An adaptive weighted self-representation method for incomplete multi-view clustering

Lishan Feng[a], Guoxu Zhou[b] and Jingya Chang [a,*]

[a]School of Mathematics and Statistics, Guangdong University of Technology, Guangzhou, 510520, China
[b]School of Automation, Guangdong University of Technology, Guangzhou, 510006, China

ARTICLE INFO

Keywords:
Multi-view clustering
Incomplete data
Subspace learning

ABSTRACT

For multi-view learning problems coming from reality, it is common that the data are incomplete due to lost files, poor data entry and so on. Thus, we consider the incomplete multi-view clustering problem. In the clustering process, the influence of each view of one object may be different and the data are often low rank. On the other hand, self-representation subspace method works well on multi-view clustering problem. Taking these factors into account, we propose an adaptive weighted self-representation subspace clustering (AWSR) method for solving the incomplete multi-view clustering problem. The AWSR method employs a weight matrix to measure the contribution of each missing view, and seeks for the low rank coefficient matrix to enhance the clustering effect. Moreover, the regularization terms are properly designed to eliminate the possible data noise and keep the convexity property of the objective function. The convergence analysis is provided for the alternate minimization algorithm of the AWSR model. Meanwhile, numerical experiments of AWSR method are implemented on five real datasets. The numerical performance shows that the AWSR method is effective and delivers superior results when compared to other six common approaches considering the clustering accuracy, normalized mutual information and purity.

1. Introduction

One object can be described from many angles. For example, we can identify a leaf by its features like shape, color, texture structure and so on. A person is distinguished from appearance (images), interpersonal relationship (graph) and social behavior, etc. On the other hand, the fast development of science and technology enable us many ways to get the information of an object. One can receive news from different news organization or medias. The same semantic can be expressed by different languages via a smart phone application program. As large amount of information is created and mixed day by day, an important task is to find the information of the same item from the miscellaneous data. Such kind of problems are categorized to multi-view clustering. In reality, due to human error and machine damage, the collected information is often missing. Data missing can be divided into three categories: feature missing, view missing and fusion missing in common sense. Hence, how to analyze object in incomplete multi-view data is crucial and valuable in real-world applications. In this paper, we study the multi-view clustering problem with incomplete data.

The existing multi-view methods employ the complementary information [5, 15, 24, 25, 28, 33, 34, 40] and different view information [2, 8, 25, 42] to build the model. Based on different mechanisms, the existing multi-view clustering approaches can be divided into the following five categories. The first category are based on collaborative learning [2, 42]. Each view uses own information to promote other view's learning. The clustering results of all views tend to be consistent after certain iteration steps, and the consensus information is maximized. The second kind of approaches are based on graph learning [15, 19, 22, 35, 36, 39, 41]. These methods first construct the initial graph of each view and then fuse all graph into a consistent graph via learning. Finally, graph segmentation or other spectral graph techniques are utilized to the segmentation problem. The third category approaches are based on subspace learning [11, 20, 27, 31, 33, 34, 38]. Their main idea is assuming that various low-dimensional sub-spaces can be extracted from high-dimensional data points. Thus the high-dimensional data can be divided into a cluster according to its subspace, and the low-dimensional features in the potential subspace is discovered and clustered. The subspace learning approaches are implemented by non-negative matrix factorization [19, 31, 34, 38] or self-representation [11, 20, 27]. For methods of non-negative matrix factorization, the high-dimensional data are divided into two low-dimensional matrices (the unitary matrix and the feature matrix), and the non-negative matrix factorization methods are employed to seek the consistent feature matrix of all views. For method of self-representation, the core idea is to reconstruct the data itself via the mutual representation of the data and to construct the similarity between the samples based on the reconstruction coefficient matrix. Then it mines the relationship among the samples and realizes the mapping representation of data from the view dimension to the sample dimension. The fourth category approaches are based on multi-kernel learning [24, 28]. These methods give each view a predefined kernel function, and seek information among views to construct a consensus kernel function to improve the clustering effect. The last category are deep learning-based methods [4, 5, 18, 25, 37]. This type of

*Corresponding author
lsfeng2022@163.com (L. Feng); gx.zhou@gdut.edu.cn (G. Zhou); jychang@gdut.edu.cn (J.C.)
ORCID(s): 0000-0002-5296-2813 (J.C.)

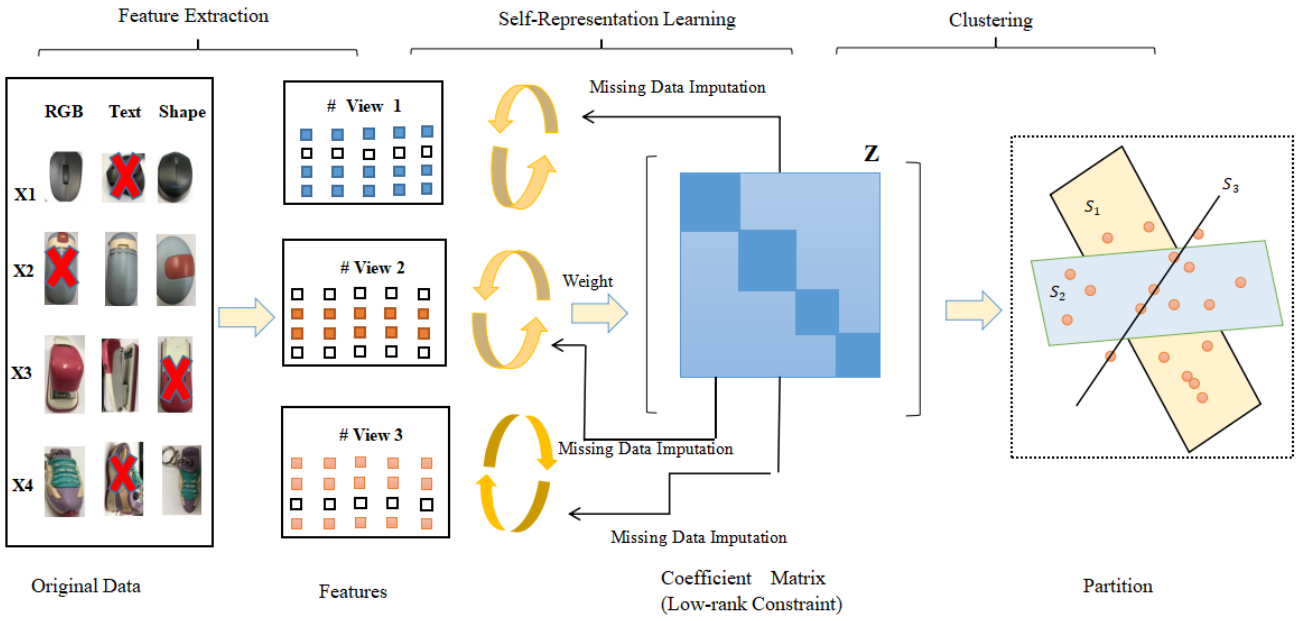

Figure 1: Overview of the algorithm. Here is an example of three views, including RGB, Text and Shape modalities. From left to right, first, feature vectors are extracted from the incomplete multi-view data X. Then the low-rank regularization constraint is imposed on the consensus coefficient matrix Z. Meanwhile, weighted self-representation learning and missing data imputation are jointly performed to obtain low-rank consensus coefficient matrix. Finally cluster assignments are given based on the resulting consensus coefficient matrix.

approach is commonly conducted based on either shared feature representation or jointly feature representation.

Among the above five categories, the subspace clustering approach has attracted much attention. Guo et al. [11] proposed a block-diagonal multi-view method, which obtained the coefficient matrix of each view, and then the consensus coefficient matrix is gained. Besides, they proposed a new block diagonal multi-view information complementary regularizer. Huang et al. [14] proposed to employ sparse representation to construct a sparse hypergraph, and then bring the hypergraph in the framework of non-negative matrix factorization to get a regularization term. Li et al. [21] combined the hypergraph with the $\ell_{2,1}$-norm of self-representation to express the complex information among data and eliminate the impact of noise. Hu et al. [13] strengthened the clustering effect by regularizing both the unitary matrix and the latent feature matrix on the basis of weighted non-negative matrix. Kang et al. [17] pointed out that affinity matrix of the self-representation subspace methods can reflect the relationship between data entries. Therefore, self-representation methods have better clustering effect than that of the non-negative matrix decomposition methods when dealing with incomplete multi-view data. Liu et al. [27] also used the ablation study to confirm that the self-representation subspace methods have better clustering effect than non-negative matrix factorization approaches for incomplete multi-view clustering. In addition, Liu et al. also adapted self-representation to incomplete data by jointly performing data imputation and self-representation learning, which solved the contradiction caused by incomplete multi-view data structure.

The existing self-representation methods treat each view as the same weight, which is inconsistent with reality. We propose an adaptive weighted self-representation subspace clustering (AWSR) method for solving the incomplete multi-view clustering problem. The idea of this approach is intuitively represented in Figure 1. Firstly, we extract the corresponding feature vectors for the incomplete multi-view data. Then, we give different weight matrix of each view to learn the consensus coefficient matrix. At the same time, the nuclear norm constraint is applied to the consensus coefficient matrix. Missing data imputation and self-representation learning are jointly performed until the algorithm converges. Finally, we cluster the data according to the coefficient matrix we have learned. Our adaptive weighted self-representation subspace clustering (AWSR) method has the following contributions:

(1) AWSR enriches the application of self-representation subspace clustering methods in incomplete multi-view data.

(2) It is clear that different views have various importance for learning consensus coefficient matrix. The adaptive weight matrices are applied to our self-representation subspace clustering methods.

(3) AWSR shares the advantages of both Least Squares Regression (LSR) and Low-Rank Representation (LRR). For the consensus coefficient matrix Z, it considers low rank of Z and imposes the F-norm constraint on Z to eliminate the noise influence of missing entries.

(4) AWSR performs well on real datasets.

Table 1
Summary of the Notations

Notation	Description
X_i	The data matrix for ith view.
Z	The consensus coefficient matrix for all the views.
$W^{(i)}$	The weight matrix for the ith view.
Z_i	The coefficient matrix for the ith view.
v	The number of the views.
U	The unitary matrix.
C	The complex matrix.

2. Notations and related work

Given a set of data vectors $X = [X_1, \cdots, X_k] \in R^{d\times n}$. Let X_i be the collection of n_i data vectors drawn from the subspace S_i, $n = \sum_{i=1}^{k} n_i$. The task of subspace clustering is to segment the data according to the underlying subspaces [23]. We visualize the process of subspace clustering through Figure 2.

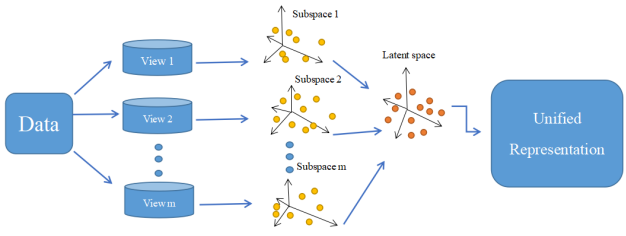

Figure 2: Subspace clustering algorithm. m views (angles) are extracted from the data to obtain a subspace for each view. Then dimension of each subspace is reduced to obtain a latent feature subspace.

The purpose of low rank representation is that it uses the correlation structure of the data to find a low rank representation, i.e., constructing a low rank coefficient matrix $Z \in R^{n\times n}$ of the original data matrix for clustering. The model of low rank representation can be expressed as

$$\min\ rank(Z), \quad \text{s.t. } X = XZ,$$

where $rank(Z)$ is the rank of matrix Z. The above problem is NP-hard. Therefore, we can transform the rank minimization constraint into the nuclear norm constraint [23]

$$\min\ \|Z\|_*, \quad \text{s.t. } X = XZ,$$

where the nuclear norm is defined as $\|Z\|_* = \sum_i^n \sigma_i$ and σ_i is the ith singular value of matrix Z. The target of sparse subspace clustering is that it reflects high-dimensional data to a low-dimensional subspace, and constructs a sparse coefficient matrix $Z \in R^{n\times n}$ for clustering. Sparse subspace clustering is described by the following problem

$$\min\ \|Z\|_0, \quad \text{s.t. } X = XZ,\ \text{diag}(Z) = 0,$$

where $\|Z\|_0$ is ℓ_0-norm of Z, i.e., the number of nonzero elements Z and the symbol diag(Z) indicates the diagonal elements of the matrix Z. Whereafter, we use diag(Z) to stand for either the diagonal elements of Z or the diagonal matrix induced by the the diagonal elements of Z, which we believe will not create ambiguous when combined with the context. The ℓ_0-norm of Z is discontinuous and the above optimization problem is non-convex and NP-hard. It is then transformed as [7]

$$\min\ \|Z\|_1, \quad \text{s.t. } X = XZ,\ \text{diag}(Z) = 0,$$

where ℓ_1-norm of Z is calculated as $\|Z\|_1 = \sum_{i=1}^{n}\sum_{j=1}^{n}|Z_{ij}|$.

Both low rank representation and sparse subspace clustering methods mentioned above belong to self-representation subspace clustering approaches. In addition, self-representation subspace clustering approach includes the least square regression (LSR) and so on. The least square regression model can be written as

$$\min\ \|Z\|_F, \quad \text{s.t. } X = XZ,\ \text{diag}(Z) = 0,$$

where the F-norm of Z is calculated as $\|Z\|_F = (\sum_{i=1}^{n}\sum_{j=1}^{n} Z_{ij}^2)^{\frac{1}{2}}$.

In single-view data set, self-representation subspace clustering approaches can be uniformly expressed as follows

$$\min_{Z}\ \ell(X - XZ) + \lambda\mathcal{R}(Z), \quad s.t.\ \Omega(Z),$$

where ℓ and $\mathcal{R}$ mean some norm and regularization term respectively.

In the complete multi-view, self-representation subspace clustering methods can be uniformly represented as

$$\min_{Z_*,\{Z_i\}_{i=1}^{v}} \sum_{i=1}^{v} \beta_i \ell(X_i - X_i Z_i) + \lambda\mathcal{R}(Z_*, \{Z_i\}_{i=1}^{v}),$$
$$\text{s.t. } \Omega(Z_*, \{Z_i\}_{i=1}^{v}, \{\beta\}_{i=1}^{v}),$$

where Z_* is the consensus coefficient matrix of all views.

3. The Proposed AWSR Model

Zhao et al. and Liu et al. in [17, 27] showed that self-representation subspace clustering performs well for incomplete multi-view data. For self-representation subspace clustering, Lu et al. [29] introduced F-norm and diagonal zero constraint to limit consensus coefficient matrix Z. The model is expressed as

$$\min_{Z}\ \|X - XZ\|_F^2 + \lambda\|Z\|_F^2, \quad \text{s.t. diag}(Z) = 0.$$

Kang et al. [16] extended the above model to multi-view clustering and obtain the following framework:

$$\min_{Z_i}\ (1/v)\sum_{i=1}^{v}\|X_i - X_i Z_i\|_F^2 + \lambda\|Z_i\|_F^2, \quad \text{s.t. diag}(Z) = 0.$$

When data information is incomplete, the incomplete data matrix is explicitly expressed and involved in the optimization process [27]:

$$\begin{aligned}\min_{Z,\{X_i^{(m)}\}_{i=1}^{v},F} \ & (1/v)\sum_{i=1}^{v}\left\|[X_i^{(0)},X_i^{(m)}]-[X_i^{(0)},X_i^{(m)}]Z\right\|_F^2 \\ & +\lambda\|Z\|_F^2+\gamma\left\|Z-F^\top F\right\|_F^2, \\ & \text{s.t. } \mathrm{diag}(Z)=0,\ FF^\top=I_k,\end{aligned}$$

where each view has the same weight, and all views share a consensus coefficient matrix Z.

Since different views have various contribution to the consensus coefficient matrix, we introduce the weight matrix $W^{(i)} \in R^{n\times n}$ to construct Z. In order to capture the data correlation structure in the same space and effectively solve the minimization problem of each data point, we place F-norm constraint on the coefficient matrix. At the same time, we impose a kernel norm constraint on Z to enhance the clustering effect. Finally, the overall objective function can be formulated as

$$\begin{aligned}\min_{Z,\{X_i^{(m)}\}_{i=1}^{v}} \ & \sum_{i=1}^{v}\left\|([X_i^{(0)},X_i^{(m)}]-[X_i^{(0)},X_i^{(m)}]Z)W^{(i)^\top}\right\|_F^2 \\ & +\gamma\|Z\|_*+(\lambda/2)\|Z\|_F^2, \\ & \text{s.t. } \mathrm{diag}(Z)=0,\end{aligned} \tag{1}$$

where

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

γ, λ are positive trade-off parameters and $X_i^{(0)}$ and $X_i^{(m)}$ refer to observed and missing entries of the ith data view. It can be observed from Eq.(1) that both observed and missing data have participated in self-representation learning, and missing data will be reconstructed in the learning process of Z. At last, we adopt the spectral clustering method to complete the cluster assignment. The Laplacian matrix in the spectral clustering approach is defined as $L=(|Z|+|Z^\top|)/2$. Here $|Z|$ is a matrix whose elements are the absolute values of elements of matrix Z.

4. Computation of the AWSR model

We introduce a separated variable J and convert Eq.(1) to the following objective function

$$\begin{aligned}\min_{J,Z,\{X_i^{(m)}\}_{i=1}^{v}} \ & \sum_{i=1}^{v}\left\|([X_i^{(0)},X_i^{(m)}]-[X_i^{(0)},X_i^{(m)}]J)W^{(i)^\top}\right\|_F^2 \\ & +\gamma\|Z\|_*+(\lambda/2)\|Z\|_F^2+(\alpha/2)\|J-Z\|_F^2, \\ & \text{s.t. } \mathrm{diag}(Z)=0.\end{aligned} \tag{2}$$

The model (1) and model (2) are equivalent when $\alpha > 0$ is large enough. The strong convexity property function implies that the solution is stable and unique. We apply the alternate strategy [26] to calculation of the above optimization problem.

4.1. Sub-Problem for J:

To update J, we fix the value of all other variables. The sub-problem related with J is

$$\begin{aligned}\min_J \ \psi(J)=&\sum_{i=1}^{v}\left\|([X_i^{(0)},X_i^{(m)}]-[X_i^{(0)},X_i^{(m)}]J)W^{(i)^\top}\right\|_F^2 \\ &+(\alpha/2)\|J-Z\|_F^2.\end{aligned} \tag{3}$$

The model (3) is equivalent to the following problem

$$\begin{aligned}\min_J \ \psi(J)=&Tr(\sum_{i=1}^{v}X_iJW^{(i)^\top}W^{(i)}J^\top X_i^\top \\ &-X_iJW^{(i)^\top}W^{(i)}X_i^\top-X_iW^{(i)^\top}W^{(i)}J^\top X_i) \\ &+(\alpha/2)Tr(J^\top J-J^\top Z-Z^\top J),\end{aligned}$$

in which $X_i=[X_i^{(o)},X_i^{(m)}]$. Take derivation of the objective function with respect to variable J, and let the derivation equal to zero. We get

$$\begin{aligned}&\sum_{i=1}^{v}X_i^\top X_iJW^{(i)^\top}W^{(i)}+(\alpha/2)J \\ &-\sum_{i=1}^{v}X_i^\top X_iW^{(i)^\top}W^{(i)}-(\alpha/2)Z=0.\end{aligned} \tag{4}$$

Since the matrix equation $AXB=C$ is equivalent to [30]

$$(B^T\otimes A)\mathrm{vec}(X)=\mathrm{vec}(C),$$

where $\otimes$ means Kronecker product, equation (4) equals to

$$\left[\sum_{i=1}^{v}W^{(i)^\top}W^{(i)}\otimes X_i^\top X_iJ+(\alpha/2)I\right]\mathrm{vec}(J)=\mathrm{vec}(C). \tag{5}$$

Here $C=\sum_{i=1}^{v}X_i^\top X_iW^{(i)^\top}W^{(i)}+(\alpha/2)Z$. The coefficient matrix in the linear equation system (5) is positive definite. Hence the vector $\mathrm{vec}(J)$ can be solved by the traditional conjugate gradient method. Finally, we get the matrix J by rearranging $\mathrm{vec}(J)$.

4.2. Sub-Problem for $X_i^{(m)}$:

To update $X_i^{(m)}$, we fix the value of all other variables. The sub-problem related with $X_i^{(m)}$ is

$$\min_{X_i^{(m)}}\ \psi(X_i^{(m)})=\sum_{i=1}^{v}\left\|([X_i^{(0)},X_i^{(m)}]-[X_i^{(0)},X_i^{(m)}]J)W^{(i)^\top}\right\|_F^2.$$

(6)

The Eq.(6) can be transformed to

$$\min_{X_i^{(m)}} \psi(X_i^{(m)}) = Tr([X_i^{(0)}, X_i^{(m)}]B_i[X_i^{(0)}, X_i^{(m)}]^\top), \quad \text{for } i = 1, \dots, \upsilon, \tag{7}$$

where

$$B_i = W^{(i)\top}W^{(i)} - JW^{(i)\top}W^{(i)} - W^{(i)\top}W^{(i)}J^\top + JW^{(i)\top}W^{(i)}J^\top.$$

Because X_i is divided into observed and missing data, i.e., $X_i = [X_i^{(0)}, X_i^{(m)}]$, we rewrite B_i in the form of block matrix. The Eq.(7) is equivalent to

$$\min_{X_i^{(m)}} \psi(X_i^{(m)}) = Tr([X_i^{(0)}, X_i^{(m)}]\begin{pmatrix} B_i^{(oo)} & B_i^{(om)} \\ B_i^{(mo)} & B_i^{(mm)} \end{pmatrix}[X_i^{(0)}, X_i^{(m)}]^\top). \tag{8}$$

The model (8) is further replaced by

$$\min_{X_i^{(m)}} \psi(X_i^{(m)}) = Tr(X_i^{(o)}(B_i^{(om)} + B_i^{(mo)\top})X_i^{(m)\top} + X_i^{(m)}B_i^{(mm)}X_i^{(m)\top}). \tag{9}$$

From the definition of B_i, we have

$$B_i^{(mm)} = \left(W^{(i)} - JW^{(i)}\right)^{(m)}\left[\left(W^{(i)} - JW^{(i)}\right)^{(m)}\right]^\top, \tag{10}$$

in which $(W^{(i)} - JW^{(i)})^{(m)}$ refers to the columns of $W^{(i)} - JW^{(i)}$ and corresponds to the weighted missing entries in the ith view. By taking the derivation of the objective function in (9) with respect to the variable $X_i^{(m)}$, we get

$$X_i^m B_i^{mm} = -X_i^o B_i^{om}. \tag{11}$$

Because $B_i^{(mm)}$ is a positive semi-definite matrix, we calculate the eigenvalue decomposition of matrix $B_i^{(mm)}$, i.e., $B_i^{(mm)} = U_i \wedge_i V_i^\top$. The numerical solution of equation system (11) is

$$X_i^m = -X_i^{(o)}B_i^{(om)}V_i \wedge_i^+ U_i^\top. \tag{12}$$

Here $\wedge_i^+$ is the pseudo inverse of $\wedge_i$.

4.3. Sub-Problem for Z:

To update Z, we fix the value of all other variables. We solve the following sub-problem:

$$\min_Z \psi(Z) = \gamma\|Z\|_* + (\lambda/2)\|Z\|_F^2 + (\alpha/2)\|J - Z\|_F^2, \quad \text{s.t. } \operatorname{diag}(Z) = 0. \tag{13}$$

The objective function in (13) is rearranged and we obtain the equivalent model as follows

$$\min_Z \psi(Z) = (\gamma/(\lambda+\alpha))\|Z\|_* + (1/2)\|Z - (\alpha/(\lambda+\alpha))J\|_F^2, \quad \text{s.t. } \operatorname{diag}(Z) = 0. \tag{14}$$

Denote the Lagrange function $\mathcal{L}(Z, y) = \psi(Z) + < y, \operatorname{diag}(Z) >$. Then the dual function of the Lagrange multiplier y is $g(y) = \inf_Z \mathcal{L}(Z, y)$. Because the optimization problem (14) is convex and satisfies the weak Slater condition, the strong duality property holds, which means

$$\sup_y \inf_Z \mathcal{L}(Z, y) = \mathcal{L}(Z^*, y^*) = \inf_Z \sup_y \mathcal{L}(Z, y).$$

We employ the Uzawa's algorithm to solve the dual problem of (14). Since the gradient of $g(y)$ is given by

$$\nabla g(y) = \frac{\partial \mathcal{L}(Z, y)}{\partial y} = \operatorname{diag}(Z).$$

The Uzawa's algorithm is achieved iteratively via the following scheme

$$\begin{cases} \mathcal{L}(Z^t, y^{t-1}) = \min_Z \mathcal{L}(Z, y^{t-1}), \\ y^t = y^{t-1} + \delta_k(\operatorname{diag}(Z^t)). \end{cases} \tag{15}$$

The parameter δ_k is a step size of the gradient direction. Next, we consider the minimization problem $\min_Z \mathcal{L}(Z, y^{t-1})$ in (15). By substituting the formula in (14) for $\psi(Z)$, we obtain

$$\begin{aligned} \mathcal{L}(Z, y^{t-1}) &= \psi(Z) + < y^{t-1}, \operatorname{diag}(Z) > \\ &= \psi(Z) + < Y^{t-1}, Z > \\ &= (1/2)\left\| Z - \left(\frac{\alpha}{\lambda+\alpha}J + Y^{t-1}\right)\right\|_F^2 \\ &\quad + (\gamma/(\lambda+\alpha))\|Z\|_*. \end{aligned} \tag{16}$$

Here Y is a matrix the same size with Z. Its diagonal elements equal to y and off diagonal elements are zero. Thus based on (16), the minimization problem in (15) can be efficiently solved by the singular value threshold operator [3] and the Uzawa's process is transformed into

$$\begin{cases} Z^t = \mathcal{D}_{\gamma/(\lambda+\alpha)}((\alpha/(\lambda+\alpha))J + Y^{t-1}) \\ y^t = y^{t-1} + \delta_k(\operatorname{diag}(Z^t)), \end{cases} \tag{17}$$

where

$$\begin{aligned} &\mathcal{D}_{\gamma/(\lambda+\alpha)}((\alpha/(\lambda+\alpha))J + Y^{t-1}) = U\mathcal{D}_{\gamma/(\lambda+\alpha)}(\Sigma)V^T \\ &\mathcal{D}_{\gamma/(\lambda+\alpha)}(\Sigma) = \operatorname{diag}\{\max(0, \sigma_i - (\gamma/(\lambda+\alpha)))\} \end{aligned} \tag{18}$$

and $U\Sigma V$ is the singular value decomposition of the matrix

$$(\alpha/(\lambda+\alpha))J + Y^{t-1} \text{ with } \quad \Sigma = \operatorname{diag}(\sigma_i).$$

Finally, we obtain the consensus coefficient matrix Z. The algorithm for solving our model (2) is summarized in Algorithm 1.

Algorithm 1 Adaptive Weighted Self-Representation Subspace Clustering (AWSR)

Require: Incomplete multi-views data $\{[X_i^{(0)}, X_i^{(m)}]\}_{i=1}^{v}$, initial weight matrix $W_0^{(i)}$, initial coefficient matrix Z_0, the number of clusters k, parameter γ, λ, α, *erro*;
Output: consensus coefficient matrix Z;
while
do
Update J by solving Eq.(5);
Update $\{X_i^{(m)}\}_{i=1}^{v}$ by Eq.(12);
Update Z by Eq.(17);
t=t+1;
end while
Compute the cluster assignments by $(|Z| + |Z^\top|)/2$;

4.4. Convergence Analysis

In this section, we prove that if the iteration sequence generated by the AWSR algorithm is infinite, it converges to a stationary point globally. Let $\tilde{X}^{(m)} = \{X_1^{(m)}, X_2^{(m)}, \cdots, X_V^{(m)}\}$. First, we demonstrate the property of $\{J^{(t)}, Z^{(t)}, \tilde{X}^{(m)(t)}\}$ in the following lemma.

Lemma 1. *The sequence $\{J^{(t)}, Z^{(t)}, \tilde{X}^{(m)(t)}\}$ generated via Algorithm 1 satisfies those properties:*

(1) The function $f(J^{(t)}, Z^{(t)}, \tilde{X}^{(m)(t)})$ is a monotonically decreasing function. Especially, we have

$$\begin{aligned} &f(J^{(t+1)}, Z^{(t+1)}, \tilde{X}^{(m)(t+1)}) \\ &\leqslant f(J^{(t)}, Z^{(t)}, \tilde{X}^{(m)(t)}) - (\alpha/2)\left\|J^{(t+1)} - J^{(t)}\right\|_F^2 \\ &- (\lambda/2)\left\|Z^{(t+1)} - Z^{(t)}\right\|_F^2 \end{aligned}$$

(2) $\lim_{t\to\infty} f(J^{(t)}, Z^{(t)}, \tilde{X}^{(m)(t)}) = c$ for some constant c.

(3) When $t \to +\infty$, $J^{(t+1)} - J^{(t)} \to 0$, $Z^{(t+1)} - Z^{(t)} \to 0$ and $\tilde{X}^{(m)(t+1)} - \tilde{X}^{(m)(t)} \to 0$.

(4) The sequence $\{J^{(t)}\}$, $\{Z^{(t)}\}$ and $\{\tilde{X}^{(m)(t)}\}$ are bounded.

PROOF OF THEOREM **??**. **(1).** The function $f(J, Z^{(t)}, \tilde{X}^{(m)(t)})$ of J in the tth updating process of J (3) is an α-strongly convex function. Therefore, we have

$$\begin{aligned} f(J^{(t+1)}, Z^{(t)}, \tilde{X}^{(m)(t)}) &\leqslant f(J^{(t)}, Z^{(t)}, \tilde{X}^{(m)(t)}) \\ &- (\alpha/2)\left\|J^{(t+1)} - J^{(t)}\right\|_F^2. \end{aligned} \tag{19}$$

The function $f(J^{(t+1)}, Z^{(t)}, \tilde{X}^{(m)})$ of $\tilde{X}^{(m)}$ in the tth updating process of $\tilde{X}^{(m)}$ (6) is a convex function, and the following inequality holds

$$f(J^{(t+1)}, Z^{(t)}, \tilde{X}^{(m)(t+1)}) \leq f(J^{(t+1)}, Z^{(t)}, \tilde{X}^{(m)(t)}). \tag{20}$$

Similarly, the function $f(J^{(t+1)}, Z, \tilde{X}^{(m)(t+1)})$ of Z is a λ-strongly convex function in the sub-problem (13) and we have

$$\begin{aligned} f(J^{(t+1)}, Z^{(t+1)}, \tilde{X}^{(m)(t+1)}) &\leqslant f(J^{(t+1)}, Z^{(t)}, \tilde{X}^{(m)(t+1)}) \\ &- (\lambda/2)\left\|Z^{(t+1)} - Z^{(t)}\right\|_F^2. \end{aligned} \tag{21}$$

Summation of the above inequality (19), (20) and (21) induces the inequality

$$\begin{aligned} f(J^{(t+1)}, Z^{(t+1)}, \tilde{X}^{(m)(t+1)}) &\leqslant f(J^{(t)}, Z^{(t)}, \tilde{X}^{(m)(t)}) \\ &- (\lambda/2)\left\|Z^{(t+1)} - Z^{(t)}\right\|_F^2 \\ &- (\alpha/2)\left\|J^{(t+1)} - J^{(t)}\right\|_F^2, \end{aligned} \tag{22}$$

which means the function $f(J^{(t)}, Z^{(t)}, \tilde{X}^{(m)(t)})$ decreases as t increases.

(2). Because $f(J^{(t)}, Z^{(t)}, \tilde{X}^{(m)(t)})$ is non-negative and monotonically decreasing, it converges as t tends to infinity. That is to say

$$\lim_{t\to+\infty} f(J^{(t)}, Z^{(t)}, \tilde{X}^{(m)(t)}) = c \text{ for some constant } c.$$

(3). Based on (22), we have

$$\begin{aligned} &\sum_{t=0}^{+\infty}\left[(\lambda/2)\left\|Z^{(t+1)} - Z^{(t)}\right\|_F^2 + (\alpha/2)\left\|J^{(t+1)} - J^{(t)}\right\|_F^2\right] \\ &\leqslant \sum_{t=0}^{+\infty}\left[f(J^{(t)}, Z^{(t)}, \tilde{X}^{(m)(t)}) - f(J^{(t+1)}, Z^{(t+1)}, \tilde{X}^{(m)(t+1)})\right] \\ &\leq f(J^{(0)}, Z^{(0)}, \tilde{X}^{(m)(0)}). \end{aligned} \tag{23}$$

Furthermore, we get $J^{(t+1)} - J^{(t)} \to 0$, $Z^{(t+1)} - Z^{(t)} \to 0$. In addition, the rule of updating $\tilde{X}^{(m)(t)}$ indicates

$$f(J^{(t+1)}, Z^{(t)}, \tilde{X}^{(m)(t+1)}) \leq f(J^{(t+1)}, Z^{(t)}, \tilde{X}^{(m)(t)}),$$

and then we obtain $\tilde{X}^{(m)(t+1)} - \tilde{X}^{(m)(t)} \to 0$ from (7).

(4). Due to the fact that the value $f(J^{(t)}, Z^{(t)}, \tilde{X}^{(m)(t)})$ is bounded and each term in (2) is nonnegative, the sequences $\{J^{(t)}\}$, $\{Z^{(t)}\}$ and $\{\tilde{X}^{(m)(t)}\}$ are all bounded.

Theorem 1. *Suppose $\{(J^{(t)}, Z^{(t)}, \tilde{X}^{(m)(t)})\}$ is the infinite iteration sequence produced by Algorithm 1. Then there exists $(J^{(*)}, Z^{(*)}, \tilde{X}^{(m)(*)})$ such that*

$$\lim_{t\to\infty}(J^{(t)}, Z^{(t)}, \tilde{X}^{(m)(t)}) = (J^{(*)}, Z^{(*)}, \tilde{X}^{(m)(*)}) \tag{24}$$

with $diag(Z^{(*)}) = 0$ and

$$\lim_{t\to\infty} \|\nabla f(J^{(t)}, Z^{(t)}, \tilde{X}^{(m)(t)})\| = 0. \tag{25}$$

PROOF OF THEOREM **??**. First we prove that the sequence $\{(J^{(t)}, Z^{(t)}, \tilde{X}^{(m)(t)})\}$ converges. Because $(J^{(t)}, Z^{(t)}, \tilde{X}^{(m)(t)})$ is bounded, there exists at least one accumulation point $(J^{(*)}, Z^{(*)}, \tilde{X}^{(m)(*)})$ such that a sub-sequence $\{J^{(t_j)}, Z^{(t_j)}, \tilde{X}^{(m)(t_j)}\}$ converges to $(J^{(*)}, Z^{(*)}, \tilde{X}^{(m)(*)})$, i.e.,

$$J^{(t_j)} \to J^{(*)}, Z^{(t_j)} \to Z^{(*)}, \tilde{X}^{(m)(t_j)} \to \tilde{X}^{(m)(*)}.$$

On the other hand, we have

$$J^{(t+1)} - J^{(t)} \to 0, Z^{(t+1)} - Z^{(t)} \to 0, \tilde{X}^{(m)(t+1)} - \tilde{X}^{(m)(t)} \to 0$$

from Lemma 1. Thus, the sequence $\{(J^{(t)}, Z^{(t)}, \tilde{X}^{(m)(t)})\}$ converges and Eq.(24) holds. From [3, Theorem 4.4], the solution $Z^{(t)}$ of sub-problem (13) converges to $Z^{(*)}$ satisfying $\mathrm{diag}(Z^{(*)}) = 0$.

Next we show that the partial derivative sequences with respect to J, Z and $\tilde{X}^{(m)}$ converge to zero. It can be deduced from the solution process of the sub-problems that

$$\begin{aligned} \nabla f_J(J^{(t+1)}, Z^{(t)}, \tilde{X}^{(m)(t)}) &= 0, \\ \nabla f_{\tilde{X}^{(m)}}(J^{(t+1)}, Z^{(t)}, \tilde{X}^{(m)(t+1)}) &= 0. \\ \nabla f_Z(J^{(t+1)}, Z^{(t+1)}, \tilde{X}^{(m)(t+1)}) &= 0. \end{aligned} \tag{26}$$

Since $\{(J^{(t)}, Z^{(t)}, \tilde{X}^{(m)(t)})\}$ converges, the conclusion (25) is obtained immediately.

5. Experiment

In this section, we demonstrate the performance of our AWSR method compared to performance of other common methods on five datasets.

5.1. Datasets

In the experiment, we implement our method as well as others with five benchmark datasets, named ORL, $Still$, $BBCSport$, $Olympics$ and $Leaves$. For process of creating missing view, we refer to the processing method in literature [27]. First, the $nm = [n * r]$ data entries of the first view is deleted, where n is the total number of entries and r is the missing rate. Then the nm data entries of the second view is deleted and those existing in the first view are removed to ensure all data are observed at least in one view. Finally, nm data entries arbitrarily chosen from the other views are deleted. In the numerical experiments, we consider the missing rate as 0.1, 0.2, 0.3, 0.4, 0.5. The details of each dataset are as follows.

ORL : The dataset [32] consists of 400 images taken by 40 different people at different times, lighting, facial expressions (eyes open, eyes closed, smiling and not smiling) and facial details (with/without glasses). In the experiment, ORL is a dual view dataset with dimensions 1024 and 288, respectively.

Still : The dataset [6] consists of 467 images taken of six different movements, including grabbing, running, walking, throwing, squatting and kicking. In the experiment, $Still$ is a three-view dataset with dimensions 200, 200 and 200 respectively.

BBCSport : The dataset [9] consists of 737 sports articles published on the BBC Sport website by journalists in five different areas including athletics, cricket, football, rugby and tennis. In the experiment, we only selected a subset of $BBCSport$ that contains news reports from 116 journalists. $BBCSport$ is a four-view dataset with dimensions 1991, 2063, 2113 and 2158.

Olympics : The dataset [10] contains images of 464 athletes or organizers at 28 different sports in the London 2012 Summer Olympic Games. In the experiment, $Olympics$ is a double view dataset with dimensions 4942 and 3097 respectively.

Table 2
Statistics of the datasets.

Dataset	Samples	Clusters	Demensions			
ORL	400	40	1024	288	-	-
Still	467	6	200	200	200	-
BBCSport	116	5	1991	2063	2113	2158
Olympics	464	28	4942	3097	-	-
Leaves	1600	100	64	64	64	-

Leaves : This dataset [1] contains 1600 leaf images, covering leaf images of 100 plant species. $Leaves$ describes leaves from shape, fine-scale edges and texture histogram features, respectively. In the experiment, $Leaves$ is a three-view dataset with dimensions 64, 64 and 64.

Numerical information about the data is presented in Table 2.

5.2. Methods for comparison

In this subsection, we provide a brief introduction to six other algorithms and two benchmarks, which we use to compare with our proposed algorithm.

(1) LSRs (single-best baseline) [23] first filled the missing data with random values, then it applied LSR algorithm to learning of each view.

(2) In LSRc (concatenated baseline) [23], the missing data were filled with arbitrary values, and data of each view was concatenated. It run the LSR algorithm on the concatenated data.

(3) IMG [41] factorized data of each view to learn the latent subspaces independently. Meanwhile, it utilized the graph laplacian term to regularize the latent subspaces of each view.

(4) DAIMC [12] was a weighted non-negative matrix factorization based method. It employed the existing instance alignment information to learn the consistency latent feature matrix of all views. Besides, in order to reduce the influence of missing entries, the $\ell_{2,1}$-norm was applied to build the consistency basis matrix of all views.

(5) AGL [36] first performed a low-rank representation of the graph for each view with spectral constraints. Then it built a consensus representation for all views using the co-regularization term.

(6) AWGF [39] factorized an weighted non-negative matrix to obtain the feature matrix of each view and next constructed the graph of each view. The feature extraction and the graph were fused in a large framework according to a certain weight.

(7) PLR [22] obtained the consensus feature matrix of all views based on non-negative matrix factorization. It then imposed $\ell_{2,1}$-norm constraints on the basis matrix of each view, and finally imposed regularization constraints on the local graph to obtain the shared feature matrix.

(8) IMSR [27] performed feature extraction and then employed missing data imputation and low-rank regularized consensus coefficient matrix via self-representation learning to obtain the complete consensus coefficient matrix.

5.3. Evaluation Metrics

Accuracy (ACC), normalized mutual information (NMI) and purity are used to measure the performance of clustering. Generally speaking, higher values represent better clustering performance. Next we introduce definitions of the metrics. The ACC is calculated by

$$ACC = (TP + TN)/(TP + TN + FP + FN),$$

where true positive (TP) indicates that similar documents put in the same cluster, true negative (TN) means that different documents are placed in various clusters, false negative (FN) indicates that similar documents are placed in different clusters and false positive (FP) indicates that various documents are put in the same cluster. The mutual information (MI) is

$$MI(C, \Theta) = \sum_{c_i \in C} \sum_{\omega_k \in \Theta} P(c_i, \omega_k) log(P(c_i, \omega_k))/(P(c_i)P(\omega_k)),$$

where $P(\cdot)$ is a probability function. If C and Θ are independent of each other, mutual information (MI) is equivalent to zero. In the clustering experiments, we generally use the normalized mutual information (NMI) to measure

$$NMI(C, \Theta) = (MI(C, \Theta))/(max(H(C), H(\Theta))),$$

where $H(\cdot)$ is the entropy function and is defined as $H(\Theta) = -\sum_k P(\omega_k) log P(\omega_k)$. The metric purity is given by

$$Purity(C, \Theta) = (1/n) \sum_{i=1}^{r} \max_{j=1}^{k} (\omega_k \cap c_i),$$

where n is the total number of sample, $\Theta = \{\omega_1, \omega_2, \cdots, \omega_K\}$ is the clustering class and real class is defined as $C = \{c_1, c_2, \cdots, c_J\}$.

5.4. Numerical results

In all experiments, we run each method for 10 times and then calculate the average value of the metrics coming from the 10 times' computation.

First we compare the ACC, NMI and Purity metric values of the proposed algorithm with that of the IMSR algorithm for different missing ratios and datasets. The results are illustrated in Figure 3 and Figure 4. We can see that for most of the data sets as well as different data missing ration, our AWSR method get higher ACC, NMI and Purity values than those values provided by the IMSR algorithm respectively.

Finally, we compare our method with two baselines and six algorithms on five datasets. In Table 3, we demonstrate the results of all above mentione

From Table 3, we can see that for each dataset, the overall performance of the proposed algorithm on ACC, NMI and Purity indicators is significantly better than that of other algorithms except the NMI metrics on ORL dataset, of which the proposed algorithm is the second best.

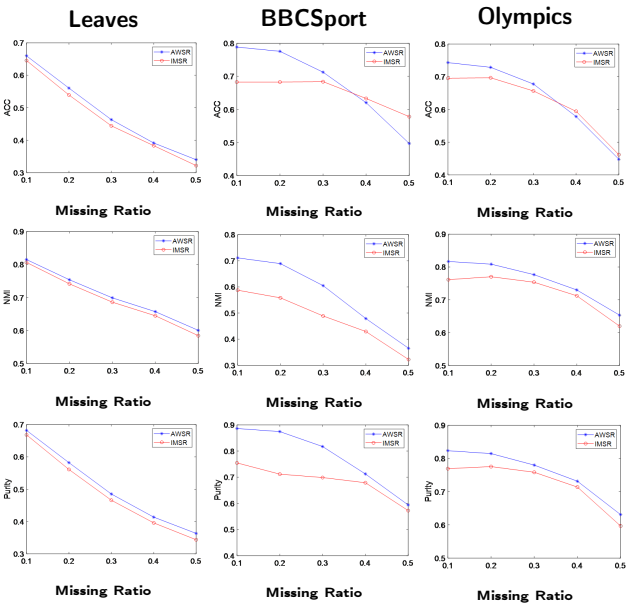

Figure 3: Performance(ACC, NMI, Purity) comparison with IMSR on *Leaves*, *BBCSport*, *Olympics*.

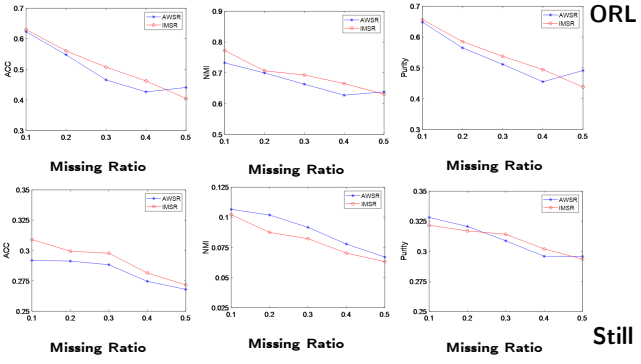

Figure 4: Performance(ACC, NMI, Purity) comparison with IMSR on *Still*, *ORL*.

6. Conclusion

For problem of incomplete multi-view data clustering, we consider different contribution of each view in self-representation learning. Adaptive weight matrices are introduced in self-representation learning. At the same time, in order to capture the data correlation structure in the same space and effectively solve the minimization problem of each data point, the F-norm constraint on consensus coefficient matrix are utilized. In order to strengthen the clustering effect, we impose the kernel norm constraint on consensus coefficient matrix. Finally, we propose our AWSR method to solve incomplete multi-view data clustering problems. The experiment performance demonstrate the effectiveness of AWSR and its superiority over other methods.

Table 3
Comparison of the performance after collation. The best results are shown in bold.

	Dataset	LSRs	LSRc	IMG	DAIMC	AGL	AWGF	PLR	IMSR	AWSR
ACC	ORL	41.50	32.20	35.30	56.10	43.20	61.27	60.50	64.63	**65.07**
	Still	29.85	27.54	29.46	32.29	28.22	29.38	31.65	33.09	**33.74**
	BBCSport	60.17	38.97	43.90	73.28	67.24	37.29	70.17	76.72	**78.15**
	Olympics	60.52	59.53	26.47	59.87	67.03	43.09	64.87	73.17	**74.44**
	Leaves	41.39	21.96	37.36	48.33	47.88	44.34	47.63	51.81	**60.93**
NMI	ORL	60.98	53.31	52.55	73.73	63.80	76.79	75.83	**78.57**	76.83
	Still	9.56	6.56	10.13	10.60	7.91	8.90	10.52	11.56	**12.01**
	BBCSport	42.88	17.16	23.04	60.49	60.10	14.51	59.51	69.06	**70.7**
	Olympics	66.45	69.67	35.12	72.44	74.28	55.92	77.09	82.16	**82.89**
	Leaves	65.06	50.75	60.75	71.59	71.32	56.40	70.61	73.49	**78.48**
Purity	ORL	43.35	34.70	40.06	59.70	45.65	65.58	63.55	67.57	**67.63**
	Still	32.38	29.98	31.15	34.90	31.13	31.76	34.30	35.20	**35.38**
	BBCSport	68.10	47.07	45.10	81.21	78.28	43.19	80.00	86.38	**87.64**
	Olympics	65.82	69.09	32.88	70.78	72.76	53.32	75.30	82.37	**83.32**
	Leaves	42.69	22.85	40.23	50.68	50.13	46.91	49.85	54.03	**63.27**

CRediT authorship contribution statement

Lishan Feng: Conceptualization, Software, Validation, Writing—original draft preparation, Visualization. **Guoxu Zhou:** Investigation,Validation, Resources, Supervision, funding acquisition. **Jingya Chang:** Methodology, Investigation, Writing—review and editing, Supervision, Funding acquisition.

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.

Acknowledgement

This work was supported by the National Natural Science Foundation of China [grant No.11901118, No.62073087].

References

[1] Beghin, T., Cope, J.S., Remagnino, P., Barman, S., 2010. Shape and texture based plant leaf classification, in: International conference on advanced concepts for intelligent vision systems, Springer. pp. 345–353.

[2] Bettoumi, S., Jlassi, C., Arous, N., 2019. Collaborative multi-view k-means clustering. Soft Computing 23, 937–945.

[3] Cai, J.F., Candès, E.J., Shen, Z., 2010. A singular value thresholding algorithm for matrix completion. SIAM Journal on optimization 20, 1956–1982.

[4] Chao, G., Wang, S., Yang, S., Li, C., Chu, D., 2022. Incomplete multi-view clustering with multiple imputation and ensemble clustering. Applied Intelligence , 1–11.

[5] Chen, J., Zhao, Z., Ye, J., Liu, H., 2007. Nonlinear adaptive distance metric learning for clustering, in: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 123–132.

[6] Delaitre, V., Laptev, I., Sivic, J., 2010. Recognizing human actions in still images: a study of bag-of-features and part-based representations, in: BMVC 2010-21st British Machine Vision Conference.

[7] Donoho, D.L., 2006. For most large underdetermined systems of linear equations, the minimal ℓ^1-norm solution is also the sparsest solution. Communications on pure and applied mathematics 59, 797–829.

[8] Gao, H., Nie, F., Li, X., Huang, H., 2015. Multi-view subspace clustering, in: Proceedings of the IEEE international conference on computer vision, pp. 4238–4246.

[9] Greene, D., Cunningham, P., 2006. Practical solutions to the problem of diagonal dominance in kernel document clustering, in: Proceedings of the 23rd international conference on Machine learning, pp. 377–384.

[10] Greene, D., Cunningham, P., 2013. Producing a unified graph representation from multiple social network views, in: Proceedings of the 5th annual ACM web science conference, pp. 118–121.

[11] Guo, J., Yin, W., Sun, Y., Hu, Y., 2019. Multi-view subspace clustering with block diagonal representation. IEEE Access 7, 84829–84838.

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

[13] Hu, M., Chen, S., 2019b. One-pass incomplete multi-view clustering, in: Proceedings of the AAAI conference on artificial intelligence, pp. 3838–3845.

[14] Huang, S., Wang, H., Ge, Y., Huangfu, L., Zhang, X., Yang, D., 2018. Improved hypergraph regularized nonnegative matrix factorization with sparse representation. Pattern Recognition Letters 102, 8–14.

[15] Kang, Z., Lin, Z., Zhu, X., Xu, W., 2021. Structured graph learning for scalable subspace clustering: From single view to multiview. IEEE Transactions on Cybernetics .

[16] Kang, Z., Zhao, X., Peng, C., Zhu, H., Zhou, J.T., Peng, X., Chen, W., Xu, Z., 2020a. Partition level multiview subspace clustering. Neural Networks 122, 279–288.

[17] Kang, Z., Zhou, W., Zhao, Z., Shao, J., Han, M., Xu, Z., 2020b. Large-scale multi-view subspace clustering in linear time, in: Proceedings of the AAAI conference on artificial intelligence, pp. 4412–4419.

[18] Kumar, A., Daumé, H., 2011. A co-training approach for multi-view spectral clustering, in: Proceedings of the 28th international conference on machine learning (ICML-11), Citeseer. pp. 393–400.

[19] Li, J., Zhou, G., Qiu, Y., Wang, Y., Zhang, Y., Xie, S., 2020. Deep graph regularized non-negative matrix factorization for multi-view clustering. Neurocomputing 390, 108–116.

[20] Li, S.Y., Jiang, Y., Zhou, Z.H., 2014. Partial multi-view clustering, in: Proceedings of the AAAI conference on artificial intelligence.

[21] Li, Y., Zhang, S., Cheng, D., He, W., Wen, G., Xie, Q., 2017. Spectral clustering based on hypergraph and self-re-presentation. Multimedia Tools and Applications 76, 17559–17576.
[22] Lian, H., Xu, H., Wang, S., Li, M., Zhu, X., Liu, X., 2021. Partial multiview clustering with locality graph regularization. International Journal of Intelligent Systems 36, 2991–3010.
[23] Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., Ma, Y., 2012. Robust recovery of subspace structures by low-rank representation. IEEE transactions on pattern analysis and machine intelligence 35, 171–184.
[24] Liu, J., Liu, X., Xiong, J., Liao, Q., Zhou, S., Wang, S., Yang, Y., 2020. Optimal neighborhood multiple kernel clustering with adaptive local kernels. IEEE Transactions on Knowledge and Data Engineering .
[25] Liu, J., Liu, X., Yang, Y., Guo, X., Kloft, M., He, L., 2021a. Multiview subspace clustering via co-training robust data representation. IEEE Transactions on Neural Networks and Learning Systems .
[26] Liu, J., Liu, X., Yang, Y., Wang, S., Zhou, S., 2021b. Hierarchical multiple kernel clustering, in: Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI, pp. 2–9.
[27] Liu, J., Liu, X., Zhang, Y., Zhang, P., Tu, W., Wang, S., Zhou, S., Liang, W., Wang, S., Yang, Y., 2021c. Self-representation subspace clustering for incomplete multi-view data, in: Proceedings of the 29th ACM International Conference on Multimedia, pp. 2726–2734.
[28] Liu, X., Dou, Y., Yin, J., Wang, L., Zhu, E., 2016. Multiple kernel k-means clustering with matrix-induced regularization, in: Proceedings of the AAAI conference on artificial intelligence.
[29] Lu, C.Y., Min, H., Zhao, Z.Q., Zhu, L., Huang, D.S., Yan, S., 2012. Robust and efficient subspace segmentation via least squares regression, in: European conference on computer vision, Springer. pp. 347–360.
[30] Merino, D.I., 1992. Topics in matrix analysis. The Johns Hopkins University.
[31] Niu, G., Yang, Y., Sun, L., 2021. One-step multi-view subspace clustering with incomplete views. Neurocomputing 438, 290–301.
[32] Samaria, F.S., Harter, A.C., 1994. Parameterisation of a stochastic model for human face identification, in: Proceedings of 1994 IEEE workshop on applications of computer vision, IEEE. pp. 138–142.
[33] Wang, Y., Lin, X., Wu, L., Zhang, W., Zhang, Q., 2014. Exploiting correlation consensus: Towards subspace clustering for multi-modal data, in: Proceedings of the 22nd ACM international conference on Multimedia, pp. 981–984.
[34] Wang, Y., Wu, L., Lin, X., Gao, J., 2018. Multiview spectral clustering via structured low-rank matrix factorization. IEEE transactions on neural networks and learning systems 29, 4833–4843.
[35] Wen, J., Fang, X., Xu, Y., Tian, C., Fei, L., 2018a. Low-rank representation with adaptive graph regularization. Neural Networks 108, 83–96.
[36] Wen, J., Xu, Y., Liu, H., 2018b. Incomplete multiview spectral clustering with adaptive graph learning. IEEE transactions on cybernetics 50, 1418–1429.
[37] Xu, H., Zhang, X., Xia, W., Gao, Q., Gao, X., 2020. Low-rank tensor constrained co-regularized multi-view spectral clustering. Neural Networks 132, 245–252.
[38] Yang, Z., Liang, N., Yan, W., Li, Z., Xie, S., 2020. Uniform distribution non-negative matrix factorization for multiview clustering. IEEE transactions on cybernetics 51, 3249–3262.
[39] Zhang, P., Wang, S., Hu, J., Cheng, Z., Guo, X., Zhu, E., Cai, Z., 2020. Adaptive weighted graph fusion incomplete multi-view subspace clustering. Sensors 20, 5755.
[40] Zhang, Z., Liu, L., Shen, F., Shen, H.T., Shao, L., 2018. Binary multi-view clustering. IEEE transactions on pattern analysis and machine intelligence 41, 1774–1782.
[41] Zhao, H., Liu, H., Fu, Y., 2016. Incomplete multi-modal visual data grouping., in: IJCAI, pp. 2392–2398.
[42] Zhou, D., Wang, J., Jiang, B., Guo, H., Li, Y., 2017. Multi-task multi-view learning based on cooperative multi-objective optimization. IEEE Access 6, 19465–19477.