Fast and General Incomplete Multi-view Adaptive Clustering

Xia Ji[1] · Lei Yang[1] · Sheng Yao[1] · Peng Zhao[1] · Xuejun Li[1]

Abstract

With the development of data collection technologies, multi-view clustering (MVC) has become an emerging research topic. The traditional MVC method cannot process incomplete views. In recent years, although many incomplete multi-view clustering methods have been proposed by many researchers, these methods still suffer from some limitations. For example, these methods all have parameters that need to be adjusted, or have high computational complexity and are not suitable for processing large-scale data. To make matters worse, these methods are not suitable for cases where there are no paired samples among multiple views. The above limitations make existing methods difficult to apply in practice. This paper proposes a Fast and General Incomplete Multi-view Adaptive Clustering (FGPMAC) method. The FGPMAC adopts an adaptive neighbor assignment strategy to independently construct the similarity matrix of each view, thereby it can handle the cases where there are no paired samples among multiple views, and eliminating the necessary to adjust the parameters. Moreover, by adopting a non-iterative approach, FGPMAC has low computational complexity and is suitable for large-scale datasets. Results of experiments on multiple real datasets fully demonstrate the advantages of FGPMAC, such as simplicity, effectiveness and superiority.

Keywords Multi-view clustering · Incomplete multi-view data · Similarity matrix spectral clustering

Introduction

With the development of information collection technology, data are often described by multiple views [1, 2]. A typical example is the classification of webpages, which can be described by two views of text and its linked information [3]. Generally speaking, different views provide complementary information to describe the data, which makes multiview learning can get better performance than single view learning [4]. As one of the most representative methods of multi-view data learning, multi-view clustering has been widely applied in many fields, such as data analysis, information retrieval and image classification [5–8]. Many advanced methods have been proposed for multi-view clustering. However, due to noise, failure of data-collecting equipment and many other unforeseen factors, data can be lost randomly in a single view or multiple views, making incomplete multi-view data widely exist [9].

Figure 1 shows two cases of incomplete multi-view data. In the first case, part of samples contain the features of all views.

In the second case, no samples contain features of all views. That is, there is no complete common samples among multiple views. For the first case, many methods have been proposed, such as PVC (Partial multi-view clustering) [9], PMH (Learning to hash on partial multi-modal data) [10] and IMG (Incomplete multi-modal visual data grouping) [11] based on matrix factorization, and APMC (Anchors Bring Ease: An Embarrassingly Simple Approach to Partial Multi-view Clustering) [12]. However, these methods cannot handle the second case, because they need to rely on common parts of multiple views. Furthermore, neither the PVC nor the IMG approach can handle multi-view data with more than two views. To our knowledge, currently only the newly proposed methods IMSC_AGL (Incomplete Multiview Spectral Clustering with Adaptive Graph Learning) [13] and AGC_IMC (Incomplete Multiview Clustering with Adaptive Graph) [14] can handle the second case of incomplete multi-view data. However, both methods are computationally complex and require many iterations to converge. In addition, in order to achieve the optimal clustering effect, they each contain three parameters that need to be adjusted. Therefore, they are difficult to use in practice.

To address the above problems, we propose a novel method called Fast and General Incomplete Multi-View Adaptive Clustering (FGPMAC). FGPMAC adopts an

✉ Lei Yang
leiyang617@163.com

1 School of Computer Science and Technology, Anhui University, Hefei 230601, China
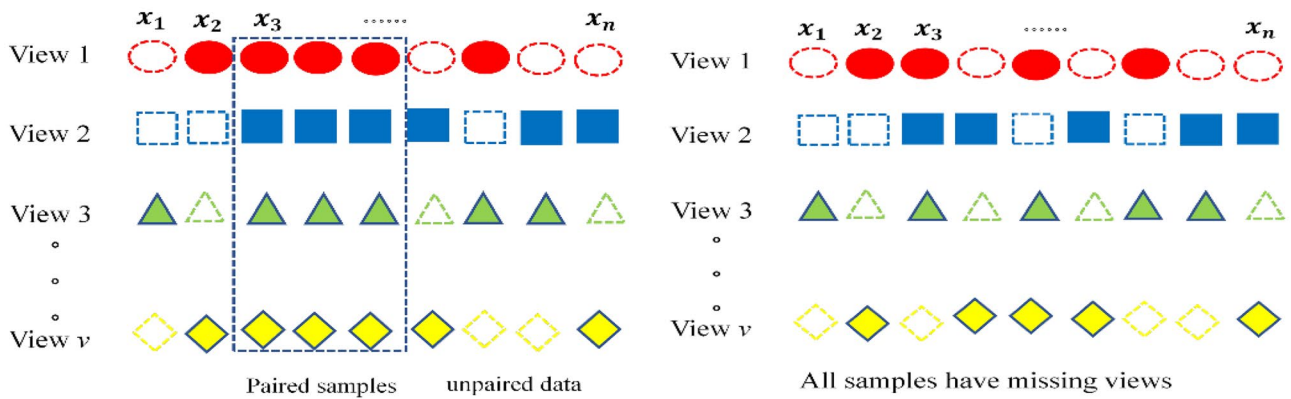
Fig. 1 Two types of the incomplete multi-view data

adaptive neighbor assignment strategy to calculate the similarity matrix of each view independently and non-iteratively, without relying on the common part of multiple views, so FGPMAC can handle the second case of the incomplete multi-view data. Experimental results on many real datasets fully demonstrate the effectiveness and efficiency of FGP-MAC. In general, FGPMAC has the following contributions.

1. FGPMAC can independently calculate the similarity matrix of each view without requiring complete common samples among multiple views. It can handle both cases of incomplete multi-view data at the same time, so FGPMAC is a general incomplete multi-view clustering method.
2. FGPMAC adopts an adaptive neighbor assignment strategy to calculate similarity, thus tedious parameter adjustment is avoided.
3. FGPMAC has a non-iterative structure with low computational complexity and is suitable for large-scale datasets.
   The rest of this paper is organized as follows. Related works are introduced in "Related Works". "The Proposed Method" describes FGPMAC in detail. "Spectral Clustering On Fused Similarities" presents the experimental resulst and the analysis. "Method Framework" is the conclusion of this paper.

Related Works

For the first case of incomplete multi-view data, many methods have been proposed. BSV (Best Single View) is the simplest and the most direct way to deal with incomplete multi-view clustering. BSV first complete the missing values with the average value of each view, and then clusters each view separately to select the best result. However, BSV does not take full use of the information of multiple views. SC[C] is a simple splicing method. SC[C] connect the features of multiple views into a long single vector, and then obtain a unified

similarity matrix and perform spectral clustering. SC[A] is a simple fusion method. SC[A] generate a similarity matrix for each view, then fuse these similarity matrices equally, and finally perform spectral clustering. MultiNMF is a clustering method that processes complete multi-view data. We first complete all missing values like BSV, and then perform MultiNMF. PVC [9] is a pioneering approach that uses nonnegative matrix factorization with l_1 sparse regularization to determine the optimal low-dimensional subspace. Meanwhile, MIC (Multiple incomplete views clustering via weighted NMF with $l_{2,1}$ regularization) [15] extends MultiNMF (Multiview clustering via joint nonnegative matrix factorization) [16] with weighted non-negative matrix factorization and $l_{2,1}$ regularization to obtain highly reliable results. IMG [11] integrates PVC and manifold learning to adaptively capture the global structure of all instances, but such integration requires additional parameters. In addition, neither PVC nor IMG can process more than two views. Trivedi et al. [17] and Gao et al. [18] proposed an incomplete multi view clustering method based on kernel regular correlation analysis respectively. However, both methods require at least one complete view as a reference. APMC [12] demonstrates a significantly improved computational efficiency, but it must ensure that there are some samples that contain all the view features.

The PVC, MIC, IMG, MultiNMF, APMC mentioned above cannot handle the second incomplete case. The newly proposed IMSC_AGL [13] and AGC_IMC [14] can handle the second case. IMSC_AGL [13] and AGC_IMC [14] can perform clustering when no samples contain all the view features. However, they have high computational complexity, respectively $O\left(\tau\left(kn^3 + n^3 + \sum_v n_v^3\right)\right)$ and $O\left(\left(\tau cn^2\right)\right)$, where τ represents the number of iterations, k represents the number of views, n represents the number of samples, and c represents the number of clusters. Due to the high computational complexity, IMSC_AGL and AGC_IMC are unsuitable for large-scale datasets. In addition, the three parameters of IMSC_AGL and those of

Table 1 Summary of the notations

n	Total number of instances
v	Total number of views
c	Number of clusters
d_v	Dimension of features in view-v
$X^{(v)}$	Data matrix of the v-th view
$Y^{(v)}$	No missing sample sets in view-v
n_v	Number of instances that appear in view-v
$Z^{(v)}$	The similarity between samples is not missing in view-v
$\overline{Z}^{(v)}$	The similarity between all samples in view-v
$G^{(v)}$	The index matrix of view-v
S	Consistency similarity matrix of multi view
$W^{(v)}$	Fusion weight of view-v

AGC_IMC are not easy to adjust in practice. The choice of parameters depends on experience and is adjusted according to the results. Inappropriate parameter selection will directly affect the accuracy of clustering. Different datasets require different optimal values for the parameters. This greatly affects the usefulness of both algorithms. Due to the limitations of previous methods, it is very necessary to do further research on incomplete multi-view clustering.

The Proposed Method

Notations

Consider dataset $X = \{X^{(1)}, X^{(2)}, ..., X^{(v)}\}$, where $X^{(v)} = \{x_1^v, x_2^v, ..., x_n^v,\} \in R^{d_v \times n}$ is the data matrix of the v-th view. v is the number of views, n is the total number of samples, and d_v is the feature dimension of the v-th view. Incomplete multi-view clustering divides all of the above samples into c clusters, where c is predefined by users. The notations used in this paper are summarized in Table 1.

Method Framework

The proposed FGPMAC involves two stages, shown in Fig. 2. At the first stage, FGPMAC adopts an adap-tive neighbor assignment strategy to construct the similarity matrix of each view indepandently, thereby eliminating the necessary to adjust the parameters. Then, FGPMAC quantifies the contribution of each view, and generates a consistent similarity matrix by fusing the similarity matrixes of multiple views. At the second stage, FGPMAC performs spectral clustering [19] on the consistent similarity matrix to obtain clustering results.

Consistent Representation Learning

Let $X^{(v)} = \left\{x_1^{(v)}, x_2^{(v)}, ..., x_n^{(v)}\right\} \in R^{d_v \times n}$ denotes the samples of the v-th view (including the missing samples). We use $Y^{(v)} = \left[y_1^{(v)}, y_2^{(v)}, ..., y_{n_v}^{(v)}\right] \in R^{d_v \times n_v} (n_v < n)$ to represent the samples that are not lost in the v-th view, where d_v and n_v are the feature dimensions and the number of unmissed samples in the v-th view, respectively. We use $Y^{(v)}$ to construct the similarity matrix $Z^{(v)}$ of the unmissed samples.

We learn similarity matrix by adaptively assigning the optimal neighbors for each sample. Nearby points have similar properties [20]. Samples with a shorter distance should have a higher probability to be neighbors. The neighbors of $x_i \in R^{d \times 1}$ can be defined as the k-nearest
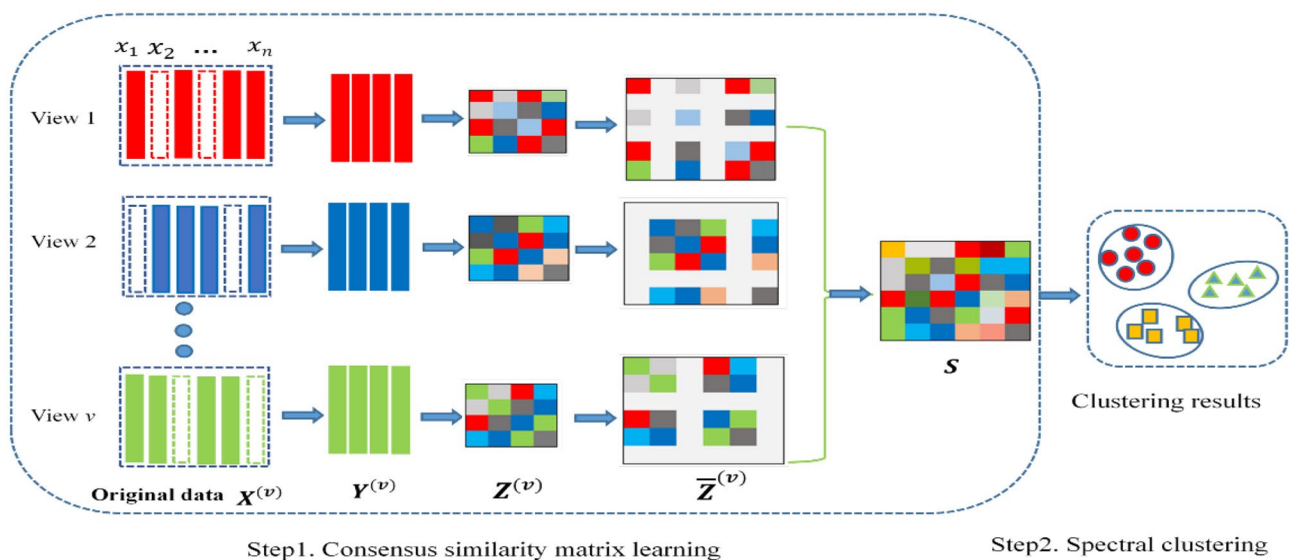

Fig. 2 The framework of the proposed incomplete multi-view clustering method

samples to x_i. We use the Euclidean distance to measure the distance between two samples. For the i-th sample x_i, all the samples $\{x_1, x_2, \ldots, x_n\}$ can be the neighbors of x_i with probability z_{ij}. Usually, a shorter distance $d_{ij} = \|x_i - x_j\|_2^2$ should be assigned a higher probability z_{ij}.

The similarity $z_{ij}^{(v)}$ stands for the probability that $y_j^{(v)}$ is the neighbor of $y_i^{(v)}$. Thereby, a good way to obtain neighbor probabilities of the i-th sample is to solve the following problem:

$$\min_{z_i^{(v)^T}1=1, z_{ij}^{(v)} \geq 0} \sum_{j=1}^{n_v} \|y_i^{(v)} - y_j^{(v)}\|_2^2 z_{ij}^{(v)} \tag{1}$$

However, there is a trivial solution in (1), i.e., only the nearest sample can be the neighbor of x_i with probability 1, and the probability of all the other samples being the neighbors of x_i is 0. This solution is obviously Pointless. To solve the problem, a regularization term is added to (1), then we have

$$\min_{z_i^{(v)^T}1=1, z_{ij}^{(v)} \geq 0} \sum_{j=1}^{n_v} \|y_i^{(v)} - y_j^{(v)}\|_2^2 z_{ij}^{(v)} + \gamma \sum_{j=1}^{n_v} z_{ij}^{(v)2} \tag{2}$$

The second term in Eq. (2) is a regularization and γ is the regularization parameter. $z_{ij}^{(v)}$ is the j-th value of $z_i^{(v)^T}$, n_v represents the number of samples appearing in view-v. Let $d_{ij}^{(v)} = \|y_i^{(v)} - y_j^{(v)}\|_2^2$, and (2) can be rewritten as

$$\min_{z_i^{(v)}} \|z_i^{(v)} + \frac{d_{ij}^{(v)}}{2\gamma}\|_2^2 \quad \text{s.t } z_i^{(v)^T}1 = 1, z_{ij}^{(v)} \geq 0 \tag{3}$$

Considering the equality and inequality constraints in (3), we use the Lagrangian function with KKT condition (Convex optimization) [21] to solve the equation. The Lagrangian function of (3) is

$$L\left(z_i^{(v)}, \eta, \beta_i\right) = \frac{1}{2}\|z_i^{(v)} + \frac{d_i^{(v)}}{2\gamma_i}\|_2^2 - \eta\left(z_i^{(v)^T} - 1\right) - \beta_i^T z_i^{(v)} \tag{4}$$

where η and β_i are the Lagrangian multipliers, η is the equality constraint coefficient, and β_i is the inequality constraint coefficient. According to the KKT condition, the following conditions (5) must be met in order to get the optimal solution.

$$\begin{cases} \frac{\partial L}{\partial z_i^{(v)}}\Big|_{z_i^{(v)} = z_i'} = 0 \\ \eta \neq 0 \\ \beta_i \geq 0 \\ \beta_i z_i' = 0 \\ z_i'^T 1 - 1 = 0 \\ z_i' \geq 0 \end{cases} \tag{5}$$

z_i' is the optimal solution and γ can be set as $\gamma = \frac{k}{2}d_{i,k+1}^{(v)} - \frac{1}{2}\sum_{j=1}^{k} d_{ij}^{(v)}$ [22] in order to obtain the optimal solution of (4):

$$z_{ij}^{(v)} = \frac{d_{i,k+1}^{(v)} - d_{ij}^{(v)}}{kd_{i,k+1}^{(v)} - \sum_{j=1}^{k} d_{ij}^{(v)}} \tag{6}$$

It is preferred to learn a sparse z_i which has exactly k nonzero values. The study [23] revealed that sparse representation is robust to noise and outliers. Because the learned Z is sparse, the computation burden of subsequent spectral clustering can be largely reduced. In addition, the number of neighbors k is much easier to adjust than the regularization parameter γ since k is an integer with an explicit meaning.

After obtaining the similarity matrix $Z^{(v)}$ among the non-missing instances in the v-th view, the similarity matrix $\overline{Z}^{(v)}$ of all samples (including the missing and non-missing samples) in the v-th view can be obtained by the following formula:

$$\overline{Z}^{(v)} = G^{(v)^T} Z^{(v)} G^{(v)} \tag{7}$$

where $G^{(v)} \in R^{n_v \times n}$ is the index matrix, in which the value associated with the missing sample is forced to 0. The definition of matrix $G^{(v)}$ is as follows:

$$G^{(v)} = \begin{cases} 1 & \text{if } y^{(v)} \text{ is the original instance } x^{(v)} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

Similar to the previous analysis, deleting those samples that suffer from missing information or filling in the missing views with the average is unreasonable. However, setting the value to 0 in the similarity of the corresponding view is reasonable. In this way, the uncertain information in the incomplete view will not play a negative role in learning the data cluster representation. By contrast, only the available information is used to guide the representation learning, which is conducive to obtaining a highly reliable data cluster representation and reducing the negative impact of missing information.

After obtaining the similarity $\overline{Z}^{(v)}$ of each view, we fuse all the similarity matrices into a unified matrix, which is also our key matrix. We use $S = \sum_1^v W^{(v)} \overline{Z}^{(v)}$ to represent the unified matrix to make full use of the information.

The non-missing samples represent the information that we can use. The more information a view provides, the greater the weight of this view is. Moreover, the view with more missing samples should be assigned a smaller weight to achieve a highly reliable and consistent representation and to reduce the negative influence of incomplete views. If the missing rates of multiple views are quite different and

the weights are equally assigned, then those views with high missing rates may provide too much inaccurate information and affect the final clustering results. Therefore, $W^{(v)}$ can be expressed as

$$W^{(v)} = n_v / \sum_1^v n_v \text{ s.t } \sum_1^v W^v = 1 \tag{9}$$

Spectral Clustering On Fused Similarities

After obtaining the fused similarity matrix S, spectral clustering is performed to gain the final clustering result. In comparison with the traditional k-means algorithm, spectral clustering is more adaptable to different data distribution. Spectral clustering learns a low-dimensional representation $F \in R^{n \times c}$ according to the fusion matrix S. Spectral clustering solves the problem (10) by eigendecomposition of L to obtain the corresponding c minimum eigenvectors, and then K-means clustering is performed to get the clustering results.

$$\min_{F^T F = I} Tr(F^T L F) \tag{10}$$

where $Tr(.)$ is the trace of the matrix, $L = D\text{-}S$ is a Laplacian matrix [24], $D \in R^{n \times n}$ is a diagonal matrix with $D_{ii} \in \sum_{j=1}^n S_{ij}$, and I is the identity matrix.

Algorithm 1 summarizes the calculation process described above.

Computational Complexity Analysis

At the first consistent similarity learning stage, the time complexity of generating the similarity matrix S is $O\left(n_v^2 \sum_v d_v\right)$, where n_v represents the number of instances that appear in view-v, and d_v represents the feature dimension of view-v. With the increase of missing samples, n_v will decrease and the algorithm complexity will be greatly reduced. The complexity of the algorithm increases approximately linearly as the number of views increases. Because every time a view is added, the similarity matrix of a view needs to be calculated accordingly.

At the second stage, spectral clustering is performed on the fused similarity matrix S, and we only need the c largest singular values. In addition, given the properties of the similarity matrix S, performing SVD (Fast svd for large-scale matrices) [25] can reduce the time complexity to $O\left(nc^2\right)$.

Algorithm 1: FGPMAC

INPUT: Incomplete multi-view data

$X = \{X^{(1)}, X^{(2)}, \dots, X^{(v)}\}$; the number of nearest neighbors k; the number of clusters c.

OUTPUT: The cluster indicators;

1: Calculate each $Z^{(v)}$ separately.

2: Construct the similarity matrix $\overline{Z}^{(v)}$ for each view.

3: Weighted fusion $\overline{Z}^{(v)}$, get $S \in R^{n \times n}$.

4: Perform spectral clustering to obtain the results.

Experiments

In this section, we conduct extensive experiments to demonstrate the effectiveness and efficiency of FGPMAC.

Datasets

Table 2 briefly describes the datasets used in the experiment. The datasets Flowers17 and USPS-MNIST with two views are used in the first incomplete case. The other datasets are used for the second incomplete case.

Table 2 Description of the used datasets

Dataset	Sample	Class	View
Flowers17	1360	17	2
USPS-MNIST	500	10	2
3Sources	169	6	3
100leaves	1600	100	3
ORL	400	40	4
NUS	1200	12	6
Caltech101	9144	102	6
NUSWIDEOBJ	30,000	31	6

The Flowers17 Dataset [26] is composed of 17 flower classes, described by color, shape, and textures. Following [12], we take the X^2 distance matrix of color and shape features as the two views.

The USPS-MNIST Dataset is the combination of two famous handwritten datasets, namely, USPS [27] and MNIST [28]. We follow [12] and randomly select 50 images from each number category from each dataset.

The 3Sources Dataset [29] contains 948 news articles collected from 3 online news sources, namely, BBC, Reuters, and The Guardian. In our experiments, we select a subset containing 169 stories reported in all 3 sources.

The 100leaves Dataset [30] contains three views with a total of 1600 samples divided into 100 categories.

The ORL Dataset [31] contains 40 categories and a total of 400 images. For each image, we generate 4 feature vectors, including GIST (512), LBP (59), HOG (864), and CENT (254).

The NUS Dataset [32] is a subset of NUS-WIDE, which contains a total of 1200 images, divided into 12 categories.

The Caltech101 Dataset [33] contains 101 objects and a background category, and each object provides 40 to 800 images.

NUS-WIDE-Object [32] is a dataset for object recognition which consists of 30,000 images in 31 classes. We use 5 features provided by the website, i.e., 65-dimension color Histogram (CH), 226-dimension color moments (CM), 145-dimension color correlation (CORR), 74-dimension edge distribution and 129 wavelet texture.

Comparison methods

1. **BSV** [11] BSV (Best Single View) first fills in the missing values with the average value of each view, and then clusters each view separately to select the best result.

2. **SC[C]** It connects the features of multiple views into a long single vector, and then obtain a unified similarity matrix and perform spectral clustering.

3. **SC[A]** It generates a similarity matrix for each view, then fuse these similarity matrices equally, and finally perform spectral clustering.

4. **MultiNMF** It first fill in all missing values like BSV, and then perform MultiNMF.

5. **PVC** [9] establishes a latent subspace where the same sample described by different views are close to each other.

6. **MIC** [15] extended MultiNMF by weighted NMF to obtain better results.

7. **IMG** [11] integrates the global structure of data into subspace learning.

8. **APMC** [12] utilizes anchors to reconstruct instance to instance relationships for clustering.

9. **IMSC_AGL** [13] exploits the graph learning and spectral clustering techniques to learn the common representation for incomplete multi-view clustering.

10. **AGC_IMC** [14] develops a joint framework for graph completion and consensus representation learning.

Experimental Settings

We construct two types of incomplete multi-view data in our experiments.

There are paired samples among multi views, i.e., some samples contain the features of all views. We set the partial data ratio (IDR) from 10 to 90% with a 20% interval, where 0% means that all views are complete. The lost samples are evenly distributed across all views, and each sample is available in at least one view.

There are no paired samples among multi views, i.e., no sample contain the features of all views. We randomly delete approximately 30%, 50%, and 70% of the samples from the multi-view dataset and then evaluate clustering performance under different missing rates.

In order to evaluate the clustering performance, two classic clustering evaluation indicators are adopted, namely,

Table 3 Running time (seconds) on different datasets with a missing rate of approximately 50%

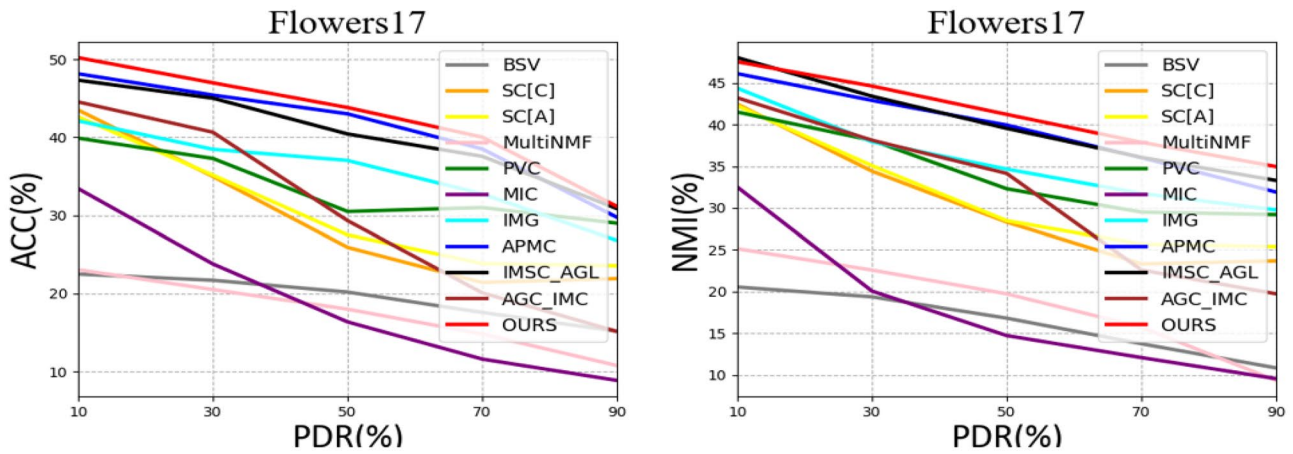| Method	3Sources	100leaves	ORL	NUS	Caltech101	NUSWIDEOBJ
BSV	1.59	10.21	1.39	23	307.14	1137.34
SC[C]	1.17	1.54	1.26	**3.01**	**134.35**	**572.43**
SC[A]	1.12	1.47	0.97	3.56	137.74	594.887
AGC_IMC	5.09	200.97	65.02	713.61	1278.09	8505.6
IMSC_AGL	10.86	410.89	79.06	1002.3	3041.25	14,487.13
OURS	**0.18**	**1.31**	**0.35**	3.78	173.07	911.7

Fig. 3 Experiment results on the Flowers17 dataset in the first incomplete case.

clustering accuracy (ACC) and standardized mutual information (NMI). The values of these indicators range from 0 to 1, and a larger value indicates a better performance of the clustering algorithm. For fairness, we perform all comparison methods 10 times on each dataset and report the average clustering results to eliminate the uncertainty caused by randomness. The parameters involved in the comparison method are set according to author's suggestion or default value. For the purpose of reproducibility, the code and datasets are released at: https://github.com/leiyang617/code_for_FGIMAC.

Results and Analysis

Figures 3 and 4 report the ACC and MNI at different missing rates in the first incomplete case, whereas Figs. 5–7 report the indicators in the second incomplete case. Table 3 shows the run time of different methods on various datasets with a missing rate of approximately 50%. Due to the limit of

space, we only report the results obtained when the missing rate is 50% given that similar trends can be observed in the other missing rates.

The clustering performance of all methods declines along with an increasing missing rate. Figures 3 and 4 show that BSV, SC[C], SC[A], MultiNMF and MIC are unsatisfactory in most cases, indicating that filling the missing samples with the average is not good enough to solve the incomplete multi-view clustering problem. BSV only uses a single view to obtain the clustering results and is unable to use the complementary information among multiple views. SC[C] connects all views into a single long view and ignores the differences in the distributions of various views. SC[A] treats all views equally without considering their integrity or credibility, which is unreasonable for incomplete multi-view clustering. PVC, IMG, APMC, IMSC_AGL and AGC_IMC demonstrate an acceptable performance, suggesting that using the complementary information among views is an effective approach.
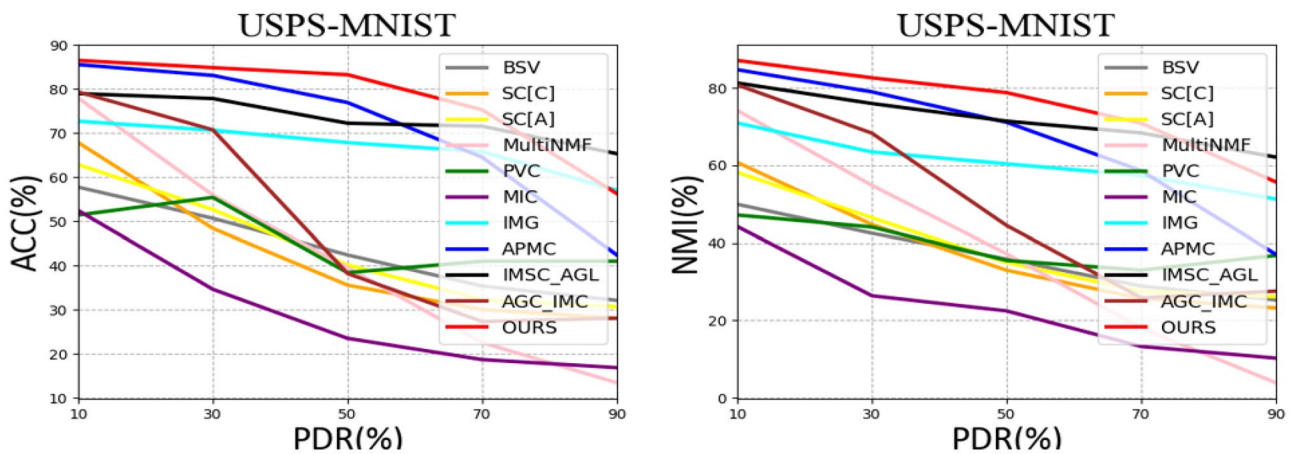

Fig. 4 Experiment results on the USPS-MNIST dataset in the first incomplete case

Fig. 5 Experiment results on different datasets with 30% missing rate in the second incomplete case
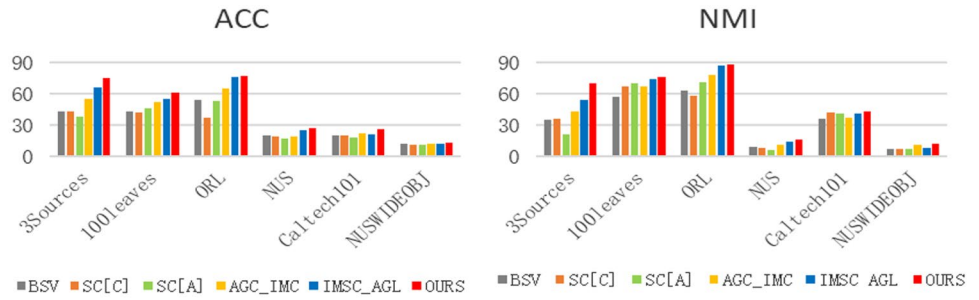

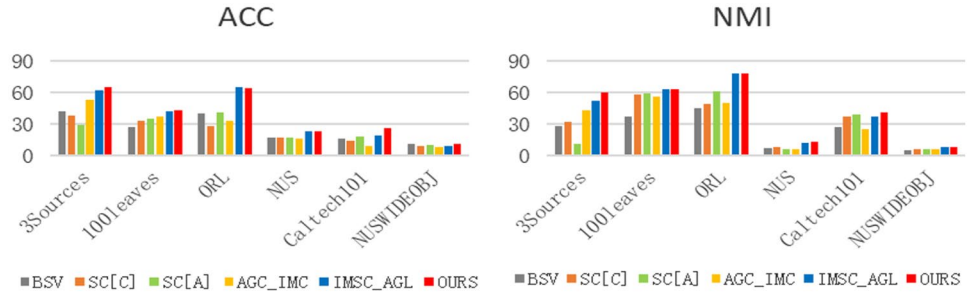**Fig. 6** Experiment results on different datasets with 50% missing rate in the second incomplete case

Fig. 7 Experiment results on different datasets with 70% missing rate in the second incomplete case
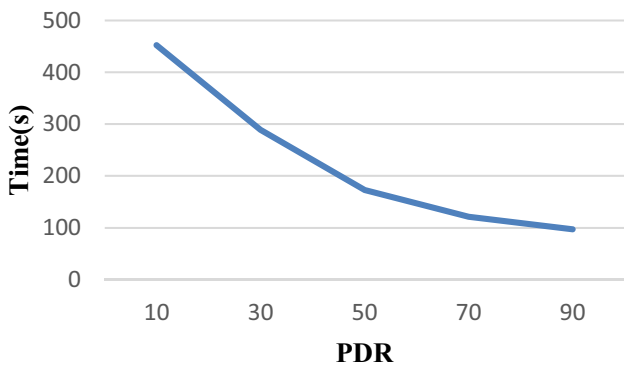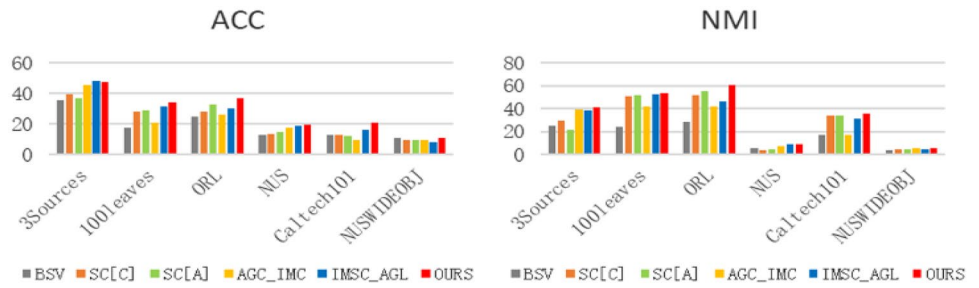

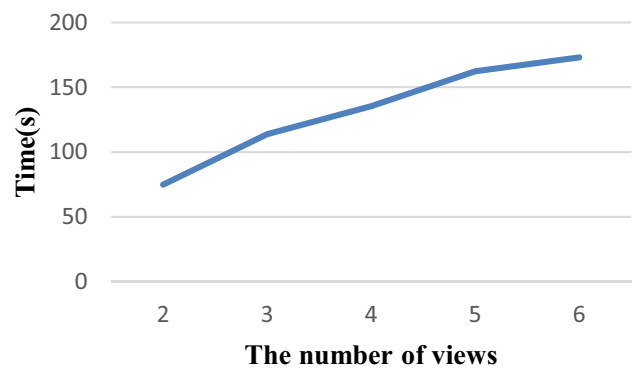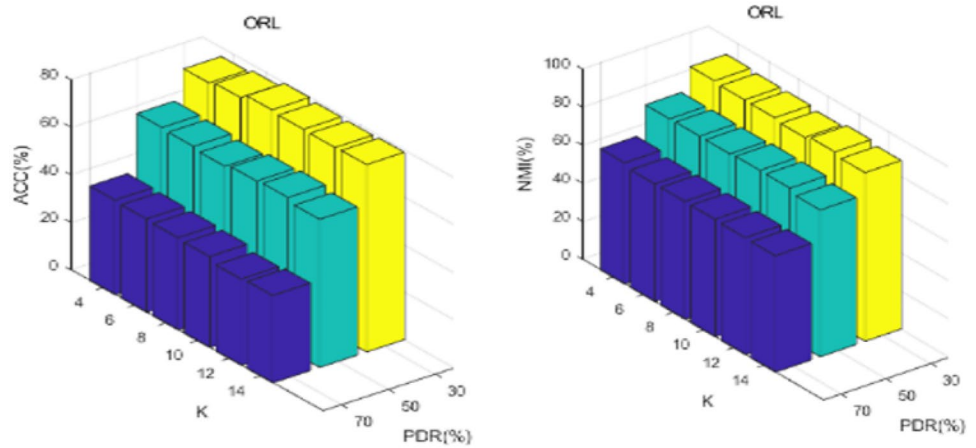**Fig. 8** Running time on Caltech101 dataset with different missing rates

Fig. 9 Running time on Caltech101 dataset with different numbers of views

Fig. 10 Influence of the number of nearest k on the ORL dataset with different PDR settings

In the first incomplete case, Fig. 3 show that FGPMAC obtains better results than other methods on the Flowers17 dataset. Figure 4 show that on the USPS-MNIST dataset, FGIMAC obtains better results than other methods except when the PDR is 90%.

In the second incomplete case, Fig. 5 shows that FGIMAC outperforms the other methods in all datasets when the missing rate is 30%. Figure 6 shows that when the missing rate is 50%, FGIMAC achieves the best results on all datasets except the ORL dataset and the NUS dataset. Figure 7 shows that when the missing rate is 70%, FGPMAC obtains the best results on all datasets except the 3Sources dataset. When FGPMAC fails to obtain the optimal result, the gap between its results and the optimal ones is very small.

Although IMSC_AGL and AGC_IMC have suboptimal clustering performance, but their computational complexity is unsatisfied.

It can be seen from Table 3 that the run time of IMSC_AGL and AGC_IMC is much higher than that of other methods. Our method has absolute advantages in terms of run time. For instance, on the 3Sources dataset, it takes IMSC_AGL 10.86 s to run, 5.09 s for AGC_IMC, while our method only use 0.18 s. On the Caltech101 dataset, the run time of IMSC_AGL is 3041.25 s, and that of AGC_IMC is 1278.09 s, while ours is just 173.07 s, especially on the dataset 100leaves, the run time of method AGC_IMC and IMSC_AGL is about hundreds of times longer than ours. In addition, the experimental results on Cal-tech101 and NUS-WIDEOBJ datasets show that the run time of IMSC_AGL and AGC_IMC increases rapidly with the enlargement of the dataset scale. For large-scale datasets, this is clearly unacceptable.

In order to further explore how the time complexity of the algorithm proposed in this paper changes with the number of views and sample missing rate, we conducted experiments on the Caltech101 dataset. The results are shown in Figs. 8 and 9, and it can be seen from the results that as the sample missing rate increases, the clustering accuracy decreases and the running time reduces significantly. In addition, ACC and NMI decrease dramatically when the missing rate reaches 70%, because the missing rate is so large that we have little information to use. Data with a missing rate of more than 70% are rare in practical applications. As the number of views increases, the accuracy increases and the runtime increases accordingly. The change in running time is consistent with what we analyzed in "Consistent Representation Learning". Furthermore, it can be seen that as the number of views increases, the ACC and NMI also increase, which indicates that combining features from multiple views helps to improve the clustering performance.

We test the sensitivity of parameters to further evaluate the performance of FGIMAC, which only needs to adjust pa-rameter k. We evaluate the clustering performance of k in the range of $\{4, 6, \ldots, 14\}$. Due to the limited space, we only report the results on the ORL dataset given that similar trends can be observed in the other datasets. As shown in Fig. 10, under the same PDR, the fluctuation of ACC and NMI is very small when different k are selected, so it has little influence on the final clustering result. In other words, FGIMAC has low sensitivity to k in a relatively wide range.

Conclusion

We propose FGIMAC, a simple and effective method for incomplete multi-view clustering that can overcome many shortcomings of the existing methods. Compared to traditional methods, FGIMAC demonstrates greater flexibility as it is able to cluster in two imcomplete cases of multi-view data. We perform experiments on multi-view datasets with

different missing rates, and experimental results show that FGIMAC achieves higher clustering performance with less run time.

However, most of the existing incomplete multi-view clustering methods (including our method) must take the number of clusters into consideration. As for future work, we plan to refer to COMIC (COMIC: Multi-view Clustering Without Parameter Selection) [34] and adjust our method so that clustering can be performed without knowing the number of clusters.

Data Availability The data and code that support the findings of this study are openly available at https://github.com/leiyang617/code_for_FGIMAC.

Declarations

Ethical approval This article does not contain any studies with human or animal subjects performed by any of the authors.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Song B, Zhou Z, Wang J, Xiang B, Qi T. Ensemble Diffusion for Retrieval. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2017.
2. Zhang Z, Xie Y, Zhang W, Tian Q. Effective Image Retrieval via Multilinear Multi-Index Fusion. IEEE Trans Multimedia. 2019;21(11):2878–90.
3. Xu W, Liu X, Gong Y. Document clustering based on non-negative matrix factorization. In: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval. 2003;267–273.
4. Yin M, Gao J, Xie S, Guo Y. Multiview subspace clustering via tensorial t-product representation. IEEE Transactions on Neural Networks and Learning Systems. 2019;30(3):851–64.
5. Salvador J, Casas JR. Multi-view video representation based on fast monte carlo surface reconstruction. IEEE Trans Image Process. 2013;22(9):3342–52.
6. Tan TY, Zhang L, Lim CP. Adaptive melanoma diagnosis using evolving clustering, ensemble and deep neural networks. Knowledge-Based Systems. 2020;187:104807.1–104807.26.
7. Sato Y, Izui K, Yamada T, Nishiwaki S. Data mining based on clustering and association rule analysis for knowledge discovery in multiobjective topology optimization. Expert Syst Appl. 2019;119:247–61.

8. Zhu X, Zhang S, He W, Hu R, Lei C, Zhu P. One-Step Multi-View Spectral Clustering. IEEE Trans Knowl Data Eng. 2019;31(10):2022–34.
9. Li SY, Jiang Y, Zhou ZH. Partial multi-view clustering. In: Proceedings of the AAAI conference on artificial intelligence. 2014;1968–1974.
10. Wang Q, Si L, Shen B. Learning to hash on partial multi-modal data. In: Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI). 2015;3904–3910.
11. Zhao H, Liu H, Fu Y. Incomplete multi-modal visual data grouping. In: Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI). 2016;2392–2398.
12. Guo J, Ye J. Anchors Bring Ease: An Embarrassingly Simple Approach to Partial Multi-view Clustering. In: Proceedings of the AAAI conference on artificial intelligence. 2019;118–125.
13. Wen J, Xu Y, Liu H. Incomplete Multiview Spectral Clustering with Adaptive Graph Learning. IEEE Transactions on Cybernetics. 2020;50(4):1418–29.
14. Wen J, Yan K, Zhang Z, Xu Y, Zhang B. Adaptive graph completion based incomplete multi-view clustering. IEEE Trans Multimedia. 2020;99:1–1.
15. Shao W, He L, Yu P. Multiple incomplete views clustering via weighted NMF with regularization. ECML/PKDD. 2015;318–334.
16. Liu J, Wang C, Gao J, Han J. Multi-view clustering via joint nonnegative matrix factorization. SDM. 2013;252–260.
17. Trivedi A, Rai P, Daume H, DuVall SL. Multiview clustering with incomplete views. In: Advances in Neural Information Processing Systems Workshop. 2010.
18. Gao H, Peng Y, Jian S. Incomplete multi-view clustering. In: International Conference on Intelligent Information Processing. Springer. 2016;245–255.
19. Ng AY, Jordan MI, Weiss Y. On Spectral Clustering: Analysis and an Algorithm. In: Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic. MIT Press. 2001.
20. Hou C, Nie F, Li X, Yi D, Yi W. Joint Embedding Learning and Sparse Regression: A Framework for Unsupervised Feature Selection. IEEE Transactions on Cybernetics. 2013;44(6).
21. Boyd S, Vandenberghe L. Convex optimization. In Cambridge University Press. 2004.
22. Nie F, Wang X, Huang H. Clustering and projected clustering with adaptive neighbors. In: Proceedings of ACM SIGKDD. 2014.
23. Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y. Robust face recognition via sparse representation. In IEEE Transactions on Pattern Analysis and Machine Intelligence. 2019;31:210–27.
24. Hagen L, Kahng AB. New spectral methods for ratio cut partitioning and clustering. IEEE Trans Comput Aided Des Integr Circuits Syst. 1992;11(9):1074–85.
25. Holmes M, Gray A, Isbell C. Fast svd for large-scale matrices. In NIPS workshop. 2007;28:249–52.
26. Nilsback ME, Zisserman A. A visual vocabulary for flower classification. In CVPR. 2006;1447–1454.
27. Hull J. A database for handwritten text recognition research. TPAMI. 1994;16(5):550–4.
28. LeCun Y, Bottou L, Bengio Y, Haaffner P. Gradient-based learning applied to document recognition. In: Proceedings of the IEEE. 1998;86(11):2278–2324.
29. Greene D, Cunningham P. A matrix factorization approach for integrating multiple data views. ECML/PKDD. 2009;423–438.
30. Mallah C, Cope J, Orwell J. Plant leaf classification using probabilistic integration of shape, texture and margin features. In: Proceedings of the IASTED International Conference Signal Processing, Pattern Recognition and Applications. 2013;279–286.

31. Samaria FS, Harter AC. Parameterisation of a stochastic model for human face identification. In: Proceedings of the Second IEEE Workshop on Applications of Computer Vision. 1994;138–142.

32. Chua TS, Tang J, Hong R, Li H, Luo Z, Zheng Y. Nus-wide: a real-world web image database from national university of Singapore. In: Proceedings of the ACM International Conference on Image and Video Retrieval, (CIVR). 2009;48.

33. Dueck D, Frey BJ. Non-metric affinity propagation for unsupervised image categorization. In: Proceedings of the IEEE International Conference on Computer Vision, IEEE. 2007;1–8.

34. Peng Xi, et al. COMIC: Multi-view clustering without parameter selection. In: International conference on machine learning. PMLR. 2019;5092-5101.