(c) WebKB (d) Wisconsin

Fig. 5 NMI (%) versus parameters λ_1 and λ_2 of the proposed CCIM-SLR on the (a) SensIT300 dataset with 10% incomplete samples of each view, (b) Statlog dataset with 10% incomplete samples of each view, (c) Wisconsin dataset with 10% incomplete samples of each view, and (d) WebKB dataset with 10% incomplete samples of each view.

2) Parameters λ_1 and λ_2: We explored the details of parameters λ_1 and λ_2 in Equation (9) by applying CCIM-SLR to the SensIT300, Starlog, Wisconsin, and WebKB datasets with a 10% incomplete-view rate, as shown in Fig. 5. In our analyses, the performance of our algorithm is shown to be insensitive to λ_1 and λ_2 parameters. As shown in Fig. 5, the values of the indicators of clustering do not change significantly as the values of the parameters change.

3) Parameters λ_3 and λ_4: We evaluated the NMI score for different values of λ_3 and λ_4 parameters in Equation (9) on the SensIT300, Statlog, Wisconsin, and WebKB datasets with an incomplete-view rate of 10%. Fig. 6 depicts the best clustering results. Based on the experiments, the most suitable values for the candidate parameters λ_3 and λ_4 range from [1,10] and [0.1,100], respectively.

4) Parameters μ and ρ: Fig. 7 shows the NMI versus μ and ρ parameters in Equation (45) on the SensIT300, Statlog, Wisconsin, and WebKB datasets with an incomplete-view rate of 10%. The experimental results show that if μ ranges from [0.5,10] and ρ from [1,1.1], our CCIM-SLR performed best.



(a) SensIT300 (b) Statlog

(c) Wisconsin (d) WebKB

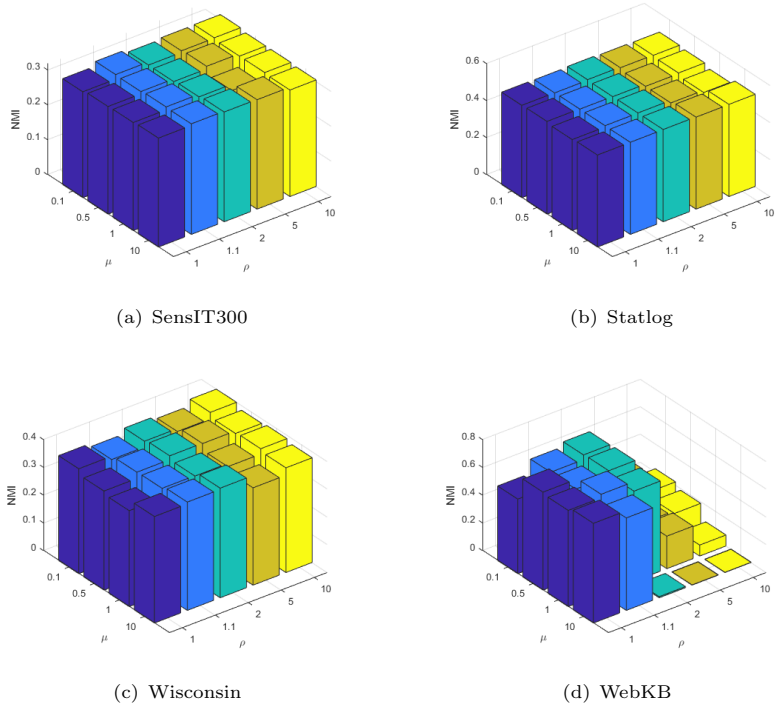
Fig. 6 NMI (%) versus parameters λ_3 and λ_4 of the proposed CCIM-SLR on the (a) SensIT300 dataset with 10% incomplete samples of each view; (b) Statlog dataset with 10% incomplete samples of each view; (c) Wisconsin dataset with 10% incomplete samples of each view; and (d) WebKB dataset with 10% incomplete samples of each view.

4.7 Ablation study

To investigate the impact of each component of CCIM-SLR on its overall performance, we performed two ablation experiments on the three datasets. Specifically, we removed the hidden view from the clustering part in Ablation 1. Ablation 2 replaced the adjustable Γ-norm with the traditional kernel norm to produce a low-rank representation. The experiment results are reported in Table 3. From this table, we can find that although the NMI score of Ablation 2 is 2.66% higher than that of CCIM-SLR at a missing rate of 30%, CCIM-SLR still performed well in all other cases. This indicates that the adjustable low-rank representation Γ-norm can obtain better filling performance than the traditional kernel norm. In addition to this, we can find that clustering with the removed hidden views produced the worst results, which indicates that the fusion results of multiview directly affect the clustering results.

4.8 Robustness experiments

To further demonstrate the robustness of the proposed CCIM-SLR algorithm, we conducted experiments on two datasets, SensIT300 and Statlog, with varying degrees



(a) SensIT300 (b) Statlog

(c) Wisconsin (d) WebKB

Fig. 7 NMI (%) versus parameters μ and ρ of the proposed CCIM-SLR on the (a) SensIT300 dataset with 10% incomplete samples of each view, (b) Statlog dataset with 10% incomplete samples of each view, (c) Wisconsin dataset with 10% incomplete samples of each view, and (d) WebKB dataset with 10% incomplete samples of each view.

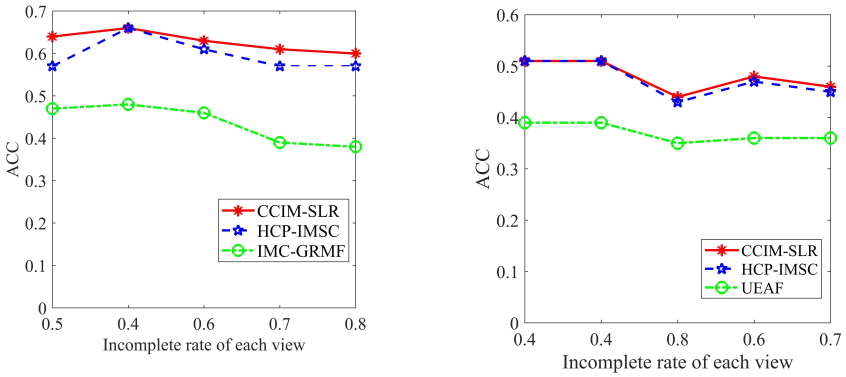
of missing data. Specifically, we set missing rates to (50%, 40%, 60%, 70%, and 80%) and (40%, 40%, 80%, 60%, and 70%) for SensIT300 and Statlog, respectively.

In Fig.8(a), we compare the clustering performance of CCIM-SLR, HCP-IMSC, and IMC-GRMF on the SensIT300 dataset across the different missing rates. Our proposed algorithm is the most stable, except for the 40% missing rate. Similarly, in Fig.8(b), we show the clustering performance of CCIM-SLR, HCP-IMSC, and UEAF on the Statlog dataset with the above-mentioned missing rates. Our proposed algorithm still performs the best, except for the 40% missing rate.

Overall, our experimental results demonstrate that CCIM-SLR is robust and performs well on datasets with varying degrees of missing data.

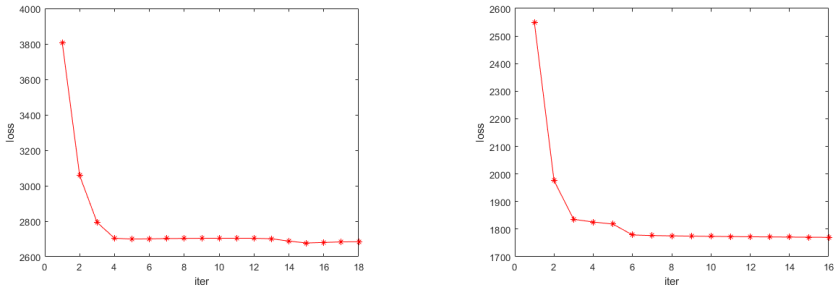
4.9 Experiments on the convergence of CCIM-SLR

As presented in Section 3.3, the objective function of CCIM-SLR is divided into several subproblems, with each subproblem being analytically solved. The CCIM-SLR algorithm adopts an alternating iterative optimization procedure. The objective function monotonically decreases until it converges. As shown in Fig. 9, the experimental results have demonstrated the correctness of the theoretical proofs in Section **??**.



(a) SensIT300 (b) Statlog

Fig. 8 Comparisons of robustness experiments on SensIT300 and Statlog Datasets.



(a) Statlog (b) WebKB

Fig. 9 Objective function values versus the iteration steps of CCIM-SLR on the (a) Statlog and (b) WebKB face databases, in which 10% samples are randomly selected as the paired samples.

4.10 The effectiveness of CCIM-SLR on the datasets

As shown in Table 4, the following presents the performance of two methods that utilize association information to constrain the reconstruction of missing view samples in terms of clustering metrics, execution time, and complexity. One employs hypergraph-induced hyper-Laplacian regularization, while the other focuses on intra-view and inter-view association information.

5 Conclusion

In reality, datasets that are collected often contain data samples with incomplete multiview, and the number of such samples varies significantly. This presents a challenge for clustering methods. To address this issue, a novel incomplete multiview clustering method called CCIM-SLR has been presented in this paper based on a sparse low-rank representation. In particular, CCIM-SLR measures the correlations between samples with the same views using sparse low-rank learning, while also capturing

the correlations between different views through shared hidden view learning. Moreover, the proposed method learns the shared hidden space, visible view, and cluster partition alternatively, thus avoiding the sensitivity of postprocessing methods like K-means to initial parameter values. This improves performance, as demonstrated through both theoretical proof and experimental comparison with advanced methods for IMVC on five representative datasets. The experimental results showed that our CCIM-SLR achieved good performance, especially on datasets with an increasing number of incomplete samples.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 62076047), and Dalian Innovation Fund (No. 2021JJ12SN44).

Conflict of Interest

We declare that we have no conflict of interest.

Database availability statement

The authors declare the availability of data and material.

References

[1] Nie F, Jing L, Li X (2017) Self-weighted multiview clustering with multiple graphs. In: Twenty-Sixth International Joint Conference on Artificial Intelligence(IJCAI), pp 2564–2570 https://doi.org/10.24963/ijcai.2017/357.

[2] Peng X, Huang Z, Lv J, Zhu H, Zhou J.T (2019) Comic: Multi-view clustering without parameter selection, In: International Conference on Machine Learning, pp 5092–5101

[3] Zhang Z, Liu L, Shen F, Shen H.T, Shao L (2018) Binary multi-view clustering. IEEE transactions on pattern analysis and machine intelligence 41(7): 1774–1782

[4] Zhao H, Ding Z, Fu Y (2017) Multi-view clustering via deep matrix factorization. In: Proceedings of the AAAI Conference on Artificial Intelligence vol, 31

[5] Chang J, Wang L, Meng G, Xiang S, Pan C (2017) Deep adaptive image clustering. In: Proceedings of the IEEE International Conference on Computer Vision, pp 5879–5887

[6] Liu G, Ge H, Su S Wang S (2022) Multi-view clustering via dual-norm and hsic, Multimedia Tools and Applications, 1–20

[7] Li S.-Y, Jiang Y, Zhou Z.-H (2014) Partial multi-view clustering, In: Proceedings of the AAAI Conference on Artificial Intelligence vol, 28

[8] Yin Q.,Wu S.,Wang L (2015) Incomplete multi-view clustering via subspace learning. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, pp 383–392

[9] Zhao H, Liu H, Fu Y (2016) Incomplete multi-modal visual data grouping, In: IJCAI, pp 2392–2398

[10] Wen J, Wu Z, Zhang Z, Fei L, Zhang B, Xu Y (2021) Structural deep incomplete multi-view clustering network, In: Proceedings of the 30th ACM International Conference on Information & Knowledge Management, pp 3538–3542

[11] Guo J, Ye J (2019) Anchors bring ease: An embarrassingly simple approach to partial multi-view clustering, In: Proceedings of the AAAI Conference on Artificial Intelligence vol, 33, pp 118–125

[12] Hu M, Chen S (2019) Doubly aligned incomplete multi-view clustering, arXiv preprint arXiv:1903.02785

[13] Wen J, Zhang Z, Zhang Z, Fei L, Wang M (2020) Generalized incomplete multiview clustering with flexible locality structure diffusion, IEEE transactions on cybernetics 51(1): 101–114

[14] Wen J, Zhang Z, Xu Y, Zhang B, Fei L, Liu H (2019) Unified embedding alignment with missing views inferring for incomplete multi-view clustering. In: Proceedings of the AAAI Conference on Artificial Intelligence,vol, 33, pp 5393–5400

[15] Wen J, Zhang Z, Zhang Z, Zhu L, Fei L, Zhang B, Xu Y (2021) Unified tensor framework for incomplete multi-view clustering and missing-view inferring, In: Proceedings of the AAAI Conference on Artificial Intelligence vol, 35, pp 10273–10281

[16] Liang N, Yang Z, Li Z, Han W (2022) Incomplete multi-view clustering with incomplete graph-regularized orthogonal non-negative matrix factorization, Applied Intelligence 52(13): 14607–14623

[17] Yin J, Sun S (2022) Incomplete multi-view clustering with cosine similarity, Pattern Recognition 123: 108371

[18] Xu C, Guan Z, Zhao W, Wu H, Niu Y, Ling B (2019) Adversarial incomplete multi-view clustering, In: IJCAI vol, 7, pp 3933–3939

[19] Xu J, Li C, Ren Y, Peng L, Mo Y, Shi X, Zhu X (2022) Deep incomplete multi-view clustering via mining cluster complementarity, In: Proceedings of the AAAI Conference on Artificial Intelligence vol, 36, pp 8761–8769

[20] Wang Q, Ding Z, Tao Z, Gao Q, Fu Y (2021) Generative partial multi-view clustering with adaptive fusion and cycle consistency, IEEE Transactions on Image

Processing 30: 1771–1783

[21] Zheng X, Liu X, Chen J, Zhu E (2022) Adaptive partial graph learning and fusion for incomplete multi-view clustering, International Journal of Intelligent Systems 37(1): 991–1009

[22] Wang Q, Ding Z, Tao Z, Gao Q, Fu Y (2018) Partial multi-view clustering via consistent gan, In: 2018 IEEE International Conference on Data Mining (ICDM), pp 1290–1295

[23] Yang W, Shi Y, Gao Y, Wang L, Yang M (2018) Incomplete-data oriented multiview dimension reduction via sparse low-rank representation, IEEE transactions on neural networks and learning systems 29(12): 6276–6291

[24] Fang X, Hu Y, Zhou P, Wu D.O (2021) Unbalanced incomplete multi-view clustering via the scheme of view evolution: Weak views are meat; strong views do eat, IEEE Transactions on Emerging Topics in Computational Intelligence 6(4): 913–927

[25] Deng Z, Liu R, Xu P, Choi K.-S, Zhang W, Tian X, Zhang T, Liang, L, Qin B, Wang S (2020) Multi-view clustering with the cooperation of visible and hidden views, IEEE Transactions on Knowledge and Data Engineering 34(2): 803–815

[26] El Hajjar S, Dornaika F, Abdallah F (2022) One-step multi-view spectral clustering with cluster label correlation graph, Information Sciences 592: 97–111

[27] Niu G, Yang Y, Sun L (2021) One-step multi-view subspace clustering with incomplete views, Neurocomputing 438: 290–301

[28] Niu X, Fu L, Zhang W, Li Y (2021) Seismic data interpolation based on simultaneously sparse and low-rank matrix recovery, IEEE Transactions on Geoscience and Remote Sensing 60: 1–13

[29] Wang L, Wang B, Zhang Z, Ye Q, Fu L, Liu G, Wang M (2019) Robust autoweighted projective low-rank and sparse recovery for visual representation, Neural Networks 117: 201–215

[30] Du H, Zhang X, Hu Q, Hou Y (2015) Sparse representation-based robust face recognition by graph regularized low-rank sparse representation recovery, Neurocomputing 164: 220–229

[31] Kang Z, Peng C, Cheng Q (2015) Robust pca via nonconvex rank approximation, In: 2015 IEEE International Conference on Data Mining, pp 211–220

[32] Richard E, Savalle P.-A, Vayatis N (2012) Estimation of simultaneously sparse and low rank matrices, arXiv preprint arXiv:1206.6474

[33] Wen J, Fang X, Cui J, Fei L, Yan K, Chen Y, Xu Y (2018) Robust sparse linear discriminant analysis, IEEE Transactions on Circuits and Systems for Video Technology 29(2): 390–403

[34] Tao P.D, An L.H (1997) Convex analysis approach to dc programming: theory, algorithms and applications, Acta mathematica vietnamica 22(1): 289–355

[35] Zhang W, Wen Y, Guan K, Kilper D, Luo H, Wu D.O (2013) Energy-optimal mobile cloud computing under stochastic wireless channel, IEEE Transactions on Wireless Communications 12(9): 4569–4581

[36] Duarte M.F, Hu Y.H (2004) Vehicle classification in distributed sensor networks. Journal of Parallel and Distributed Computing 64(7): 826–838

[37] Qin Y.P, Qin P.D, Wang Y, Lun S.X (2013) A new optimal binary tree svm multiclass classification algorithm, In: Applied Mechanics and Materials vol, 373, pp 1085–1088

[38] Liu J.-W, Wang Y.-F, Lu R.-K, Luo X.-L (2020) Multi-view non-negative matrix factorization discriminant learning via cross entropy loss, In: 2020 Chinese Control And Decision Conference (CCDC), pp 3964–3971

[39] Craven M, McCallum A, PiPasquo D, Mitchell T, Freitag D (1998) Learning to extract symbolic knowledge from the world wide web, Technical report Carnegiemellon univ pittsburgh pa school of computer Science

[40] Sharif M, Mohsin S, Jamal M.J, Raza M (2010) Illumination normalization preprocessing for face recognition, In: Environmental Science and Information Application Technology (ESIAT) 2010 International Conference

[41] Jie W, Zheng Z, Yong X, Zhong Z (2019) Incomplete Multi-view Clustering Via Graph Regularized Matrix Factorization: Munich Germany September 8-14, 2018 Proceedings Part IV, Computer Vision – ECCV 2018 Workshops, pp 593–608

[42] Wen J, Xu Y, Liu H (2018) Incomplete multiview spectral clustering with adaptive graph learning, IEEE transactions on cybernetics 50(4): 1418–1429

[43] Li Z, Tang C, Zheng X, Liu X, Zhang W, Zhu E (2022) High-order correlation preserved incomplete multi-view subspace clustering, IEEE Transactions on Image Processing 31: 2067–2080

[44] Est´evez P.A, Tesmer M, Perez C.A, Zurada J.M (2009) Normalized mutual information feature selection, IEEE Transactions on neural networks 20(2): 189–201

[45] Cai D, He X, Han J (2005) Document clustering using locality preserving indexing. IEEE Transactions on Knowledge and Data Engineering 17(12): 1624–1637

[46] Romano S, Vinh N.X, Bailey J, Verspoor K (2016) Adjusting for chance clustering comparison measures, The Journal of Machine Learning Research 17(1): 4635–4666

[47] Wang Y, Li J, Li Y, Wang R, Yang X (2015) Confidence interval for f1 measure of algorithm performance based on blocked 3×2 cross-validation, IEEE Transactions on Knowledge and Data Engineering 27(3): 651–659