



Incomplete multi-view clustering based on hypergraph

Jin Chen ^a✉, Huafu Xu ^b✉, Jingjing Xue ^a✉, Quanxue Gao ^a✉, Cheng Deng ^c✉, Ziyu Lv ^d✉

Show more ▾

Outline | Share | Cite

<https://doi.org/10.1016/j.inffus.2024.102804>

[Get rights and content](#)

Full text access

Highlights

- Propose a cohesive model to solve missing data, learn representation and clustering.
- A hypergraph is used to explore the data structure to reconstruct the missing parts.
- Non-negative matrix factorization to equate K-means for clustering in one step.
- Tensor Schatten p -norm captures the complementary information in different views.



Abstract

The graph-based incomplete multi-view clustering **aims** at integrating information from multiple views and utilizes graph models to capture the global and local structure of the data for reconstructing missing data, which is suitable for processing complex data. **However**, ordinary graph learning methods usually only consider pairwise relationships between data points and cannot unearth **higher-order relationships latent in the data**. And existing graph clustering methods often divide the process of learning the representations and the clustering process into two separate steps, which may lead to unsatisfactory clustering results. Besides, they also tend to consider only intra-view similarity structures and overlook inter-view ones. To this end, this paper introduces an innovative **one-step incomplete multi-view clustering based on hypergraph (IMVC_HG)**. Specifically, we use a hypergraph to reconstruct missing views, which can better explore the local structure and higher-order information between sample points. Moreover, we use non-negative matrix factorization with orthogonality constraints to equate K-means, which eliminates post-processing operations and avoids the problem of suboptimal results caused by the two-step approach. In addition, the tensor Schatten p -norm is



used to better capture the complementary information and low-rank structure between the cluster label matrices of multiple views. Numerous experiments verify the superiority of IMVC_HG.



Keywords

Incomplete multi-view clustering; Graph learning; Hypergraph; Non-negative matrix factorization

1. Introduction

In recent years, the rapid development across major fields has precipitated the emergence of complex data in diverse formats, including text, audio, and video. It has become a major challenge to process the multifaceted data to unearth the hidden usable information. Clustering, a pivotal unsupervised learning technique, has attracted significant attention within the machine learning community for its cost-effectiveness and its ability to mitigate the need for manual labeling [1]. And many techniques have been used to solve the clustering problem [2], [3], [4], [5], including various meta-heuristic algorithms, such as artificial bee colony (ABC) algorithm, interactive autodidactic school (IAS), and artificial gorilla troop optimization (AGTO), etc.

Real-world applications such as medicine, transportation and education frequently **encompass such multi-view data, wherein a single entity is depicted through various data modalities**. Multi-view clustering transcends the traditional single-view approach by effectively amalgamating data from multiple perspectives to distill **consistent information**. Nevertheless, disturbed by many factors, the acquired data is often incomplete, which will lead to the destruction of the natural alignment properties between multiple views. Thus, the study of incomplete multi-view clustering is of great importance.



It is worth noting that graph-based approaches have risen to prominence within incomplete multi-view clustering. **These methods excel at capturing the intrinsic geometric structure of data, offering a compelling direction for research into effectively analyzing incomplete multi-view datasets**. They predominantly involve constructing similarity graphs from data to uncover relationships between missing data in views, effectively harnessing the data's geometric structure [6], [7], [8], [9], [10], [11]. While these approaches enhance clustering outcomes, they primarily address pairwise data point correlations. As data point interrelations are often complex, surpassing mere pairings in numerous applications, it is especially important to explore higher-order information between data samples and local structural information in the view. Notably, most graph-based clustering algorithms initially learn an adaptive nearest neighbor graph, a subspace graph, or a depth graph [12], [13], [14], and then apply spectral clustering or other clustering methods to obtain the final clustering results. Such methods of separating the process of learning the representation from the clustering process may result in initial representations inadequately suited for subsequent clustering. Moreover, some existing methods neglect the inter-view relationships, thereby failing to exploit the complementary information and low-rank structures across different views [15], [16], [17], [18]. For multi-view clustering, fully leveraging this inter-view complementarity and low-rank structure is imperative.



To address the challenges above, this paper concentrates on exploiting graph structures for uncovering higher-order relationships among data samples, integrating the learning of representations with the clustering process, and effectively considering the complementary information and low-rank structure across multiple views. Considering that hypergraphs have more complex structural, nodal, and edge attributes information than traditional graphs, we propose a one-step incomplete multi-view clustering algorithm that combines the use of hypergraphs to reconstruct views and orthogonal non-negative matrix factorization. Specifically, our



approach begins by generating individual affinity matrices for each view through self-representation and then deriving a consistent affinity matrix. Moreover, constructing a hypergraph from this matrix captures extensive higher-order correlation and the local structure, which is instrumental in reconstructing missing data. Besides, non-negative matrix factorization with orthogonal constraint is applied to each view to yield the respective clustering label matrices. When they are assembled into a tensor, the tensor Schatten p -norm is used to reveal the low-rank structure and complementary information of different label matrices. This methodology obtains a final clustering label matrix by weighting the label matrices corresponding to different views.

The principal contributions of this study are summarized as follows.

- We use a consistent affinity matrix to construct a hypergraph and use the hypergraph to reconstruct the missing data in the view. By considering the similarity relationship between points on the same hyperedge and their nearest neighbors, we can better explore the local structure and higher-order information of the view to better recover the missing parts.
- The non-negative matrix factorization with orthogonal constraints is equivalent to K-means approach, which simplifies the clustering process and eliminates the post-processing operation of the two-step strategy.
- Considering the tensor Schatten p -norm for the tensor composed of the clustering label matrices of different views better exploits the complementary information and low-rank structure among these views. All these help to construct a joint model that simultaneously handles the missing data reconstruction, representation learning and clustering problems, thereby improving the overall clustering performance of the model.

Notations. In this paper, we use bold calligraphic letters to denote third-order tensors, e.g., $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, bold uppercase letters to denote matrices, e.g., \mathbf{Z} , bold lowercase letters to denote vectors, e.g., \mathbf{z} , and italic uppercase letters such as Z_{ij} for the entries of \mathbf{Z} . The trace of matrix \mathbf{Z} is expressed as $\text{tr}(\mathbf{Z})$. The i th frontal slice of \mathcal{X} is $\mathcal{X}^{(i)}$. $\overline{\mathcal{X}}$ represents the discrete Fast Fourier Transform (FFT) of \mathcal{X} along the third dimension, i.e., $\overline{\mathcal{X}} = \text{fft}(\mathcal{X}, [], 3)$. Thus, $\mathcal{X} = \text{ifft}(\overline{\mathcal{X}}, [], 3)$.

2. Related works

In recent years, many incomplete multi-view clustering methods have been proposed. This section mainly reviews two types: matrix factorization-based and graph-based incomplete multi-view clustering. Moreover, the relevant knowledge of hypergraph and tensor is also introduced in this section.

2.1. Incomplete multi-view clustering based on matrix factorization

The methodologies in question generally entail a multi-stage process: initially, filling in the gaps within incomplete multi-view data, then restoring matrices via matrix decomposition to extract the fundamental and representative matrices. This is followed by learning consistency to secure coherent representative matrices, and it concludes with the application of a clustering algorithm to generate the final outcomes.

To solve the problem of missing data, a foundational model in this domain, termed Partial Multi-View Clustering (PVC), was devised by Li et al. [19]. PVC harnesses non-negative matrix decomposition to distill low-dimensional feature representations from incomplete and complete multi-view data, employing fully observed instances to compensate for those with missing views. Progressing from PVC, Zhao et al. [20], and Xu et al. [21] amalgamated graph embedding techniques. Through the imposition of varied constraints, they delved into the data's geometric structure, which yielded enhancements in efficacy. Nonetheless, these approaches presuppose a particular alignment in incomplete multi-view data samples (as depicted in Fig.

1(b)) and do not extend to situations characterized by randomly distributed missing view information (illustrated in Fig. 1(a)). Besides, they only work when there are two views, but not for more than two views, and do not fully consider the importance of different views.

In order to solve the problem of random missing data in all views, expand the number of views that the algorithm can handle and fully consider the importance of different views, recent advances have concentrated on weighted NMF-based methods for incomplete multi-view clustering. Shao et al. [22] introduced the MIC approach, utilizing a weighted NMF algorithm to discern latent feature matrices for each view. This method aims to minimize the divergence between these matrices and a unified matrix, modulate the weighting of views according to the fraction of missing instances per view, and incorporate an $\ell_{2,1}$ regularization to bolster robustness against noise and outliers. Moreover, methods like Doubly Aligned Incomplete Multi-View Clustering (DAIMC) [23] and Graph Regularized Partial Multi-View Clustering (GPMVC) [24] have probed additional dimensions of data through orthogonal or graph constraints, thereby augmenting their performance. However, these methods are usually accompanied by high computational complexity and cannot be used to process large-scale data. Furthermore, in order to reduce the computational complexity, some online methods have been proposed. Online Multi-View Clustering with Incomplete Views (OMVC) [25] and One-Pass Incomplete Multi-View Clustering (OPIMC) [26] present improvements in efficiency and memory usage courtesy of their block-optimized decomposition framework. However, since online methods usually do not have enough historical data to handle abnormal situations, they may be sensitive to noise and outliers. Table 1 summarizes the above IMVC methods based on matrix factorization.



[Download: Download high-res image \(151KB\)](#)

[Download: Download full-size image](#)

Fig. 1. Missing types of incomplete multi-view data.

Table 1. Summary of IMVC methods based on matrix factorization.

Methods	Advantages	Disadvantages
PVC [19]	Fast convergence speed.	Only suitable for two views.
IMG [20]	Introducing a new graph Laplacian term.	Only suitable for two views.
MIC [22]	Using weighted NMF to consider the importance of different views; Suitable for more than two views.	Not suitable for large-scale data.
DAIMC [23]	Using weighted Semi-NMF; Suitable for more than two views.	Not suitable for large-scale data.
GPMVC [24]	Using a new weighted view-specific graph Laplacian regularization; Suitable for more than two views.	Not suitable for large-scale data.
OMVC [25]	Weighting each view; Reducing memory requirements for large-scale data.	Allowing only one pass through the data and may not fully utilize all data.
OPIMC [26]	Weighting each view; Reducing memory requirements; Getting clustering results directly.	Cannot exploit the information of the entire dataset like offline methods.

Table 2. Summary of IMVC methods based on graphs.

Methods	Advantages	Disadvantages
SCIMC [27]	Spectral clustering helps capture global structure.	Strong initialization dependency.
PIC [28]	Reducing the risk of spectral perturbations by optimizing the similarity matrix.	Decoupling graph construction from the overall process; Sensitive to the initial graph.
IMSC_AGL [16]	Integrating subspace learning and graph construction process.	Ignoring the importance of different views; High computational complexity.
IMVTSC-MVI [29]	Integrating inference of missing views and learning similarity graphs.	Ignoring the importance of different views; High computational complexity.
HCP_IMSC [30]	Applying hypergraph theory to reconstruction of missing views.	Ignoring the importance of different views; High computational complexity.
BGIMVSC [31]	Introducing a weighted multi-view learning mechanism.	High computational complexity.
IMVC-CBG [32]	Weighting each view; Using bipartite graph to improve computational efficiency.	Requiring additional clustering process.
OGIMC [33]	Weighting each view; Getting clustering results directly.	High computational complexity.

2.2. Incomplete multi-view clustering based on graphs

In contrast to other incomplete multi-view clustering methodologies, graph-based strategies excel in handling data with more intricate distributions and effectively leveraging the geometric structure of incomplete views. These methods endeavor to derive consensus graphs or representations from incomplete graphs learned from the incomplete views. Many preceding studies predicated their techniques on pre-established incomplete graphs, wherein absent elements were assigned zero or imputed with the mean value. Gao et al. [27] introduced Spectral clustering-based incomplete multi-view clustering (SCIMC), which retrieves potential representations by leveraging information from alternative views and subsequently attains a unified representation by minimizing the discrepancies across different views. Wang et al. [28] presented a spectral perturbation model for incomplete multi-view data (PIC), where rows aligned with missing instances are populated with the average of rows from accessible views. This perturbation-focused model learns coefficients for graph fusion and attains a unified graph by fusing multiple graphs with these coefficients. However, methods that decouple graph construction from the overall learning process can be highly susceptible to the initial graph configurations, thus posing practical difficulties.

To remedy this, Wen et al. [16] proposed incomplete multi-view spectral clustering with adaptive graph learning (IMSC_AGL). This approach amalgamates adaptive graph construction with consensus representation learning into a cohesive optimization framework. It commences with the adaptive construction of graphs for each view using a low-dimensional representation, followed by the attainment of low-dimensional representations through spectral constraints inspired by spectral clustering, and subsequently acquires consensus potential representations of all views via concordance terms, finally applying K-means clustering to these consensus representations. IMVTSC-MVI [29] amalgamates the inference of missing views and the learning of similarity graphs into a unified framework, allowing these dual tasks to mutually reinforce and efficiently unravel the concealed information within missing views. Furthermore, in order to enhance the reconstruction capability of missing views, HCP_IMSC [30] introduces hypergraph theory to reconstruct incomplete data to preserve the high-order geometric structure of incomplete view data. Samples with

missing views are restricted to be reconstructed by their neighboring samples. Nevertheless, different views contribute differently to the clustering task, and these models cannot adaptively select informative views for subsequent clustering.

Therefore, Sun et al. [31] proposed BGIMVSC, which introduces a weighted multi-view learning mechanism to automatically balance the role of different views in model optimization. In this way, the intrinsic information of different views can be fully mined. These methods, while powerful, are memory-intensive and grapple with the challenges posed by large-scale, incomplete multi-view datasets. To address this, Wang et al. [32] devised a highly-efficient technique tailored for large-scale incomplete multi-view clustering, named IMVC-CBG. This technique orchestrates multi-view anchor learning with incomplete bipartite graph learning within a singular framework, tackling the challenges of incomplete multi-view clustering by employing bipartite graphs and ensuring linear complexity about the number of instances, hence suitable for large-scale applications. Moreover, most algorithms detach the learning of representations from the clustering process, and the graph or latent representation acquired in the initial phase may not fully serve the following clustering phase. Hence, Zhou et al. [33] proposed OGIMC, which imposes a rank constraint on the Laplacian matrix of the fused graph so that the number of connected components is equal to the required number of clusters, and the clustering results can be obtained simultaneously while generating a unified graph. We give a summary of these IMVC methods based on graphs in Table 2.

Taking the above issues into full consideration, our proposed algorithm employs non-negative matrix factorization with orthogonal constraints to parallel K-means, thus bypassing the post-processing stage inherent in the “two-step” method. It uses a hypergraph to reconstitute the missing views, thereby more adeptly capitalizing on the local structure and higher-order information inherent in the views. Additionally, applying the tensor Schatten p -norm further assists in duly considering the complementary information and low-rank structure among the clustering label matrices of each view.

2.3. Hypergraph

For a hypergraph [34] $\mathbf{G} = (\mathbf{V}, \mathbf{E}, \mathbf{W})$, where $\mathbf{V} = \{v_1, v_2, \dots, v_n\}$ denotes the vertex set consisting of n vertices and $\mathbf{E} = \{e_1, e_2, \dots, e_m\}$ is the set consisting of m hyperedges, the weight of hyperedge e_i is denoted by $w(e_i)$. The incidence matrix $\mathbf{H}^{n \times m}$ of a hypergraph is defined as

$$\mathbf{H}_{ij} = \begin{cases} 1 & \text{if } v_i \in e_j, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Given an affinity matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$, where each vertex and the K nearest neighbors centered on it form a hyperedge, the incidence matrix can be obtained by

$$\mathbf{H}_{ij} = \begin{cases} S_{ij} & \text{if } v_i \in e_j, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

In the weight matrix \mathbf{W} , the weight $w(e_i)$ of each hyperedge is

$$w(e_i) = \sum_{v_i \in e_j} S_{ij}. \quad (3)$$

In the degree matrix \mathbf{D}_V of the vertices, the degree $d(v_i)$ of each vertex is

$$d(v_i) = \sum_{e_j \in \mathbf{E}} w(e_j) h(v_i, e_j). \quad (4)$$

In the degree matrix \mathbf{D}_E of the hyperedges, the degree $d(e_j)$ of each hyperedge is

$$d(e_j) = \sum_{v_i \in \mathbf{V}} h(v_i, e_j). \quad (5)$$

Then, the super Laplacian matrix can be obtained by

$$\mathbf{L}_h = \mathbf{D}_V - \mathbf{H}\mathbf{W}\mathbf{D}_E^{-1}\mathbf{H}^\top. \quad (6)$$

2.4. Tensor Schatten p -norm

Given a third-order tensor $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the tensor Schatten p -norm [35] of tensor \mathcal{Z} is defined as

$$\|\mathcal{Z}\|_{\mathfrak{S}^p} = \left(\sum_{i=1}^{n_3} \|\bar{\mathcal{Z}}^{(i)}\|_{\mathfrak{S}^p}^p \right)^{\frac{1}{p}} = \left(\sum_{i=1}^{n_3} \sum_{j=1}^{\min(n_1, n_2)} \sigma_j(\bar{\mathcal{Z}}^{(i)})^p \right)^{\frac{1}{p}}, \quad (7)$$

where $p \in (0, 1]$. $\sigma_j(\bar{\mathcal{Z}}^{(i)})$ denotes the j th singular value of $\bar{\mathcal{Z}}^{(i)}$. When $p = 1$, the tensor Schatten p -norm of \mathcal{Z} is equivalent to the tensor nuclear norm, i.e.,

$$\|\mathcal{Z}\|_{\mathfrak{S}^1} = \|\mathcal{Z}\|_* = \sum_{i=1}^{n_3} \sum_{j=1}^{\min(n_1, n_2)} \sigma_j(\bar{\mathcal{Z}}^{(i)}). \quad (8)$$

When $p \rightarrow 0$, $\sigma_j(\bar{\mathcal{Z}}^{(i)})^p \rightarrow 1$. Thus, the tensor Schatten p -norm actually calculates the number of singular values that are not equal to 0, that is, the rank of the tensor, i.e.,

$$\lim_{p \rightarrow 0} \|\mathcal{Z}\|_{\mathfrak{S}^p} = \text{rank}(\mathcal{Z}). \quad (9)$$

Therefore, it should be pointed out that for $0 < p \leq 1$, the Schatten p -norm is introduced for rank approximation, which approximates the rank function better when p is well chosen. In this paper, we impose a minimization constraint on the tensor Schatten p -norm to obtain a spatial low-rank structure, thereby fully exploiting the complementary information between label matrices of different views.

3. Proposed model

In this section, we first introduce the motivation of the proposed model and the step-by-step construction of the objective function in the first subsection. The detailed solution process and algorithm flow for the objective function are given in the second subsection. And in the last subsection, we give the computational complexity analysis of our algorithm.

3.1. Motivation and objective function

Existing multi-view clustering methods are usually based on data completeness. However, due to interference from many factors, the acquired data often has information missing in real world. Therefore, we divide one view of a multi-view dataset $\{\mathbf{X}^{(v)}\}_{v=1}^V$ into an observable part $\mathbf{X}_o^{(v)} \in \mathbb{R}^{d_v \times n_v}$ and a missing part $\mathbf{E}^{(v)} \in \mathbb{R}^{d_v \times (n - n_v)}$, where n is the total number of samples contained in the dataset, d_v and n_v are the number of features and the number of observable samples corresponding to the v th view, respectively. Each view's data is represented as

$$\mathbf{X}^{(v)} = \mathbf{X}_o^{(v)} \mathbf{W}_1^{(v)} + \mathbf{E}^{(v)} \mathbf{W}_2^{(v)}, \quad (10)$$

where the projection matrices $\mathbf{W}_1^{(v)} \in \mathbb{R}^{n_v \times n}$ and $\mathbf{W}_2^{(v)} \in \mathbb{R}^{(n - n_v) \times n}$ combine $\mathbf{X}_o^{(v)}$ and $\mathbf{E}^{(v)}$ to form a complete feature matrix. And they are defined as

$$\begin{aligned} \mathbf{W}_1^{(v)}_{ij} &= \begin{cases} 1, & \text{if } \mathbf{x}_o^{(v)}{}_i \text{ corresponds to } \mathbf{x}^{(v)}{}_j, \\ 0, & \text{otherwise,} \end{cases} \\ \mathbf{W}_2^{(v)}_{ij} &= \begin{cases} 1, & \text{if } \mathbf{e}^{(v)}{}_i \text{ corresponds to } \mathbf{x}^{(v)}{}_j, \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \quad (11)$$



where $\mathbf{x}_o^{(v)}_i$ and $\mathbf{e}^{(v)}_i$ denote the i th column vector of $\mathbf{X}_o^{(v)}$ and $\mathbf{E}^{(v)}$ respectively. $\mathbf{x}^{(v)}_j$ denotes the j th column vector of $\mathbf{X}^{(v)}$.

The objective function for incomplete multi-view subspace clustering is given by

$$\begin{aligned} \min_{\mathbf{E}^{(v)}, \mathbf{Z}^{(v)}} \sum_{v=1}^V \|\mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)}\|_F^2 + \Omega(\mathbf{Z}^{(v)}), \\ \text{s.t. } \mathbf{Z}^{(v)}_{ii} = 0, \mathbf{Z}^{(v)}_{ij} \geq 0, \mathbf{Z}^{(v)} = (\mathbf{Z}^{(v)})^T, \end{aligned} \quad (12)$$

where $\Omega(\cdot)$ denotes the regularization term. The condition $\mathbf{Z}^{(v)}_{ij} \geq 0$ encourages positivity within the same subspace while suppressing data from different subspaces. The symmetry constraint $\mathbf{Z}^{(v)} = (\mathbf{Z}^{(v)})^T$ eliminates double-counting operations.

Graph learning can obtain richer information about geometric structures, with hypergraphs containing even more information than ordinary graphs. Each edge of a traditional graph is connected by two nodes, i.e., an ordinary graph only describes the similarity between two points. While hypergraphs have at least two nodes in each hyperedge, i.e., hypergraphs can take into account more higher-order relationships and local structures between nodes, which makes hypergraphs have more complex structural, nodal, and edge attributes information than traditional graphs [30]. Therefore, we fuse the affinity matrices of different views to form a consistent affinity matrix and build a hypergraph based on this matrix to incorporate high-order information between different views. We use the GSPBOX tool to build the k -nearest neighbors hypergraph [36]. Hyper-Laplacian regularization is applied to constrain missing samples to be reconstructed in their neighborhoods. The regularization is expressed as

$$\min_{\mathbf{E}^{(v)}} \frac{1}{2} \sum_{e \in \mathbf{E}_{\text{hyper}}} \sum_{(i,j) \in e} \frac{\omega(e)}{d(e)} \|\mathbf{e}_i^{(v)} - \mathbf{x}_j^{(v)}\|^2, \quad (13)$$

where $\omega(e)$ is the weight of the hyperedge, $d(e)$ is the degree of the hyperedge, $\mathbf{E}_{\text{hyper}}$ is the set of all hyperedges. This can further be expressed as

$$\min_{\mathbf{E}^{(v)}} \text{Tr}(\mathbf{X}^{(v)} \mathbf{L}_{\text{hyper}} \mathbf{X}^{(v)T}), \quad (14)$$

where $\mathbf{L}_{\text{hyper}}$ is a hyper-Laplacian matrix constructed from the consistent affinity matrix \mathbf{Z}^* . The consistent affinity matrix \mathbf{Z}^* is obtained by

$$\min_{\mathbf{Z}^*} \|\mathbf{Z}^{(v)} - \mathbf{Z}^*\|_F^2. \quad (15)$$

To address the issue of suboptimal clustering results from the “two-step method”, non-negative matrix factorization with orthogonality constraints is employed to equate K-means [37], unifying the representation learning and clustering processes into a single framework. This approach eliminates post-processing and improves clustering performance. The relevant model is

$$\begin{aligned} \min_{\mathbf{F}^{(v)}, \mathbf{G}^{(v)}} \sum_{v=1}^V \|\mathbf{Z}^{(v)} - \mathbf{F}^{(v)} \mathbf{G}^{(v)T}\|_F^2 + \|\mathbf{F}\|_{\mathbb{S}^p}^p, \\ \text{s.t. } \mathbf{F}^{(v)} \geq 0, \mathbf{G}^{(v)T} \mathbf{G}^{(v)} = \mathbf{I}, \end{aligned} \quad (16)$$

where $\mathbf{F}^{(v)} \in \mathbb{R}^{n \times c}$ is the v th matrix corresponding to the third-order tensor, which is subsequently rotated to the tensor \mathcal{F} .

Remark 1

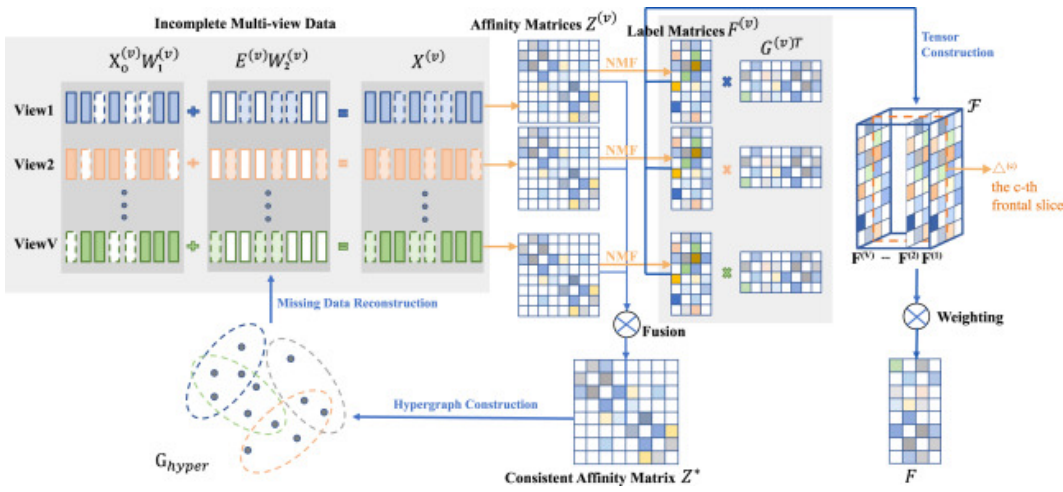
The tensor construction process is illustrated in Fig. 2, for tensor $\mathcal{F} \in \mathbb{R}^{n \times V \times c}$, the c th frontal slice $\Delta^{(c)}$ describes the similarity between n sample points and the c clusters in different views. Ideally, the label

matrices $\mathbf{F}^{(v)}$ are as similar as possible. But different views often show different cluster structures, imposing tensor Schatten p -norm minimization constraint on \mathcal{F} can make sure $\Delta^{(c)}$ has spatial low-rank structure [38], which helps exploit the complementary information embedded in inter-views.

Combining the above objective functions, the final model is given by

$$\begin{aligned}
 & \min_{\mathbf{E}^{(v)}, \mathbf{Z}^{(v)}, \mathbf{F}^{(v)}, \mathbf{G}^{(v)}} \sum_{v=1}^V \frac{1}{\alpha_v} \left(\lambda_1 \left\| \mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} \right\|_F^2 \right. \\
 & \quad + \lambda_2 \text{Tr}(\mathbf{X}^{(v)} \mathbf{L}_{\text{hyper}} \mathbf{X}^{(v)T}) + \left\| \mathbf{Z}^{(v)} - \mathbf{Z}^* \right\|_F^2 \\
 & \quad \left. + \lambda_3 \left\| \mathbf{Z}^{(v)} - \mathbf{F}^{(v)} \mathbf{G}^{(v)T} \right\|_F^2 \right) + \left\| \mathcal{F} \right\|_{\mathcal{S}\mathcal{P}}^p, \\
 & \text{s.t. } \mathbf{X}^{(v)} = \mathbf{X}_o^{(v)} \mathbf{W}_1^{(v)} + \mathbf{E}^{(v)} \mathbf{W}_2^{(v)}, \mathbf{Z}_{ii}^{(v)} = 0, \mathbf{Z}_{ij}^{(v)} \geq 0, \\
 & \quad \mathbf{Z}^{(v)} = \mathbf{Z}^{(v)T}, \mathbf{F}^{(v)} \geq 0, \mathbf{G}^{(v)T} \mathbf{G}^{(v)} = \mathbf{I}, \\
 & \quad \sum_{v=1}^V \alpha_v = 1, \alpha_v \geq 0,
 \end{aligned} \tag{17}$$

where α_v is the weight parameter used to reflect the importance of each view. The λ_1 , λ_2 and λ_3 are hyperparameters.



[Download: Download high-res image \(540KB\)](#)

[Download: Download full-size image](#)

Fig. 2. The framework of our proposed IMVC_HG. $\mathbf{X}^{(v)}$ is an incomplete multi-view dataset, and it can be divided into observable part $\mathbf{X}_o^{(v)}$ and missing part $\mathbf{E}^{(v)}$. $\mathbf{W}_1^{(v)}$ and $\mathbf{W}_2^{(v)}$ are the projection matrices. $\mathbf{Z}^{(v)}$ and \mathbf{Z}^* are the affinity matrix for each view and the consistent affinity matrix. $\mathbf{G}_{\text{hyper}}$ is the hypergraph. $\mathbf{F}^{(v)}$ is the clustering label matrix for each view. $\|\mathcal{F}\|_{\mathcal{S}\mathcal{P}}^p$ is the tensor consisting of the clustering label matrices and \mathbf{F} is the view-weighted clustering indicator matrix.

The flow of the model is illustrated in Fig. 2. Initially, the affinity matrix for each view denoted as $\{\mathbf{Z}^{(v)}\}_{v=1}^V$, which is individually learned using self-representation method. Subsequently, a consistent affinity matrix \mathbf{Z}^* is derived from these matrices. Based on \mathbf{Z}^* , a hypergraph $\mathbf{G}_{\text{hyper}}$, encapsulating higher-order correlation information, is constructed. This hypergraph is then utilized to reconstruct the missing data. Furthermore, for each view, a clustering label matrix $\{\mathbf{F}^{(v)}\}_{v=1}^V$ is obtained through non-negative matrix factorization with orthogonal constraints. Then these matrices are assembled into a tensor and the tensor Schatten p -norm is employed to reveal the low-rank structure and complementary information of different label matrices. Finally, a view-weighted clustering indicator matrix \mathbf{F} is attained.

3.2. Optimization

The model (17) is solved optimally as follows. The Augmented Lagrange Multiplier (ALM) algorithm and Alternate Direction Minimization (ADM) strategy are utilized. Initially, to facilitate the solution, the following objective function is derived after introducing \mathcal{J} to equate \mathcal{F} :

$$\begin{aligned}
& \min_{\mathbf{E}^{(v)}, \mathbf{Z}^{(v)}, \mathbf{F}^{(v)}, \mathbf{G}^{(v)}} \sum_{v=1}^V \frac{1}{\alpha_v} \left(\lambda_1 \left\| \mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} \right\|_F^2 \right. \\
& \quad + \lambda_2 \text{Tr} \left(\mathbf{X}^{(v)} \mathbf{L}_{\text{hyper}} \mathbf{X}^{(v)T} \right) + \left\| \mathbf{Z}^{(v)} - \mathbf{Z}^* \right\|_F^2 \\
& \quad \left. + \lambda_3 \left\| \mathbf{Z}^{(v)} - \mathbf{F}^{(v)} \mathbf{G}^{(v)T} \right\|_F^2 \right) + \|\mathcal{J}\|_{\mathbb{S}^p}^p, \\
& \text{s.t. } \mathbf{X}^{(v)} = \mathbf{X}_o^{(v)} \mathbf{W}_1^{(v)} + \mathbf{E}^{(v)} \mathbf{W}_2^{(v)}, \mathbf{Z}_{ii}^{(v)} = 0, \mathbf{Z}_{ij}^{(v)} \geq 0, \\
& \mathbf{Z}^{(v)} = \mathbf{Z}^{(v)T}, \mathbf{F}^{(v)} \geq 0, \mathbf{G}^{(v)T} \mathbf{G}^{(v)} = \mathbf{I}, \\
& \sum_{v=1}^V \alpha_v = 1, \alpha_v \geq 0, \mathcal{J} = \mathcal{F}.
\end{aligned} \tag{18}$$

Furthermore, the augmented Lagrangian term is introduced to obtain the following objective function:

$$\begin{aligned}
& \mathcal{L}(\mathbf{Z}^{(v)}, \mathbf{F}^{(v)}, \mathbf{E}^{(v)}, \mathbf{G}^{(v)}, \mathcal{J}, \mathcal{F}, \mathcal{Q}_1) = \|\mathcal{J}\|_{\mathbb{S}^p}^p \\
& + \sum_{v=1}^V \frac{1}{\alpha_v} \left(\lambda_1 \left\| \mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} \right\|_F^2 + \lambda_2 \text{Tr} \left(\mathbf{X}^{(v)} \mathbf{L}_{\text{hyper}} \mathbf{X}^{(v)T} \right) \right) \\
& + \sum_{v=1}^V \frac{1}{\alpha_v} \left\| \mathbf{Z}^{(v)} - \mathbf{Z}^* \right\|_F^2 + \sum_{v=1}^V \frac{\lambda_3}{\alpha_v} \left\| \mathbf{Z}^{(v)} - \mathbf{F}^{(v)} \mathbf{G}^{(v)T} \right\|_F^2 \\
& + \frac{\mu}{2} \left\| \mathcal{F} - \mathcal{J} + \frac{\mathcal{Q}_1}{\mu} \right\|_F^2,
\end{aligned} \tag{19}$$

where \mathcal{Q}_1 represents the Lagrange multiplier and μ is the penalty parameter. The ADM strategy is then employed to update all unknown variables one by one iteratively. The specific steps are as follows.

Step 1: Fix other variables and update variable $\mathbf{Z}^{(v)}$.

The subproblem for solving $\mathbf{Z}^{(v)}$ is as follows

$$\begin{aligned}
& \min_{\mathbf{Z}^{(v)}} \sum_{v=1}^V \left(\lambda_1 \left\| \mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} \right\|_F^2 + \left\| \mathbf{Z}^{(v)} - \mathbf{Z}^* \right\|_F^2 \right. \\
& \quad \left. + \lambda_3 \left\| \mathbf{Z}^{(v)} - \mathbf{F}^{(v)} \mathbf{G}^{(v)T} \right\|_F^2 \right), \\
& \text{s.t. } \mathbf{Z}_{ii}^{(v)} = 0, \mathbf{Z}_{ij}^{(v)} \geq 0, \mathbf{Z}^{(v)} = \mathbf{Z}^{(v)T},
\end{aligned} \tag{20}$$

Neglecting its constraints for the moment, its objective function can be written as

$$\begin{aligned}
& \mathcal{L}(\mathbf{Z}^{(v)}) = \lambda_1 \left\| \mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} \right\|_F^2 + \left\| \mathbf{Z}^{(v)} - \mathbf{Z}^* \right\|_F^2 \\
& + \lambda_3 \left\| \mathbf{Z}^{(v)} - \mathbf{F}^{(v)} \mathbf{G}^{(v)T} \right\|_F^2.
\end{aligned} \tag{21}$$

Then, taking the partial derivative of Eq. (21) concerning $\mathbf{Z}^{(v)}$ and setting it to zero, we obtain

$$\bar{\mathbf{Z}}^{(v)} = \left((1 + \lambda_3) \mathbf{I} + \lambda_1 \mathbf{M}_1^{(v)} \right)^{-1} \left(\lambda_1 \mathbf{M}_1^{(v)} + \mathbf{Z}^* + \lambda_3 \mathbf{M}_2^{(v)} \right), \tag{22}$$

where $\mathbf{M}_1^{(v)} = \mathbf{X}^{(v)T} \mathbf{X}^{(v)}$, $\mathbf{M}_2^{(v)} = \mathbf{F}^{(v)} \mathbf{G}^{(v)T}$.

Considering the previously ignored constraints, the following objective function is derived

$$\begin{aligned}
& \min_{\mathbf{Z}^{(v)}} \left\| \mathbf{Z}^{(v)} - \hat{\mathbf{Z}}^{(v)} \right\|_F^2, \\
& \text{s.t. } \mathbf{Z}_{ij}^{(v)} \geq 0, \mathbf{Z}^{(v)} = \mathbf{Z}^{(v)T},
\end{aligned} \tag{23}$$

where the constraint $\mathbf{Z}_{ii}^{(v)} = 0$ is incorporated such that $\hat{\mathbf{Z}}^{(v)} = \bar{\mathbf{Z}}^{(v)} - \text{diag} \left(\text{diag} \left(\bar{\mathbf{Z}}^{(v)} \right) \right)$.

Finally, the solution formula for $\mathbf{Z}^{(v)}$ is given by

$$\mathbf{Z}^{(v)} = \left[\frac{\widehat{\mathbf{Z}}^{(v)} + \widehat{\mathbf{Z}}^{(v)T}}{2} \right]_+ . \quad (24)$$

Step 2: Fix other variables and update \mathbf{Z}^* .

The objective function with respect to \mathbf{Z}^* is given by

$$\min_{\mathbf{Z}^*} \sum_{v=1}^V \frac{1}{\alpha_v} \|\mathbf{Z}^{(v)} - \mathbf{Z}^*\|_F^2. \quad (25)$$

To find the optimal \mathbf{Z}^* , we compute the partial derivative of the objective function concerning \mathbf{Z}^* , set it to zero, and solve for \mathbf{Z}^* . This yields

$$\mathbf{Z}^* = \frac{\sum_{v=1}^V \frac{1}{\alpha_v} \mathbf{Z}^{(v)}}{\sum_{v=1}^V \frac{1}{\alpha_v}}. \quad (26)$$

Step 3: Fix other variables and update variable $\mathbf{E}^{(v)}$.

Given $\mathbf{X}^{(v)} = \mathbf{X}_o^{(v)} \mathbf{W}_1^{(v)} + \mathbf{E}^{(v)} \mathbf{W}_2^{(v)}$, the objective function with respect to $\mathbf{E}^{(v)}$ is

$$\begin{aligned} & \mathcal{L}(\mathbf{E}^{(v)}) \\ &= \lambda_1 \left\| \left(\mathbf{X}_o^{(v)} \mathbf{W}_1^{(v)} + \mathbf{E}^{(v)} \mathbf{W}_2^{(v)} \right) - \left(\mathbf{X}_o^{(v)} \mathbf{W}_1^{(v)} + \mathbf{E}^{(v)} \mathbf{W}_2^{(v)} \right) \mathbf{Z}^{(v)} \right\|_F^2 \\ &+ \lambda_2 \text{Tr} \left(\left(\mathbf{X}_o^{(v)} \mathbf{W}_1^{(v)} + \mathbf{E}^{(v)} \mathbf{W}_2^{(v)} \right) \mathbf{L}_{\text{hyper}} \left(\mathbf{X}_o^{(v)} \mathbf{W}_1^{(v)} + \mathbf{E}^{(v)} \mathbf{W}_2^{(v)} \right)^\top \right). \end{aligned} \quad (27)$$

To find the optimal $\mathbf{E}^{(v)}$, we compute the partial derivative of the objective function concerning $\mathbf{E}^{(v)}$, set it to zero, and solve for $\mathbf{E}^{(v)}$. This results in

$$\mathbf{E}^{(v)} = \left(-\mathbf{X}_o^{(v)} \mathbf{W}_1^{(v)} \mathbf{M}_3^{(v)} \mathbf{W}_2^{(v)\top} \right) \left(\mathbf{W}_2^{(v)} \mathbf{M}_3^{(v)} \mathbf{W}_2^{(v)\top} \right)^{-1}, \quad (28)$$

where $\mathbf{M}_3^{(v)} = \lambda_1 \left(\mathbf{Z}^{(v)} - \mathbf{I} \right) \left(\mathbf{Z}^{(v)\top} - \mathbf{I} \right) + \lambda_2 \mathbf{L}_{\text{hyper}}$.

Step 4: Fix other variables and update variable \mathcal{J} .

With the other variables fixed, the objective function concerning the variable \mathcal{J} is

$$\min_{\mathcal{J}} \|\mathcal{J}\|_{\mathbb{S}^p}^p + \frac{\mu}{2} \left\| \mathcal{F} - \mathcal{J} + \frac{\mathcal{Q}_1}{\mu} \right\|_F^2. \quad (29)$$

Here, relevant theorems will be presented.

Theorem 1

Theorem 1 in [35]

Let us consider a third-order tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, which has t-SVD $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$. The optimal solution to the following model

$$\arg \min_{\mathcal{X}} \|\mathcal{X}\|_{\mathbb{S}^p}^p + \frac{1}{2} \|\mathcal{X} - \mathcal{A}\|_F^2 \quad (30)$$

is

$$\mathcal{X}^* = \Gamma_{\tau^* \omega}[\mathcal{A}] = \mathcal{U} * \text{ifft} \left(P_{\tau^* \omega} \left(\overline{\mathcal{A}} \right) \right) * \mathcal{V}^T, \quad (31)$$

where $\mathbf{P}_{\tau \cdot \mathbf{n}_3}(\overline{\mathcal{A}})$ is an $\mathbf{n}_1 \times \mathbf{n}_2 \times \mathbf{n}_3$ f -diagonal tensor, whose diagonal elements are found using the generalized shrinkage thresholding (GST) algorithm introduced in Lemma 1 of [35].

According to Theorem 1, such that $\mathcal{F} + \frac{\mathcal{Q}_1}{\mu} = \mathcal{U} * \Sigma * \mathcal{V}^T$, we can get the solution of Eq. (29) as

$$\begin{aligned} \mathcal{J}^* &= \Gamma_{\frac{1}{\mu} * \omega} \left[\mathcal{F} + \frac{\mathcal{Q}_1}{\mu} \right] \\ &= \mathcal{U} * \text{ifft} \left(P_{\frac{1}{\mu} * \omega} \left(\overline{\mathcal{F} + \frac{\mathcal{Q}_1}{\mu}} \right) \right) * \mathcal{V}^T. \end{aligned} \quad (32)$$

Step 5: Fix other variables and update $\mathbf{G}^{(v)}$.

The sub-problem on $\mathbf{G}^{(v)}$ is

$$\min_{\mathbf{G}^{(v)}} \left\| \mathbf{Z}^{(v)} - \mathbf{F}^{(v)} \mathbf{G}^{(v)T} \right\|_F^2. \quad (33)$$

Eq. (33) can be expanded into

$$\begin{aligned} \min_{\mathbf{G}^{(v)}} & \text{Tr} \left(\left(\mathbf{Z}^{(v)} \right)^T \mathbf{Z}^{(v)} \right) - 2 \text{Tr} \left(\mathbf{G}^{(v)} \left(\mathbf{F}^{(v)} \right)^T \mathbf{Z}^{(v)} \right) \\ & + \text{Tr} \left(\mathbf{G}^{(v)} \left(\mathbf{F}^{(v)} \right)^T \mathbf{F}^{(v)} \left(\mathbf{G}^{(v)} \right)^T \right). \end{aligned} \quad (34)$$

Due to the constraint $\mathbf{G}^{(v)T} \mathbf{G}^{(v)} = \mathbf{I}$, it can be proven that

$$\text{Tr} \left(\mathbf{G}^{(v)} \left(\mathbf{F}^{(v)} \right)^T \mathbf{F}^{(v)} \left(\mathbf{G}^{(v)} \right)^T \right) = \text{Tr} \left(\left(\mathbf{F}^{(v)} \right)^T \mathbf{F}^{(v)} \right). \quad (35)$$

Thus, Eq. (34) can be written as

$$\max_{\left(\mathbf{G}^{(v)} \right)^T \mathbf{G}^{(v)} = \mathbf{I}} \text{Tr} \left(\mathbf{G}^{(v)} \left(\mathbf{F}^{(v)} \right)^T \mathbf{Z}^{(v)} \right). \quad (36)$$

According to Theorem 2, the optimal solution of Eq. (36) can be found as

$$\mathbf{G}^{(v)} = \mathbf{V}^{(v)} \left(\mathbf{\Lambda}^{(v)} \right)^T. \quad (37)$$

Theorem 2

Given the matrices \mathbf{A} and \mathbf{B} , and considering the constraint for \mathbf{A} that $\mathbf{A}^T \mathbf{A} = \mathbf{I}$, we aim to maximize $\text{Tr}(\mathbf{A}\mathbf{B})$ under the condition that $\mathbf{A}^T \mathbf{A} = \mathbf{I}$. The optimal solution for matrix \mathbf{A} can be expressed as $\mathbf{A} = \mathbf{V}(\mathbf{\Lambda})^T$, where $\mathbf{\Lambda} \mathbf{S} \mathbf{V}^T = \text{svd}(\mathbf{B})$ [39].

Step 6: Fix other variables and update variable $\mathbf{F}^{(v)}$.

The sub-problem on $\mathbf{F}^{(v)}$ is given by

$$\min_{\mathbf{F}^{(v)}} \frac{\lambda_3}{\alpha_v} \left\| \mathbf{Z}^{(v)} - \mathbf{F}^{(v)} \mathbf{G}^{(v)T} \right\|_F^2 + \frac{\mu}{2} \left\| \mathbf{F}^{(v)} - \mathbf{J}^{(v)} + \frac{1}{\mu} \mathbf{Q}_1^{(v)} \right\|_F^2. \quad (38)$$

Finding the partial derivative of $\mathbf{F}^{(v)}$ with respect to Eq. (38) and setting it to zero, we obtain

$$\bar{\mathbf{F}}^{(v)} = \frac{\mathbf{M}_4^{(v)}}{\frac{2\lambda_3}{\alpha_v} + \mu}, \quad (39)$$

where $\mathbf{M}_4^{(v)} = \mu \mathbf{J}^{(v)} - \mathbf{Q}_1^{(v)} + \frac{2\lambda_3}{\alpha_v} \mathbf{Z}^{(v)} \mathbf{G}^{(v)}$. The solution of $\mathbf{F}^{(v)}$ can be found as

$$\mathbf{F}^{(v)} = \begin{cases} \bar{\mathbf{F}}^{(v)}, & \text{if } \bar{\mathbf{F}}^{(v)} \geq 0, \\ \mathbf{0}, & \text{if } \bar{\mathbf{F}}^{(v)} < 0. \end{cases} \quad (40)$$

Step 7: Fix other variables and update variables \mathcal{Q}_1 , α_v , and μ .

The sub-problem on α_v is

$$\begin{aligned} \min_{\alpha_v} \quad & \sum_{v=1}^V \frac{1}{\alpha_v} \Gamma(v), \\ \text{s.t.} \quad & \sum_{v=1}^V \alpha_v = 1, \quad \alpha_v \geq 0, \end{aligned} \quad (41)$$

where

$$\begin{aligned} \Gamma(v) = & \lambda_1 \|\mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)}\|_F^2 + \lambda_2 \text{Tr} \left(\mathbf{X}^{(v)} \mathbf{L}_{\text{hyper}} \mathbf{X}^{(v)T} \right) \\ & + \|\mathbf{Z}^{(v)} - \mathbf{Z}^*\|_F^2 + \lambda_3 \|\mathbf{Z}^{(v)} - \mathbf{F}^{(v)} \mathbf{G}^{(v)T}\|_F^2. \end{aligned}$$

Applying the Cauchy–Schwarz inequality, the view weights are updated as

$$\alpha_v^+ = \frac{\Gamma^{\frac{1}{2}}(v)}{\sum_{v=1}^V \Gamma^{\frac{1}{2}}(v)}. \quad (42)$$

The Lagrange multipliers are updated using the rule

$$\mathcal{Q}_1 = \mathcal{Q}_1 + \mu(\mathcal{F} - \mathcal{J}). \quad (43)$$

The penalty parameters are updated following the rule

$$\mu = \min(\rho\mu, \mu_0), \quad (44)$$

where ρ and μ_0 are constants.

The algorithm flow is shown in Algorithm 1.

Algorithm 1: Incomplete multi-view clustering based on hypergraphs algorithm process

Input: Incomplete multi-view data $\{\mathbf{X}_o^{(v)}\}_{v=1}^V$, projection matrices $\{\mathbf{W}_1^{(v)}\}_{v=1}^V$ and $\{\mathbf{W}_2^{(v)}\}_{v=1}^V$;

Parameters: $\lambda_1, \lambda_2, \lambda_3$ and p ; $\mu_0 = 10^{10}$, $\rho = 1.2$, $\mu = 10^{-2}$;

Output: Clustering label \mathbf{Y} ;

Initialization: Initialize $E_{ij}^{(v)} = 0$, $\mathbf{F}^{(v)}$ as a random positive definite matrix and $\mathbf{G}^{(v)}$ as a random orthogonal matrix. $\alpha_v = V$.

while not converged do

Calculate $\{\mathbf{Z}^{(v)}\}_{v=1}^V$ via Eq. (24);
 Calculate \mathbf{Z}^* via Eq. (26);
 Update $\mathbf{L}_{\text{hyper}}$;
 Calculate $\{\mathbf{E}^{(v)}\}_{v=1}^V$ via Eq. (28);
 Calculate \mathcal{J} via Eq. (32);
 Calculate $\{\mathbf{G}^{(v)}\}_{v=1}^V$ via Eq. ??;
 Calculate $\{\mathbf{F}^{(v)}\}_{v=1}^V$ via Eq. (40);
 Calculate α_v , \mathbf{Q}_1 , and μ via Eq. (42), (43), and (44);

end

Calculate the weighted clustering indicator matrix $\mathbf{F} = \sum_{v=1}^V \left(\frac{1}{\alpha_v} \mathbf{F}^{(v)} \right)$, where $\mathbf{F} \in \mathbb{R}^{n \times c}$;

Calculate clustering label $\mathbf{Y} \in \mathbb{R}^{n \times 1}$. The corresponding cluster label is the index of the maximum value of each row in the matrix \mathbf{F} .

return \mathbf{Y} ;

[Download: Download high-res image \(487KB\)](#)

[Download: Download full-size image](#)

3.3. Computational complexity analysis

The computational complexity of our method lies mainly in solving $\{\mathbf{Z}^{(v)}\}_{v=1}^V, \{\mathbf{E}^{(v)}\}_{v=1}^V, \mathcal{J}, \{\mathbf{G}^{(v)}\}_{v=1}^V$ and $\{\mathbf{F}^{(v)}\}_{v=1}^V$. For updating $\{\mathbf{Z}^{(v)}\}_{v=1}^V$, the cost of solving the inverse matrix for each iteration is $\mathcal{O}(V(n^3 + n^2 d_v))$. For updating $\{\mathbf{E}^{(v)}\}_{v=1}^V$, it costs $\mathcal{O}(V(n - n_v)^3)$ to solve inverse matrix in each iteration. Iterative updating $\mathcal{J}, \{\mathbf{G}^{(v)}\}_{v=1}^V$ and $\{\mathbf{F}^{(v)}\}_{v=1}^V$, needs $\mathcal{O}(Vnc \log(Vn) + V^2 nc)$, $\mathcal{O}(Vnc^2)$ and $\mathcal{O}(Vn^2 c)$ respectively. Since c and V are relatively small, the complexity of the rest of the calculation process can be neglected. Therefore, the total complexity of Algorithm 1 is $\mathcal{O}(V(n^3 + n^2 d_v) + V(n - n_v)^3 + Vnc \log(Vn) + V^2 nc + Vnc^2 + Vn^2 c)$ per iteration. It is obvious that the computational complexity of our algorithm is relatively high, so we discuss possible methods to reduce the computational complexity. As we all know, anchor-based methods first generate representative anchors from the original data, whose number is much less than the number of original data, then build an anchor graph to represent the data structure, and finally implement clustering on the anchor graph, which is conducive to reducing the computational complexity and achieving an improvement in the algorithm speed [40], [41], [42], [43]. Therefore, in future research, we can explore the anchor strategy to reduce the limitation of the high computational complexity of our algorithm so that it can be applied to large-scale datasets.

4. Experimental procedure

In this section, the experimental settings will be introduced in detail, including the five datasets we use, data pre-processing, the evaluation metrics and experimental environment, etc. In addition, we select some compared methods for the proposed model and briefly introduce them.

4.1. Datasets and evaluation metrics

To assess the effectiveness of the proposed algorithm, we conducted experimental analyses using the following five datasets. The details of the datasets are given in Table 3.

BDGP¹: The BDGP (Berkeley Drosophila Genome Project) dataset contains a large number of Drosophila embryo microscopy images. We use the data processed in [44] to conduct our experiments. Consisting of 2,500 samples divided into five categories, this dataset includes two modalities – image and text – each representing a separate view. In our experiments, we treat these modalities as distinct views, where the image feature view dimension is 1750 and the text feature view dimension is 79.

HW²: The digits used in the experiment were extracted from nine original maps from a Dutch public utility. The maps were processed through a series of operations such as scanning, sharpening and thresholding, and manually labeled. From a set of 2000 digits containing 0-9, 200 samples were extracted for each of the 10 categories, and four types of features were extracted for each sample [45]. In our experiments, we select two types of feature sets as two views, the first view consists of Fourier coefficients of data, and the second view consists of profile correlations of digits. Each sample has 216 features in the first view and 76 features in the second view.

mnist4³: The MNIST dataset is a widely used handwritten digit dataset containing 60,000 training samples and 10,000 test samples. Each sample is a 28×28 pixel grayscale image representing a handwritten digit from 0 to 9 [46]. We focus on the first 4000 samples in the training set, denoted as mnist4. For our experiments, we select two modal data as two separate views, which are 30-dimensional ISO feature and 9-dimensional LDA feature.

BBCSport⁴: This dataset consists of 737 documents which are described by 2–5 views. All documents are sampled from the BBC Sport website corresponding to sports news articles in five topical areas [47]. In our study, we extracted 116 samples to construct the multi-view dataset. Four distinct feature types describe each

sample, thus establishing four views per sample. The dimensions of the four features are 1991, 2063, 2113, and 2158 respectively.

3-Sources⁵: The dataset is collected from three well-known online news sources: BBC, Reuters, and The Guardian. It contains a total of 948 news articles, 169 of which are reported in all three sources, and each story is manually annotated with one or more of six topic labels⁶ Our experiment selects these 169 samples, which consist of three views. The dimensions of the features of the three views are 3560, 3631, and 3068 respectively.

For the HW, BDGP, and mnist4 datasets, 30%, 50%, and 70% of the samples were randomly selected as paired samples, respectively. For instance, when the pairing rate is 30%, only 30% of the samples for each view do not lack any feature information, leaving 70% of the samples with feature information for only one view, and feature information for the other view is deleted in the preprocessing stage. For the 3-Sources and BBCSport datasets, in order to construct incomplete multi-view datasets, we randomly remove about 55% of the sample features in each view, i.e., only 45% of the sample features in each view are left.

Table 3. Details of datasets. “Sample”, “Class”, “View” and “Feature” denote the number of samples, classes contained in the samples, views and features contained in each view, respectively.

Dataset	Class	View	Sample	Feature
BDGP	5	2	2500	79/1750
HW	10	2	2000	216/76
mnist4	4	2	4000	30/9
BBCSport	5	4	116	1991/2063/2113/2158
3-Sources	6	3	169	3560/3631/3068

Various criteria are employed to evaluate the clustering performance of different methods [48]. In this study, three well-established evaluation metrics were utilized: clustering accuracy (**ACC**), normalized mutual information (**NMI**), and purity (**PUR**). Generally, higher values of these metrics signify better clustering performance. Our experiments are conducted on a standard Windows 10 Server with two Intel (R) Xeon (R) Gold 6230 CPUs at 2.1 GHz and 128 GB RAM, and MATLAB R2020b.

4.2. Compared methods

To assess the performance of our proposed method (IMVC_HG), we selected the following methods for comparison:

SC [49]: A standard unsupervised clustering approach, Spectral Clustering (SC), which generates clusters by forming a graph representation for each view.

ConSC [49]: An extension of SC that aggregates features from multiple views by concatenating each view’s feature space and then applying spectral clustering to the unified feature space.

AMGL [50]: An adaptive multi-view graph learning method that optimizes the weights for each view’s adjacency matrix to achieve a consensus graph for the clustering structure of the multi-view data.

RMSC [51]: Utilizes robust principal component analysis to learn a shared adjacency matrix that captures the global clustering structure.

GPMVC [24]: Employs graph-regularized Non-negative Matrix Factorization (NMF) for low-dimensional data representation, followed by K-means clustering to enhance the representation of the intrinsic data structure.

IMG [20]: Accomplishes latent view representation through matrix factorization and smoothing penalties, then used for K-means clustering of incomplete multi-view data.

IMSC_AGL [16]: Aims to adaptively learn multiple graphs from incomplete data to construct a direct consensus representation for clustering.

BSV [20]: Imputes missing instances by averaging existing instances, independently executes K-means on each view, and selects the best clustering outcome.

Concat [20]: Concatenates the features from each view into a single feature map and directly applies K-means clustering to this composite feature array.

HCP_IMSC [30]: Integrates affinity matrix learning with tensor factorization and hyper-Laplacian regularization within a unified framework.

IMVC_TSN [52]: Concentrates on learning a stable, low-dimensional view representation through tensor Schatten p-norm, which takes into account the inherent distribution of view-specific data and distinguishes between principal and minor information during the learning process.

PIMVC [53]: Solves IMVC problems using projection learning. Introduces an uncorrelated constraint based on scatter matrices and a graph constraint based on geometric structure to avoid trivial solutions for projections.

BGIMVSC [31]: Learns a common consensus graph from all incomplete views and introduces a relaxed spectral clustering to obtain a probability consensus representation that contains all positive elements reflecting the clustering results. A weighted multi-view learning mechanism is also introduced to automatically balance the different views.

OGIMC [33]: The strategy of local structure preservation and adaptive weighting are introduced into the model and a rank constraint is applied to the Laplacian matrix of the fused graph, integrating the separate objectives into a unified framework.

5. Experimental results and analysis

In this section, we give the clustering results of various methods on five datasets and analyze the significance of differences using t -test for them. Confusion matrix visualization and t-SNE visualization are used to compare the model in this paper with other methods to confirm its superiority. For the four parameters in our model, we perform a parametric sensitivity analysis to confirm the effectiveness of the parameters on the clustering performance. Finally, we analyze the convergence of the model on two datasets.

5.1. Clustering validity analysis

Tables 4, 5, 6, 7, Table 8, Table 9 present the experimental results obtained from applying various methods to those above incomplete multi-view datasets. Moreover, in order to demonstrate that our improvements are statistically significant, we repeat each algorithm 50 times and perform a one-tailed t -test with a significance level of 0.05 on the 50 values of each clustering metric to show the significant differences between our method and each competitor. An analysis of these results reveals that:

(1) It is acknowledged that single-view clustering methods typically do not surpass multi-view clustering approaches when processing multi-view data. Nevertheless, exceptions exist. For instance, in Table 4, the single-view clustering (SC) outperforms the multi-view clustering (AMGL) on BDGP dataset. The multi-view clustering algorithm AMGL endeavors to impute missing instances with their average values to mitigate the impact of missing data on clustering outcomes. However, this study reveals that merely employing averages for data imputation is ineffective in enhancing performance. In contrast, as proposed in this paper, the

hypergraph-based reconstruction of missing views leverages the higher-order information inherent in multi-view data. This strategy proficiently utilizes the complex interrelations among two or more sample data points, thereby enhancing clustering outcomes beyond those of the top-performing competitors. For instance, with a 30% sample pairing rate in the BDGP dataset, the proposed algorithm’s *ACC*, *NMI*, and *PUR* metrics exhibit improvements of 4.63%, 31.11%, and 12.11% respectively, compared to the *IMVC_TSN*, which does not involve the hypergraph theory. Moreover, our method has better clustering results compared to multi-view clustering methods such as *IMG*, *IMSC_AGL*, *PIMVC*, *BGIMVSC*, *OGIMC*, etc., which may be attributed to the fact that the tensor Schatten p -norm is used to reveal the low-rank structure and complementary information of different label matrices.

(2) Table 5 presents the clustering results on BBCSport and 3-Sources datasets. The proposed algorithm demonstrates a marked improvement in clustering effectiveness over traditional single-view clustering methods. Specifically, on the BBCSport dataset, our algorithm achieves *ACC*, *NMI*, and *PUR* scores that surpass the *BSV* algorithm by 42.22%, 55.27%, and 41.79%, respectively. When juxtaposed with multi-view clustering algorithms, the proposed method exhibits superior performance. For instance, on the 3-Sources dataset, *ACC*, *NMI*, and *PUR* of our algorithm are 7.41%, 7.10%, and 9.08% higher than those of *IMSC_AGL*, respectively. It is worth noting that our algorithm performs better than *HCP_IMSC*, which also uses hypergraph theory. This may be attributed to the fact that compared to *HCP_IMSC*, which requires an additional clustering process, we use non-negative matrix factorization with orthogonality constraints to equate K-means approach to eliminate the need for additional post-processing and avoid the problem of sub-optimal results caused by the two-step approach. This approach significantly bolsters the model’s clustering accuracy.

Table 4. Clustering results on the BDGP, HW and mnist4 datasets. Bold and underlined lines indicate optimal and suboptimal results, respectively. The p -value of t -test with a significance level of 0.05 is obtained and “*” means that the difference is significant.

Datasets	Methods	0.3			0.5			0.7		
		ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
BDGP	SC1 [49]	0.3539*	0.1347*	0.3572*	0.3845*	0.1626*	0.3878*	0.4103*	0.1884*	0.4122*
	SC2 [49]	0.5169*	0.3158*	0.5235*	0.5692*	0.3745*	0.5764*	0.6139*	0.4364*	0.6194*
	ConSC [49]	0.2230*	0.0228*	0.2268*	0.2139*	0.0126*	0.2165*	0.2106*	0.0098*	0.2126*
	AMGL [50]	0.2357*	0.0260*	0.2386*	0.2538*	0.0420*	0.2558*	0.2807*	0.0756*	0.2826*
	RMSC [51]	0.3683*	0.1487*	0.3712*	0.3907*	0.1684*	0.3934*	0.4233*	0.1962*	0.4246*
	GPMVC [24]	0.5424*	0.3907*	0.5615*	0.6277*	0.4827*	0.6451*	0.6833*	0.5741*	0.7062*
	IMG [20]	0.4508*	0.2809*	0.4570*	0.4868*	0.3147*	0.4916*	0.5055*	0.3246*	0.5099*
	IMSC_AGL [16]	0.6281*	0.3914*	0.6022*	0.7258*	0.4851*	0.7258*	0.7524*	0.5561*	0.7524*
	IMVC_TSN [52]	<u>0.6409*</u>	0.4065*	<u>0.6409*</u>	<u>0.7536*</u>	<u>0.5272*</u>	<u>0.7536*</u>	<u>0.7678*</u>	0.5588*	<u>0.7678*</u>
	PIMVC [53]	0.4036*	0.4177*	0.4992*	0.5156*	0.5142*	0.5924*	0.6568*	<u>0.5894*</u>	0.7128*
	BGIMVSC [31]	0.3884*	0.3131*	0.3960*	0.4064*	0.2953*	0.4140*	0.3028*	0.1698*	0.3364*
	OGIMC [33]	0.5120*	<u>0.4416*</u>	0.5308*	0.4192*	0.3383*	0.4380*	0.2868*	0.1764*	0.3404*
	IMVC_HG	0.6872	0.7176	0.7620	0.7908	0.7535	0.7908	0.8252	0.8226	0.8252
HW	SC1 [49]	0.5097*	0.4419*	0.5175*	0.5742*	0.5083*	0.5820*	0.6432*	0.5717*	0.6511*
	SC2 [49]	0.3279*	0.2614*	0.3366*	0.3527*	0.2973*	0.3658*	0.3766*	0.3296*	0.3914*

Datasets	Methods	0.3			0.5			0.7		
		ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
mnist4	ConSC [49]	0.4825*	0.5202*	0.5148*	0.5453*	0.5765*	0.5767*	0.6454*	0.6661*	0.6587*
	AMGL [50]	0.4221*	0.3576*	0.4373*	0.5024*	0.4633*	0.5266*	0.6037*	0.5840*	0.6222*
	RMSC [51]	0.4625*	0.4461*	0.4782*	0.5564*	0.5020*	0.5651*	0.6330*	0.5650*	0.6404*
	GPMVC [24]	0.3077*	0.3787*	0.3410*	0.3419*	0.4256*	0.3692*	0.4236*	0.4993*	0.4425*
	IMG [20]	0.5455*	0.4532*	0.5565*	0.5457*	0.4590*	0.5585*	0.5529*	0.4798*	0.5690*
	IMSC_AGL [16]	0.7884*	0.6737*	0.7884*	0.8328*	0.7285*	0.8328*	0.8477*	0.7506*	0.8477*
	IMVC_TSN [52]	<u>0.7958</u>	0.6826*	<u>0.7958</u>	<u>0.8385</u> *	0.7377*	<u>0.8385</u> *	<u>0.8568</u> *	<u>0.7612</u> *	<u>0.8568</u> *
	PIMVC [53]	0.6870*	<u>0.6961</u> *	0.7275*	0.7285*	<u>0.7421</u> *	0.7730*	0.7495*	0.7415*	0.7870*
	BGIMVSC [31]	0.6425*	0.6283*	0.6735*	0.5285*	0.5642*	0.5680*	0.5095*	0.5569*	0.5760*
	OGIMC [33]	0.6335*	0.6278*	0.6760*	0.6125*	0.6449*	0.6580*	0.5440*	0.5696*	0.5895*
	IMVC_HG	0.7975	0.7019	0.7975	0.8543	0.7631	0.8543	0.8965	0.9300	0.8965
	SC1 [49]	0.6500*	0.6354*	0.7381*	0.6550*	0.6453*	0.7430*	0.6544*	<u>0.6442</u> *	0.7424*
	SC2 [49]	0.5292*	0.1379*	0.5292*	0.6780*	0.3143*	0.6780*	0.7515*	0.4355*	0.7515*
	ConSC [49]	0.5446*	0.1529*	0.5446*	0.6243*	0.5410*	0.7083*	0.6421*	0.5747*	0.7232*
	AMGL [50]	<u>0.8592</u> *	<u>0.6312</u> *	<u>0.8592</u> *	0.8080*	0.5203*	0.8080*	<u>0.8435</u> *	0.5992*	<u>0.8435</u> *
	RMSC [51]	0.5657*	0.1694*	0.5657*	0.7100*	0.3501*	0.7100*	0.7566*	0.4576*	0.7643*
	GPMVC [24]	0.6745*	0.5124*	0.6745*	0.5795*	0.5133*	0.6780*	0.6305*	0.5904*	0.7235*
	IMG [20]	0.6973*	0.5254*	0.7003*	0.7058*	0.5326*	0.7058*	0.7033*	0.5327*	0.7040*
	IMSC_AGL [16]	0.6209*	0.5206*	0.6875*	0.6332*	0.4719*	0.6857*	0.6190*	0.5230*	0.7075*
	IMVC_TSN [52]	0.6060*	0.5107*	0.6980*	0.6125*	0.5074*	0.7000*	0.6202*	0.5334*	0.7092*
	PIMVC [53]	0.7170*	0.4722*	0.7170*	<u>0.8980</u> *	<u>0.7439</u> *	<u>0.8980</u> *	0.7553*	0.5739*	0.7553*
	BGIMVSC [31]	0.6590*	0.5764*	0.7050*	0.5990*	0.4622*	0.6508*	0.6308*	0.6181*	0.7345*
	OGIMC [33]	0.6438*	0.5267*	0.6663*	0.7018*	0.5184*	0.7018*	0.7470*	0.5571*	0.7470*
	IMVC_HG	0.9343	0.8536	0.9343	0.9473	0.8816	0.9473	0.9385	0.8630	0.9385

Table 5. Clustering results on the 3-Sources and BBCSport datasets. Bold and underlined lines indicate optimal and suboptimal results, respectively. The p -value of t -test with a significance level of 0.05 is obtained and “*” means that the difference is significant.

Method	3-Sources			BBCSport		
	ACC	NMI	PUR	ACC	NMI	PUR
BSV [20]	0.3598*	0.2148*	0.4936*	0.4312*	0.1860*	0.4700*
Concat [20]	0.4268*	0.2608*	0.5513*	0.4290*	0.1948*	0.4634*
IMSC_AGL [16]	0.4762*	0.2862*	0.5542*	0.6914*	0.5376*	0.7810*

Method	3-Sources			BBCSport		
	ACC	NMI	PUR	ACC	NMI	PUR
HCP_IMSC [30]	0.4201*	0.2593*	0.5799*	<u>0.7672*</u>	<u>0.6088*</u>	<u>0.8190*</u>
IMVC_TSN [52]	<u>0.4806*</u>	0.2773*	0.5539*	0.7586*	0.6060*	<u>0.8190*</u>
PIMVC [53]	0.2959*	0.1624*	0.4793*	0.7500*	0.5766*	0.8017*
BGIMVSC [31]	0.3254*	0.2935*	0.5858*	0.6121*	0.3981*	0.6293*
OGIMC [33]	0.3432*	<u>0.3354*</u>	<u>0.6272*</u>	0.4741*	0.3030*	0.5948*
IMVC_HG	0.5503	0.3572	0.6450	0.8534	0.7387	0.8879

Table 6. Clustering results using same missing rates for each view on the 3-Sources dataset. “missing rates” denotes the missing rate corresponding to all views.

Missing rates	3-Sources		
	ACC	NMI	PUR
0.1	0.5562	0.4863	0.6982
0.2	0.6331	0.4343	0.6391
0.3	0.5325	0.4974	0.6805
0.4	0.6568	0.5190	0.6982
0.5	0.5325	0.3811	0.6213
0.6	0.4260	0.3141	0.5858
0.7	0.3373	0.1662	0.4970
0.8	0.3728	0.1796	0.5385
0.9	0.3018	0.0664	0.3965

Table 7. Clustering results using same missing rates for each view on the BBCSport dataset. “missing rates” denotes the missing rate corresponding to all views.

Missing rates	BBCSport		
	ACC	NMI	PUR
0.1	0.8879	0.8891	0.9569
0.2	0.8534	0.8380	0.9224
0.3	0.8534	0.7630	0.8966
0.4	0.7759	0.6589	0.8103
0.5	0.8362	0.7510	0.8621
0.6	0.7155	0.6367	0.8362
0.7	0.5431	0.3880	0.6810
0.8	0.4828	0.2712	0.5517

Missing rates	BBCSport		
	ACC	NMI	PUR
0.9	0.5431	0.3106	0.6293

Table 8. Clustering results using different missing rates for different views on the 3-Sources dataset. “missing rates” denotes the missing rate corresponding to each view.

Missing rates	3-Sources		
	ACC	NMI	PUR
0.15, 0.95, 0.55	0.6036	0.4916	0.7041
0.25, 0.85, 0.55	0.5680	0.3622	0.6213
0.35, 0.75, 0.55	0.5680	0.3163	0.5680
0.45, 0.65, 0.55	0.5621	0.3970	0.6450
0.55, 0.55, 0.55	0.5503	0.3572	0.6450
0.65, 0.45, 0.55	0.5621	0.3799	0.6450
0.75, 0.35, 0.55	0.6036	0.3953	0.6154
0.85, 0.25, 0.55	0.5976	0.4255	0.6391
0.95, 0.15, 0.55	0.5858	0.4976	0.6746

Table 9. Clustering results using different missing rates for different views on the BBCSport dataset. “missing rates” denotes the missing rate corresponding to each view.

Missing rates	BBCSport		
	ACC	NMI	PUR
0.55, 0.15, 0.55, 0.95	0.8707	0.7856	0.9052
0.55, 0.25, 0.55, 0.85	0.7845	0.6217	0.8190
0.55, 0.35, 0.55, 0.75	0.7155	0.5945	0.7759
0.55, 0.45, 0.55, 0.65	0.7845	0.6908	0.8276
0.55, 0.55, 0.55, 0.55	0.8534	0.7387	0.8879
0.55, 0.65, 0.55, 0.45	0.7241	0.5370	0.7586
0.55, 0.75, 0.55, 0.35	0.8017	0.7225	0.8621
0.55, 0.85, 0.55, 0.25	0.8793	0.7802	0.8879
0.55, 0.95, 0.55, 0.15	0.8276	0.6852	0.8534

(3) More importantly, the statistical test shows that the clustering metrics of our method are significantly different from those of the compared methods, and only very few differences are not significant. Nevertheless, our method always obtains the best results. This is enough to show that our method really improves the

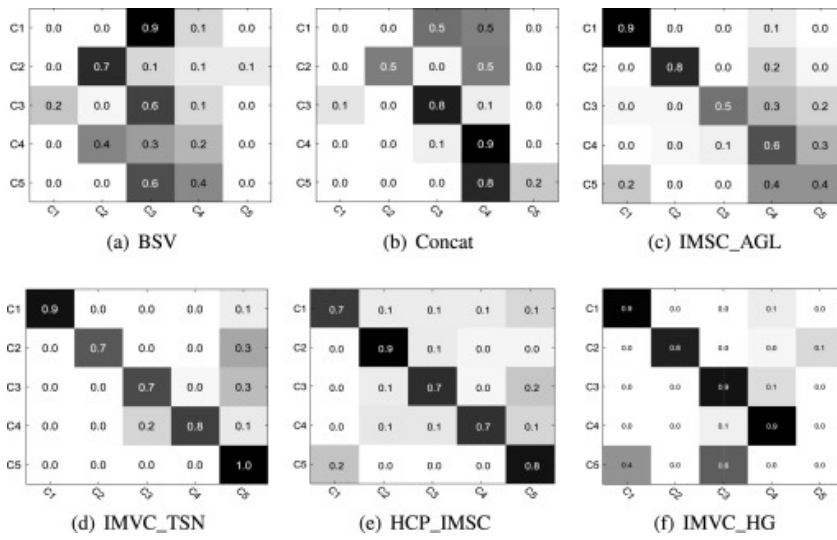
clustering performance by handling the missing data reconstruction, representation learning and clustering problems simultaneously, and also reflects the robustness of our method.

4) Inspired by [54], [55], we vary the missing rate for the experiment. In Table 4, Table 5, by setting the missing rate of BDGP, HW and Mnist4 datasets to 70%, 50%, and 30%, respectively, it is obvious that the clustering results with lower missing rate are generally better than those with higher missing rate. Similarly, in Table 6, Table 7, we make a more detailed division of the missing rate on 3-Sources and BBCSport datasets. The results show that when the missing rate is extremely high, there are too few observable data, which may not be conducive to using hypergraph to reconstruct the missing data, thus posing greater challenges to our algorithm and having a negative impact on the clustering results. Moreover, we experiment with different missing rates for different views, setting the missing rates of some views to be extremely high and some views to be very low, so as to ensure that the overall amount of missing data in each experiment remains basically unchanged. Table 8, Table 9 show the clustering results using different missing rates for different views on 3-Sources and BBCSport datasets, from which we can conclude that our algorithm still obtains relatively good clustering results. Even in the extreme case of missing data, where the data for a certain view is extremely missing, our algorithm remains robust. This may be due to the fact that our algorithm has a weight parameter α_v that balances the importance of different views. In the face of extreme missing data in a certain view, α_v automatically adjusts the contribution of this view to the clustering process, thereby obtaining better clustering results.

5.2. Visualization of results

Confusion matrix visualization: In Fig. 3, we illustrate the confusion matrix of the proposed algorithm *IMVC_HG*, along with those generated by benchmark algorithms such as *BSV*, *Concat*, *IMSC_AGL*, *IMVC_TSN*, and *HCP_IMSC*. The abscissa (horizontal axis) denotes the ground truth categories, and the ordinate (vertical axis) represents the clusters identified by the algorithms. Visualization of confusion matrices on the BBCSport dataset allows to visually assess the clustering acumen of each algorithm. It is clear that our method has clustering correctness that is difficult to be surpassed by other methods in the first four classes which have a large number of samples, even though it is less accurate in the fifth class which has a very small number of samples. Thus, relative to the competing techniques, the algorithm proposed herein achieves markedly enhanced clustering correctness in nearly all discernible classes.

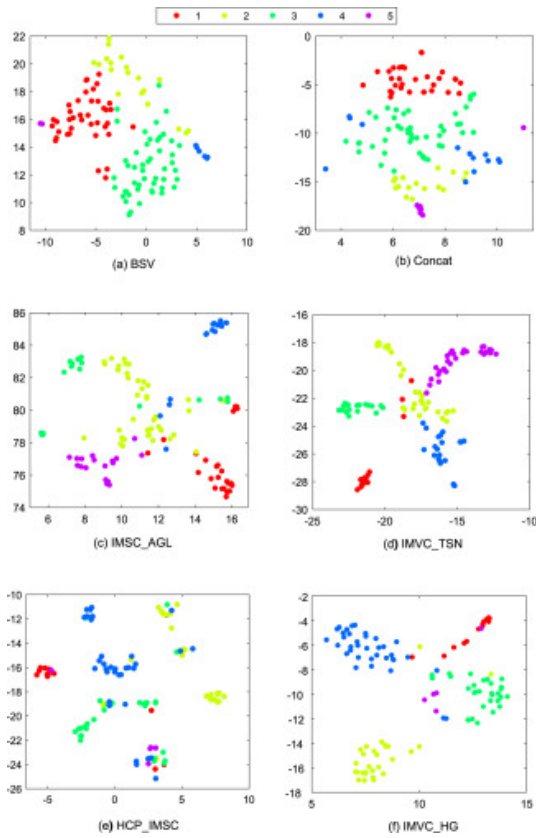
Potential representation visualization: In Fig. 4, we present the t-SNE visualization graphs (t-SNE was proposed by Felzenszwalb et al. [56]) to compare the consensus potential representations on the BBCSport dataset learned by our proposed algorithm, *IMVC_HG*, with those learned by several comparison algorithms (*BSV*, *Concat*, *IMSC_AGL*, *IMVC_TSN* and *HCP_IMSC*). Similarly, in Fig. 5, we display the t-SNE visualization graphs for the consensus potential representations on the BDGP dataset learned by *IMVC_HG* and other compared algorithms, namely *IMG*, *GPMVC*, *AMGL*, *IMSC_AGL* and *IMVC_TSN*. Based on these visualizations, it is evident that compared to other methods where data points of different classes are mixed together, our method has a clearer delineation of the different classes. Therefore, our proposed algorithm has the most superior clustering performance.



[Download: Download high-res image \(308KB\)](#)

[Download: Download full-size image](#)

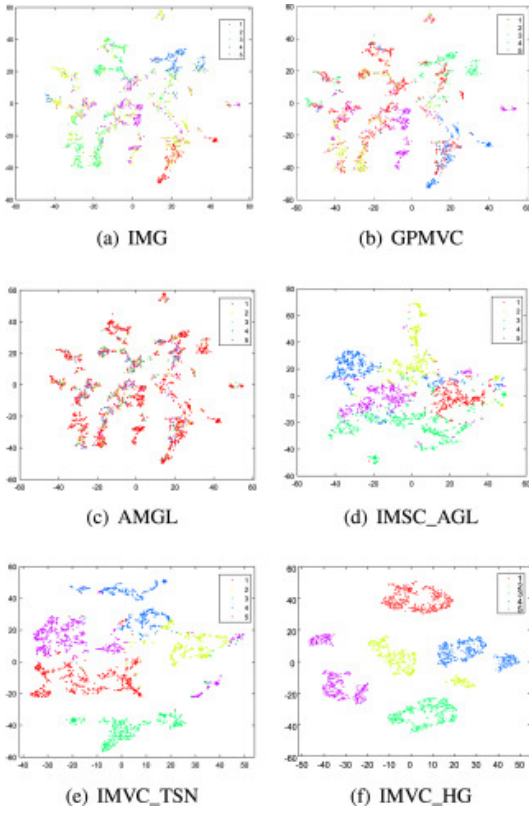
Fig. 3. Visualization of confusion matrices on the BBCSport dataset.



[Download: Download high-res image \(383KB\)](#)

[Download: Download full-size image](#)

Fig. 4. Visualization of t-SNE for consensus potential representations on the BBCSport dataset.



[Download: Download high-res image \(640KB\)](#)

[Download: Download full-size image](#)

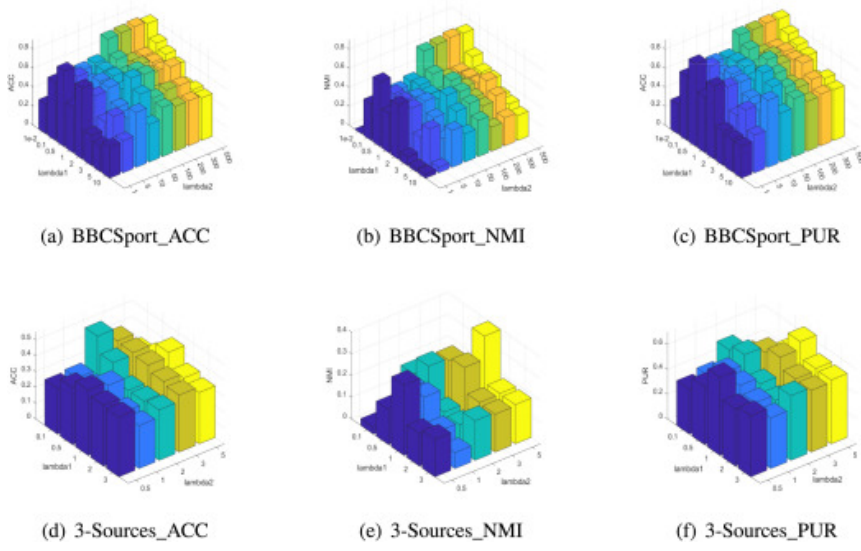
Fig. 5. Visualization of t-SNE for consensus potential representation of the BDGP dataset at 50% deletion rate.

5.3. Parametric sensitivity analysis

In this paper, the proposed model is defined by a quartet of parameters: λ_1 , λ_2 , λ_3 , and p . Parameters λ_1 , λ_2 , and λ_3 serve as balancing factors, with permissible values delineated by the discrete set $\{0.00001, 0.0001, 0.001, 0.01, 0.1, 0.5, 1, 2, 3, 5, 10, 50, 100, 200, 300, 500, 1000\}$. And we set the value of p to vary from 0.1 to 1 in intervals of 0.1. The sensitivity of these parameters to the performance of the proposed algorithm is analyzed as follows.

The impact of parameters λ_1 , λ_2 , and λ_3 on the clustering outcomes for the BBCSport and 3-Sources datasets is quantitatively evaluated. Fixing other parameters, Fig. 6 offers a visual analysis through panels (a)–(c) for the BBCSport dataset, illustrating the sensitivity of clustering performance to changes in λ_1 and λ_2 . Similarly, panels (d)–(f) in the same figure address their influence on the 3-Sources dataset. It is clear that when λ_1 and λ_2 are changed, the values of clustering metrics ACC, NMI, PUR also change significantly.

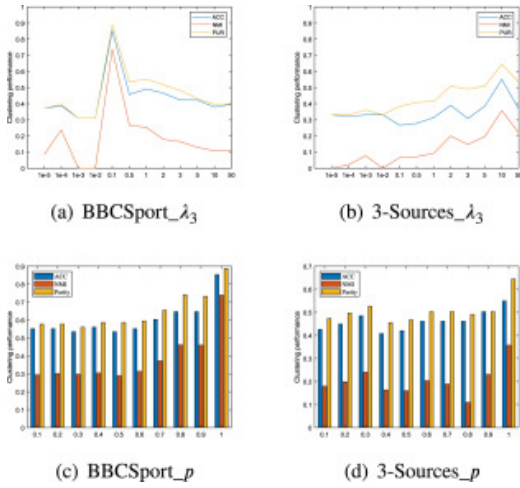
Furthermore, Fig. 7 articulates the integral roles played by λ_3 and p in modulating clustering efficiency across both datasets. With other parameters held stationary, it can be found that the value of p has a significant influence on the clustering performance. This is probably because that the Schatten p -norm approximates the target rank well. The significant changes in the clustering metrics ACC, NMI, PUR also confirm the sensitivity of λ_3 to the performance of the proposed algorithm. Concretely, corresponding to the clustering results in Table 4, Table 5, the parameter settings of our method are shown in Table 10.



[Download: Download high-res image \(489KB\)](#)

[Download: Download full-size image](#)

Fig. 6. (a)–(c) illustrate the impact of parameters λ_1 and λ_2 on the clustering outcomes for the BBCSport dataset. Similarly, (d)–(f) depict how these parameters affect the clustering results for the 3-Sources dataset.



[Download: Download high-res image \(297KB\)](#)

[Download: Download full-size image](#)

Fig. 7. (a) and (b) show the effect of the parameter λ_3 on the clustering results of the BBCSport and 3-Sources datasets, respectively. Similarly, (c) and (d) illustrate the impact of the parameter p on the clustering results of the same datasets.

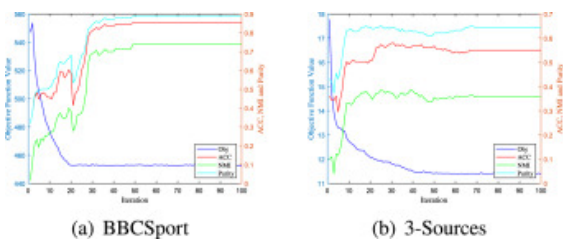
Table 10. Parameter settings correspond to the clustering results in [Tables 4](#) and [5](#).

Dataset				
3-Sources	3	10	10	1
BBCSport	0.1	300	0.1	1
BDGP(0.3)	0.001	0.001	0.1	1
BDGP(0.5)	0.001	0.001	2	1
BDGP(0.7)	0.001	0.001	3	1

Dataset	λ_1	λ_2	λ_3	p
HW(0.3)	0.01	100	0.5	1
HW(0.5)	0.1	100	0.1	1
HW(0.7)	0.01	0.01	0.1	1
mnist4(0.3)	0.01	0.001	0.5	1
mnist4(0.5)	0.001	0.01	0.1	1
mnist4(0.7)	0.001	0.1	1	1

5.4. Convergence analysis

To substantiate the convergence properties of our proposed algorithm, we designed a pair of experiments that scrutinize the interplay among objective function values, clustering performance metrics, and iteration count. Fig. 8 delineates the convergence trajectories of the objective function over iterative refinements on two distinct datasets: the BBCSport dataset is depicted in subfigure (a), while the 3-Sources dataset is represented in subfigure (b). The depicted trajectories reveal an inflection point beyond which the objective function values plateau, corroborating that our algorithm consistently arrives at a local minimum within a bounded number of iterations. The clustering metrics also gradually stabilize after a finite number of iterations. All of these confirms that our proposed algorithm satisfies our theoretical analysis with significant convergence properties.



[Download: Download high-res image \(180KB\)](#)

[Download: Download full-size image](#)

Fig. 8. Convergence curves on BBCSport dataset(a) and 3-Sources dataset(b).

6. Conclusion and future work

In this paper, we propose an innovative joint model that simultaneously handles the missing data reconstruction, representation learning and clustering problems. We use the consistent affinity matrix to construct a hypergraph to reconstruct missing data, which can better explore the local structure and higher-order information between sample points. Furthermore, unlike most methods that require additional clustering algorithms, such as K-means, to obtain the final clustering indicator matrix, we use non-negative matrix factorization with orthogonality constraints to equate K-means approach, which simplifies clustering process by eliminating post-processing. The use of tensor Schatten p -norm for the tensor composed of the clustering label matrices of different views can better exploit the complementary information and low-rank structure among these views. Extensive experiments are conducted on various datasets to verify the superiority of our method.

Nevertheless, it still has some limitations. A high computational complexity makes it difficult to be directly applied to clustering on larger datasets, which limits it in some real applications that require efficient processing of large data, such as social network analysis, image segmentation and object recognition tasks, etc.

In future research, we will aim to reduce the computational complexity to make it suitable for application scenarios that require analysis of larger-scale data.

CRediT authorship contribution statement

Jin Chen: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, [Data curation](#). **Huafu Xu:** Writing – review & editing, Funding acquisition. **Jingjing Xue:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition. **Quanxue Gao:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Cheng Deng:** Writing – review & editing. Ziyu [Lv](#): Visualization, Validation, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the anonymous reviewers and AE for their constructive comments and suggestions. This work is supported in part by Open Project Program of Guangxi Key Laboratory of Digital Infrastructure under Grant Number [GXDIOP2023010](#), in part by National NSFC under Grants [62176203](#), in part by the Postdoctoral Fellowship Program of CPSF under Grant Number [GZC20241318](#), in part by the Open Projects Program of State Key Laboratory of Multimodal Artificial Intelligence Systems under Grant Number [MAIS2024114](#), and in part by the Fundamental Research Funds for the Central Universities, China, the Innovation Fund of Xidian University .




[Recommended articles](#)

Data availability

Data will be made available on request.

References

- [1] Xue J., Nie F., Liu C., Wang R., Li X.
Co-clustering by directly solving bipartite spectral graph partitioning
IEEE Trans. Cybern. (2024)
[Google Scholar ↗](#)
- [2] Sorkhabi L.B., Gharehchopogh F.S., Shahamfar J.
A systematic approach for pre-processing electronic health records for mining: Case study of heart disease
Int. J. Data Min. Bioinform., 24 (2) (2020), pp. 97-120
[View in Scopus ↗](#) [Google Scholar ↗](#)
- [3] Rahnema N., Gharehchopogh F.S.
An improved artificial bee colony algorithm based on whale optimization algorithm for data clustering
Multimedia Tools Appl., 79 (43) (2020), pp. 32169-32194
[Crossref ↗](#) [View in Scopus ↗](#) [Google Scholar ↗](#)

- [4] Piri J., Mohapatra P., Acharya B., Gharehchopogh F.S., Gerogiannis V.C., Kanavos A., Manika S.
Feature selection using artificial gorilla troop optimization for biomedical data: A case analysis with COVID-19 data
Mathematics, 10 (15) (2022), p. 2742
[Crossref ↗](#) [View in Scopus ↗](#) [Google Scholar ↗](#)
- [5] Gharehchopogh F.S., Khargoush A.A.
A chaotic-based interactive autodidactic school algorithm for data clustering problems and its application on COVID-19 disease detection
Symmetry, 15 (4) (2023), p. 894
[Crossref ↗](#) [View in Scopus ↗](#) [Google Scholar ↗](#)
- [6] Li L., Wan Z., He H.
Incomplete multi-view clustering with joint partition and graph learning
IEEE Trans. Knowl. Data Eng., 35 (1) (2021), pp. 589-602
[Crossref ↗](#) [View in Scopus ↗](#) [Google Scholar ↗](#)
- [7] Cui J., Fu Y., Huang C., Wen J.
Low-rank graph completion-based incomplete multiview clustering
IEEE Trans. Neural Netw. Learn. Syst. (2022)
[Google Scholar ↗](#)
- [8] Wang H., Wang Q., Miao Q., Ma X.
Joint learning of data recovering and graph contrastive denoising for incomplete multi-view clustering
Inf. Fusion, 104 (2024), Article 102155
 [View PDF](#) [View article](#) [View in Scopus ↗](#) [Google Scholar ↗](#)
- [9] Liu C., Wu S., Li R., Jiang D., Wong H.-S.
Self-supervised graph completion for incomplete multi-view clustering
IEEE Trans. Knowl. Data Eng. (2023)
[Google Scholar ↗](#)
- [10] Li X., Sun Y., Sun Q., Ren Z., Sun Y.
Cross-view graph matching guided anchor alignment for incomplete multi-view clustering
Inf. Fusion, 100 (2023), Article 101941
 [View PDF](#) [View article](#) [Google Scholar ↗](#)
- [11] Wang Z., Li L., Ning X., Tan W., Liu Y., Song H.
Incomplete multi-view clustering via structure exploration and missing-view inference
Inf. Fusion, 103 (2024), Article 102123
 [View PDF](#) [View article](#) [View in Scopus ↗](#) [Google Scholar ↗](#)
- [12] Wen J., Yan K., Zhang Z., Xu Y., Wang J., Fei L., Zhang B.
Adaptive graph completion based incomplete multi-view clustering
IEEE Trans. Multimed., 23 (2020), pp. 2493-2504
[Google Scholar ↗](#)
- [13] Wong W.K., Liu C., Deng S., Fei L., Li L., Lu Y., Wen J.
Neighbor group structure preserving based consensus graph learning for incomplete multi-view clustering
Inf. Fusion, 100 (2023), Article 101917

 [View PDF](#) [View article](#) [View in Scopus ↗](#) [Google Scholar ↗](#)

- [14] Chen Z., Fu L., Yao J., Guo W., Plant C., Wang S.
Learnable graph convolutional network and feature fusion for multi-view learning
Inf. Fusion, 95 (2023), pp. 109-119

 [View PDF](#) [View article](#) [View in Scopus ↗](#) [Google Scholar ↗](#)

- [15] Trivedi A., Rai P., Daumé H. III, DuVall S.L.
Multiview clustering with incomplete views
NIPS Workshop, Vol. 224, Citeseer (2010), pp. 1-8

[Crossref ↗](#) [Google Scholar ↗](#)

- [16] Wen J., Xu Y., Liu H.
Incomplete multiview spectral clustering with adaptive graph learning
IEEE Trans. Cybern., 50 (4) (2018), pp. 1418-1429

[Google Scholar ↗](#)

- [17] Xia W., Gao Q., Wang Q., Gao X.
Tensor completion-based incomplete multiview clustering
IEEE Trans. Cybern., 52 (12) (2022), pp. 13635-13644

[Crossref ↗](#) [View in Scopus ↗](#) [Google Scholar ↗](#)

- [18] Xia W., Gao Q., Wang Q., Gao X., Ding C., Tao D.
Tensorized bipartite graph learning for multi-view clustering
IEEE Trans. Pattern Anal. Mach. Intell., 45 (4) (2022), pp. 5187-5202

[Google Scholar ↗](#)

- [19] S.-Y. Li, Y. Jiang, Z.-H. Zhou, Partial multi-view clustering, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 28, No. 1, 2014.

[Google Scholar ↗](#)

- [20] H. Zhao, H. Liu, Y. Fu, Incomplete multi-modal visual data grouping, in: IJCAI, 2016, pp. 2392–2398.

[Google Scholar ↗](#)

- [21] N. Xu, Y. Guo, X. Zheng, Q. Wang, X. Luo, Partial multi-view subspace clustering, in: Proceedings of the 26th ACM International Conference on Multimedia, 2018, pp. 1794–1801.

[Google Scholar ↗](#)

- [22] Shao W., He L., Yu P.S.
Multiple incomplete views clustering via weighted nonnegative matrix factorization with regularization
Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer (2015), pp. 318-334

[Crossref ↗](#) [View in Scopus ↗](#) [Google Scholar ↗](#)

- [23] Hu M., Chen S.
Doubly aligned incomplete multi-view clustering
(2019)

arXiv preprint [arXiv:1903.02785 ↗](#)

[Google Scholar ↗](#)

- [24] Rai N., Negi S., Chaudhury S., Deshmukh O.
Partial multi-view clustering using graph regularized NMF

[View in Scopus ↗](#) [Google Scholar ↗](#)

- [25] Shao W., He L., Lu C.-t., Philip S.Y.

Online multi-view clustering with incomplete views

2016 IEEE International Conference on Big Data, Big Data, IEEE (2016), pp. 1012-1017

[View in Scopus ↗](#) [Google Scholar ↗](#)

- [26] M. Hu, S. Chen, One-pass incomplete multi-view clustering, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, No. 01, 2019, pp. 3838–3845.

[Google Scholar ↗](#)

- [27] Gao H., Peng Y., Jian S.

Incomplete multi-view clustering

Intelligent Information Processing VIII: 9th IFIP TC 12 International Conference, IIP 2016, Melbourne, VIC, Australia, November 18-21, 2016, Proceedings 9, Springer (2016), pp. 245-255

[Crossref ↗](#) [View in Scopus ↗](#) [Google Scholar ↗](#)

- [28] Wang H., Zong L., Liu B., Yang Y., Zhou W.

Spectral perturbation meets incomplete multi-view data

(2019)

arXiv preprint [arXiv:1906.00098 ↗](#)

[Google Scholar ↗](#)

- [29] J. Wen, Z. Zhang, Z. Zhang, L. Zhu, L. Fei, B. Zhang, Y. Xu, Unified tensor framework for incomplete multi-view clustering and missing-view inferring, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35, No. 11, 2021, pp. 10273–10281.

[Google Scholar ↗](#)

- [30] Li Z., Tang C., Zheng X., Liu X., Zhang W., Zhu E.

High-order correlation preserved incomplete multi-view subspace clustering

IEEE Trans. Image Process., 31 (2022), pp. 2067-2080

[Crossref ↗](#) [View in Scopus ↗](#) [Google Scholar ↗](#)

- [31] Sun L., Wen J., Liu C., Fei L., Li L.

Balance guided incomplete multi-view spectral clustering

Neural Netw., 166 (2023), pp. 260-272

 [View PDF](#) [View article](#) [View in Scopus ↗](#) [Google Scholar ↗](#)

- [32] S. Wang, X. Liu, L. Liu, W. Tu, X. Zhu, J. Liu, S. Zhou, E. Zhu, Highly-efficient incomplete large-scale multi-view clustering with consensus bipartite graph, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 9776–9785.

[Google Scholar ↗](#)

- [33] Zhou B., Ji J., Gu Z., Zhou Z., Ding G., Feng S.

One-step graph-based incomplete multi-view clustering

Multimedia Syst., 30 (1) (2024), p. 32


[Crossref ↗](#) [Google Scholar ↗](#)

- [34] Zhou D., Huang J., Schölkopf B.


Learning with hypergraphs: Clustering, classification, and embedding

Adv. Neural Inf. Process. Syst., 19 (2006)

[Google Scholar ↗](#)

- [35] Gao Q., Zhang P., Xia W., Xie D., Gao X., Tao D.
Enhanced tensor RPCA and its application
IEEE Trans. Pattern Anal. Mach. Intell., 43 (6) (2020), pp. 2133-2140
[Google Scholar ↗](#)
- [36] Perraudin N., Paratte J., Shuman D., Martin L., Kalofolias V., Vandergheynst P., Hammond D.K.
GSPBOX: A toolbox for signal processing on graphs
(2014)
arXiv preprint [arXiv:1408.5781 ↗](#)
[Google Scholar ↗](#)
- [37] C. Ding, T. Li, W. Peng, H. Park, Orthogonal nonnegative matrix t-factorizations for clustering, in: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2006, pp. 126–135.
[Google Scholar ↗](#)
- [38] Xia W., Gao Q., Wang Q., Gao X., Ding C., Tao D.
Tensorized bipartite graph learning for multi-view clustering
IEEE Trans. Pattern Anal. Mach. Intell., 45 (4) (2023), pp. 5187-5202
[Crossref ↗](#) [View in Scopus ↗](#) [Google Scholar ↗](#)
- [39] Gao Q., Xu S., Chen F., Ding C., Gao X., Li Y.
 R_1 -2-DPCA and face recognition
IEEE Trans. Cybern., 49 (4) (2018), pp. 1212-1223
[View in Scopus ↗](#) [Google Scholar ↗](#)
- [40] Li J., Gao Q., Wang Q., Xia W.
Tensorized label learning on anchor graph
Wooldridge M.J., Dy J.G., Natarajan S. (Eds.), Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2024, February 20-27, 2024, Vancouver, Canada, AAAI Press (2024), pp. 13537-13544, [10.1609/AAAI.V38I12.29257 ↗](#)
[View in Scopus ↗](#) [Google Scholar ↗](#)
- [41] Wang J., Ma Z., Nie F., Li X.
Efficient discrete clustering with anchor graph
IEEE Trans. Neural Networks Learn. Syst., 35 (10) (2024), pp. 15012-15020
[Crossref ↗](#) [View in Scopus ↗](#) [Google Scholar ↗](#)
- [42] Nie F., Xue J., Yu W., Li X.
Fast clustering with anchor guidance
IEEE Trans. Pattern Anal. Mach. Intell., 46 (4) (2024), pp. 1898-1912
[Crossref ↗](#) [View in Scopus ↗](#) [Google Scholar ↗](#)
- [43] Lei Y., Niu Z., Wang Q., Gao Q., Yang M.
Anchor graph-based multiview spectral clustering
Neurocomputing, 583 (2024), Article 127579
 [View PDF](#) [View article](#) [View in Scopus ↗](#) [Google Scholar ↗](#)
- [44] Cai X., Wang H., Huang H., Ding C.
Joint stage recognition and anatomical annotation of drosophila gene expression patterns
Bioinformatics, 28 (12) (2012), pp. i16-i24

[Crossref ↗](#) [View in Scopus ↗](#) [Google Scholar ↗](#)

- [45] van Breukelen M., Duin R.P., Tax D.M., Den Hartog J.
Handwritten digit recognition by combined classifiers
Kybernetika, 34 (4) (1998), pp. 381-386
[View in Scopus ↗](#) [Google Scholar ↗](#)
- [46] Deng L.
The mnist database of handwritten digit images for machine learning research [best of the web]
IEEE Signal Process. Mag., 29 (6) (2012), pp. 141-142
[View in Scopus ↗](#) [Google Scholar ↗](#)
- [47] D. Greene, P. Cunningham, Practical solutions to the problem of diagonal dominance in kernel document clustering, in: Proceedings of the 23rd International Conference on Machine Learning, 2006, pp. 377–384.
[Google Scholar ↗](#)
- [48] J. Huang, F. Nie, H. Huang, Spectral rotation versus k-means in spectral clustering, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 27, No. 1, 2013, pp. 431–437.
[Google Scholar ↗](#)
- [49] Ng A., Jordan M., Weiss Y.
On spectral clustering: Analysis and an algorithm
Adv. Neural Inf. Process. Syst., 14 (2001)
[Google Scholar ↗](#)
- [50] F. Nie, J. Li, X. Li, et al., Parameter-free auto-weighted multiple graph learning: a framework for multiview clustering and semi-supervised classification, in: IJCAI, Vol. 9, 2016.
[Google Scholar ↗](#)
- [51] R. Xia, Y. Pan, L. Du, J. Yin, Robust multi-view spectral clustering via low-rank and sparse decomposition, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 28, No. 1, 2014.
[Google Scholar ↗](#)
- [52] Lv Z., Gao Q., Zhang X., Li Q., Yang M.
View-consistency learning for incomplete multiview clustering
IEEE Trans. Image Process., 31 (2022), pp. 4790-4802
[Crossref ↗](#) [View in Scopus ↗](#) [Google Scholar ↗](#)
- [53] Deng S., Wen J., Liu C., Yan K., Xu G., Xu Y.
Projective incomplete multi-view clustering
IEEE Trans. Neural Netw. Learn. Syst. (2023)
[Google Scholar ↗](#)
- [54] Entezami A., Sarmadi H., Behkamal B.
Long-term health monitoring of concrete and steel bridges under large and missing data by unsupervised meta learning
Eng. Struct., 279 (2023), Article 115616
 [View PDF](#) [View article](#) [View in Scopus ↗](#) [Google Scholar ↗](#)
- [55] Behkamal B., Entezami A., De Michele C., Arslan A.

Investigation of temperature effects into long-span bridges via hybrid sensing and supervised regression models

Remote Sens., 15 (14) (2023), p. 3503

[Crossref ↗](#) [View in Scopus ↗](#) [Google Scholar ↗](#)

[56] Felzenszwalb P., McAllester D., Ramanan D.

A discriminatively trained, multiscale, deformable part model

2008 IEEE Conference on Computer Vision and Pattern Recognition, Ieee (2008), pp. 1-8

[Crossref ↗](#) [Google Scholar ↗](#)

Cited by (2)

Discriminative approximate regression projection for feature extraction

2025, Information Fusion

[Show abstract ↘](#)

Advanced unsupervised learning: a comprehensive overview of multi-view clustering techniques

[↗](#)

2025, Artificial Intelligence Review



Jin Chen received the B. Eng. degree in Electronic Information Engineering at Jiangsu Normal University, Xuzhou, China, in 2023. She is currently pursuing the M.S. degree in Electronic Information at Xidian University, Xi'an, China. Her research interests include pattern recognition and machine learning.



Huafu Xu, Ph.D. degree, graduated from Shanghai Jiaotong University in 2016, is now presiding over the institute of Big Data Development of Guangxi Zhuang Autonomous Region Information Center, with research interests in big data and cloud computing.



Jingjing Xue received the Ph.D. degree from the School of Computer Science and School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China, in 2023. She is currently a lecturer with the School of Telecommunications Engineering, Xidian University, China. Her current research interests are machine learning and its applications, including matrix factorization, clustering and model optimization.



Quanxue Gao received the B.Eng. degree from Xi'an Highway University, Xi'an, China, in 1998, the M.S. degree from the Gansu University of Technology, Lanzhou, China, in 2001, and the Ph.D. degree from Northwestern Polytechnical University, Xi'an China, in 2005. He was an associate research with the Biometrics Center, The Hong Kong Polytechnic University, Hong Kong from 2006 to 2007. From 2015 to 2016, he was a visiting scholar with the department of computer science, The University of Texas at Arlington, Arlington USA. He is currently a professor with the School of Telecommunications Engineering, Xidian University, and also a key member of State Key Laboratory of Integrated Services Networks. He has authored around 100 technical articles in refereed journals and proceedings, including IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Image Processing, IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Cybernetics, CVPR, NeurIPS, AAAI, and IJCAI. His current research interests include pattern recognition and machine learning.



Cheng Deng received the B.E., M.S., and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China. He is currently a Full Professor with the School of Electronic Engineering at Xidian University. His research interests include computer vision, pattern recognition, and information hiding. He is the author and coauthor of more than 200 scientific articles at top venues, including IEEE TNNLS, TIP, TCYB, TMM, TSMC, ICCV, CVPR, ICML, NIPS, IJCAI, and AAAI.



Ziyu Lv received the B.S. degree in communication engineering from the Xi'an University of Posts and telecommunications, Xi'an, China, in 2020, the master's degree in information and communication engineering from Xidian University, Xi'an, China, in 2023. She is currently an engineer with the Automotive Engineering Research Institute of the BYD Company. Her research interests include multi-view clustering and partial multi-view clustering.

- 1 [Publicly Available] <http://ranger.uta.edu/%7eheng/Drosophila/> ↗.
- 2 [Publicly Available] <https://aistudio.baidu.com/datasetdetail/137482/0> ↗.
- 3 [Publicly Available] <http://yann.lecun.com/exdb/mnist/> ↗.
- 4 [Publicly Available] <https://github.com/GPMVCDummy/GPMVC/tree/master/partialMV/PVC/recreateResults/data> ↗.
- 5 [Publicly Available] <https://github.com/GPMVCDummy/GPMVC/tree/master/partialMV/PVC/recreateResults/data> ↗.
- 6 <http://erdos.ucd.ie/datasets/3sources.html> ↗.

[View Abstract](#)



All content on this site: Copyright © 2025 Elsevier B.V., its licensors, and contributors. All rights are reserved, including those for text and data mining, AI training, and similar technologies. For all open access content, the relevant licensing terms apply.

