

Received September 3, 2020, accepted September 11, 2020, date of publication September 22, 2020, date of current version October 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3025881

Adaptive Anchor-Based Partial Multiview Clustering

XIA JI¹, LEI YANG¹, AND SHENG YAO¹

School of Computer Science and Technology, Anhui University, Hefei 230601, China

Corresponding author: Xia Ji (jixia1983@163.com)

This work was supported in part by the Natural Science Foundation of China under Grant 61402004 and Grant 61672034, in part by the Key Research and Development Program of Anhui Province under Grant 1804d8020309, and in part by the Natural Science Foundation of Anhui Province under Grant 1908085MF188.

ABSTRACT Partial multiview clustering, which aims to effectively merge multiple prespecified incomplete views to improve clustering performance, is a research hotspot and difficulty in the field of machine learning. Guo *et al.* proposed a partial multiview clustering method (APMC) based on anchor graph, which uses a Gauss kernel function to solve the similarity matrix. The Gaussian kernel function is sensitive to parameter σ , and it is difficult to find the optimal value only by stepwise adjustment in practical applications. This undoubtedly affects the practicality of the APMC algorithm. To address this issue, an adaptive partial multiview clustering method based on anchor graph (AAPMC) is developed in this paper, which proposes an adaptive neighbor assignment strategy and utilizes it to improve the anchor-based similarity matrix computation of each view. The method proposed in this paper uses an adaptive method to solve the similarity matrix, eliminating the tediousness of parameter adjustment. In addition, the non-iterative method based on anchors is used to solve the optimal solution with low time complexity and is suitable for large-scale datasets. In short, our method is simple and effective, and it is easier to implement in practice. Extensive experiments show that our model can not only effectively solve the parameter setting problem, but also performs better than the state-of-the-art partial multiview clustering methods.

INDEX TERMS Adaptive neighbor, anchor graph, partial multiview clustering.

I. INTRODUCTION

With the development of information technology, multiview data is becoming more and more common in reality [1], [2]. For example, a patient's physical condition can be monitored simultaneously by multiple sensors in a home care system, or an image can be represented by visual features or text annotations. In general, different views provide complementary information to describe the data, which makes multiview learning get better performance than single view method [3]. As one of the most representative methods of multiview learning, multiview clustering can obtain better clustering results by exploring the consistency and complementarity of different views, and has been widely applied in data analysis, image classification, significance detection and information retrieval [4]–[6]. Many advanced methods have been proposed for multi-view clustering. Zhou *et al.* [7] propose a novel incremental multi-view spectral clustering

method (IMSC). In IMSC, instead of ensembling the collection of all views simultaneously, they integrate them one by one in an incremental way. This method is scalable and applicable to streaming views. Wang *et al.* [8] propose a Multi-view Clustering via Late Fusion Alignment Maximization (MVC-LFA). MVC-LFA proposes to maximally align the consensus partition with the weighted base partitions. Such a criterion is beneficial to significantly reduce the computational complexity and simplify the optimization procedure. 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 partial view data widely exist [9]. Traditional multiview clustering methods cannot directly process this data because they work under the assumption that all views are complete. In recent years, partial multi-view clustering has attracted extensive attention.

Researchers have proposed a variety of methods for partial multiview clustering [10]–[20], which can be generally divided into two categories: matrix factorization-based

The associate editor coordinating the review of this manuscript and approving it for publication was Qichun Zhang¹.

clustering method and graph-based clustering method. The matrix factorization technique directly learns the low dimensional consistent representation of all views. For example, PVC [10] uses nonnegative matrix factorization (NMF) and L_1 regularization to learn a common latent subspace of all views where instances of different views are forced to have the same representation. MIC [11] first fills the missing samples with the average of all instances in the corresponding view, and then extends MultiNMF [12] through $L_{2,1}$ regularized NMF to learn the consistent representation of all views. DAIMC [13] further extends MIC by seminegative matrix factorization and $L_{2,1}$ regularization regression. However, they only pay attention to the consistent representation of learning and ignore the internal structure of the data, and cannot guarantee the distinguishability and compactness of the representation. Graph based methods aim to learn low dimensional representations from different graphs, which reveal the relationships among all samples. Compared with the method based on matrix factorization, the geometric structure of data can be effectively used. Trivedi *et al.* [14] proposed to complete the view lacking instances by referring to the Laplacian matrix of the complete view, and then learn the low-dimensional representation of different views through the kernel CCA. It requires at least a complete view, which is the biggest disadvantage of this method. IMG [15] integrates latent subspace generation and compact global structure into a unified framework through a Laplacian graph on a complete data instance, and this integration brings more parameters. DCNMF [16] develops a dual constraint framework by combining cluster similarity and manifold keeping constraints. Gao *et al.* [17] proposed to fill the missing views with the average of the instances in the corresponding views for the construction of the graph and the learning of the subspace. However, this method will not work when the multiview data has a large number of missing instances in all views. Liu *et al.* [18] unifies the estimation and clustering of partial multiview into a single optimization process, proposes MKKM-IK algorithm, which has good performance, but has high computational and storage complexity at the same time. On the basis of MKKM-IK, LF-IMVC [19] proposed a late fusion method to simultaneously clustering and imputing the incomplete base clustering matrices. Though both MKKM-IK and LF-IMVC unify the imputation and clustering into a single optimization, they are different in the manner of imputation: the former is early fusion (or kernel-level imputation), while the latter is a kind of late fusion (or decision-level imputation).

Inspired by anchor-based strategies [21], [22], Guo and Ye [20] proposed an anchor-based partial multiview clustering (APMC) method. APMC uses anchors to reconstruct clustering relationships between instances, and integrates intra-view and inter-view similarities through anchors. This anchor-based approach effectively improves the calculation efficiency, and easy to be extended for more than two partial views.

However, there are still some limitations in APMC. Firstly, there are two parameters involved in the intra-view similarity computation, one is k , which controls how many nearest anchors are selected as the basis for sample representation, the other is σ , which adjusts the local influence range of the Gaussian kernel function. Especially the Gaussian kernel function used for solving the similarity matrix is very sensitive to the parameter σ , but APMC does not give effective setting criteria or method. Parameters problem leads to poor operability of APMC in practical applications [23],[24]. Secondly, ignoring of the credibility of view itself. At the stage of synthesizing inter-view similarity, APMC treats all views equally. Affected by factors such as missing samples or noise, the credibility of different views should be different.

In this paper, to solve the limitations of APMC, we introduce an adaptive strategy to solve the similarity matrix, which is effective and easy to operate for partial multiview clustering. On this basis, an *Adaptive Anchor-based Partial multiview Clustering* (AAPMC) method is presented. The experimental results verify its superiority.

The rest of this paper is organized as follows. Section 2 briefly introduces some works related to the proposed method. Section 3 introduces AAPMC in detail. Section 4 presents the experimental results. Section 5 concludes the paper.

II. RELATED WORK

In this section, we review relevant research of partial multiview clustering and anchor-based similarity reconstruction as our foundations for the subsequent discussion.

A. NOTATIONS AND PROBLEM DEFINITION

We summarize the notations used in this paper, which is shown in the Table 1.

TABLE 1. Summary of the notations.

Symbol	Definition and description
n	Total number of instances.
c	Number of clusters.
d_v	Dimension of features in view- v .
n_c	Number of common instances.
n_v	Number of instances that only appear in view- v .
$X^{(1,2)}$	The common instances present in both views.
$X^{(v)}$	The instances that appear only in view- v .
l	Number of anchors.
$Z^{(v)}$	Similarity matrix between instance and anchor of view- v .
Z	Integrated similarity matrix between instance and anchor.
S	Consistency similarity matrix of multiview.
$W^{(v)}$	Fusion weight of view- v .

To solve the missing examples clustering problem, partial multiview clustering method always assume that instances share the common feature space in each individual view and the two different views are bridged by the shared common examples. To facilitate the discussion and without loss of generality, we take two views for illustration. Suppose there are n samples $\{x_1, x_2, \dots, x_n\}$ in total. We separate the original data of the two partial views as $\{X^{(1,2)}, X^{(1)}, X^{(2)}\}$. $X^{(1,2)} \in R^{n_c \times (d_1 + d_2)}$ represents the common samples present

in both views, n_c is the number of common samples. $X^{(1)} \in \mathbb{R}^{n_1 \times d_1}$ represents the samples that appear only in view-1, n_1 is the number of samples that only appear in view-1, $X^{(2)} \in \mathbb{R}^{n_2 \times d_2}$ denotes the samples only appearing in view-2, n_2 is the number of samples only appearing in view-2. Here d_1 and d_2 denote the feature dimension of view-1 and view-2 respectively. Obviously, the formula $n = n_c + n_1 + n_2$ can be derived.

As shown in Fig 1, the paired common part of the two views are $X^{(1,2)} = \{x_3, x_4, x_5\}$. The samples appear only in view-1 are $X^{(1)} = \{x_1, x_2\}$ and the samples appear only in view-2 are $X^{(2)} = \{x_6, x_7\}$.

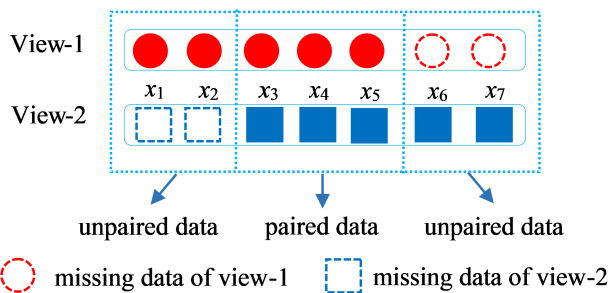


FIGURE 1. Partial multiview data.

The purpose of partial multiview clustering is to group all the above samples into c clusters, where c is given in advance by the user.

B. THE SIMILARITY MATRIX CONSTRUCTION IN APMC

The APMC algorithm uses spectral clustering method to classify the data at last, and as known, spectral clustering classifies the data according to the similarity matrix of the input data, so the clustering results of APMC depend on the similarity learning of the data to a great extent.

APMC algorithm firstly select the k nearest anchor points as a set of bases to represent every sample, and then uses a similarity calculation method adopt Gaussian kernel function to calculate the intra-view similarity. At the stage of solving the inter-view similarity, the APMC synthesize the similarities between common instances and anchors in the proportion of two views each accounting for half. The steps of anchor-based similarity construction in APMC algorithm are briefly described as follows.

(1) Intra-view similarity construction. Suppose there are l pairs of anchor points by selecting the common instances that appear in both views, i.e., $l = n_c$. Denote the set of all instances in the v -th view as $(x_i^{(v)})_{i=1}^{n_c+n_v}$, the anchor points in the v -th view as $(u_i^{(v)})_{i=1}^{n_c}$. The similarity between the instance $x_i^{(v)}$ and anchor point $u_j^{(v)}$ is defined as

$$Z_{ij}^{(v)} = \begin{cases} \frac{\exp(-D^2(x_i^{(v)}, u_j^{(v)})/\sigma^2)}{\sum_{j \in \langle i \rangle^v} \exp(-D^2(x_i^{(v)}, u_j^{(v)})/\sigma^2)} & \forall j \in \langle i \rangle^v \\ 0 & \text{otherwise} \end{cases}$$

where $\langle i \rangle^v$ is an index set of k ($k \ll l$) nearest anchors of $x_i^{(v)}$ according to a distance function $D^2(x, u)$ and σ is the parameter controlling the neighborhood width.

(2) Inter-view similarity solving. The final unified similarity matrix is composed of three parts $Z = [\tilde{Z}; \tilde{Z}^{(1)}; \tilde{Z}^{(2)}] \in \mathbb{R}^{n \times l}$, here $\tilde{Z}^{(v)} \in \mathbb{R}^{n_v \times l}$ ($v = 1, 2$) indicates the similarity between the samples appear only in the v -th view and anchors, which is consist of the last n_v rows of the $Z^v \in \mathbb{R}^{(n_c+n_v) \times l}$. $\tilde{Z} \in \mathbb{R}^{n_c \times l}$ indicates the similarity between the common instances and anchors. In APMC, to leverage the information of both views, the element of \tilde{Z} is defined as $\tilde{z}_{ij} = \frac{1}{2}(\tilde{z}_{ij}^{(1)} + \tilde{z}_{ij}^{(2)})$.

The consistent similarity matrix $S \in \mathbb{R}^{n \times n}$ among all instances is approximated as $S = Z\Lambda^{-1}Z^T$ in APMC, and here $\Lambda = \text{diag}(Z^T 1) \in \mathbb{R}^{l \times l}$.

C. MOTIVATIONS AND CONTRIBUTIONS

Clustering is a solution process from the local structure of the sample to the global structure of the sample set. The accurate expression of the local structure of the sample is the prerequisite to ensure the final clustering quality.

In partial multiview clustering, it is more challenging to accurately express and fuse the local structure of samples with incomplete information through samples with complete information. APMC provides a solution, but it involves multiple parameters, and there is a problem of parameter sensitivity. In complete single view clustering problem, Nie et al. [25] proposed an adaptive neighbor assignment strategy to learn the data similarity matrix by assigning the adaptive and optimal neighbors for each data point based on the local connectivity. We improve this strategy and apply it to the calculation of sample similarity in partial multiview clustering.

To be specific, first we should select the common parts of the views as anchors to build an anchor graph and use an adaptive neighbor assignment strategy to construct a similarity matrix between samples and anchors, then we integrate intra-view and inter-view similarities. Finally, perform spectral clustering on the consensus matrix to obtain clustering results.

In general, our approach has the following advantages:

(1) Our method uses the parameter-free adaptive neighborhood assignment strategy to construct the similarity matrix, which avoids the tedious parameter adjustment and is easier to use. Finally, perform spectral clustering on the consensus matrix to obtain clustering results.

(2) Our method adopts a non-iterative approach with low time complexity. In addition, this method solves the similarity matrix only involving basic operations of addition, subtraction, multiplication, and division. It does not require the calculation of Gaussian kernel functions and other more operations, so it is more efficient.

(3) This method has good generalization ability and can be extended to more than two views.

III. THE PROPOSED METHOD

A. METHOD FRAMEWORK

The partial multiview clustering method proposed in this paper can be divided into two stages. Fig 2 shows the

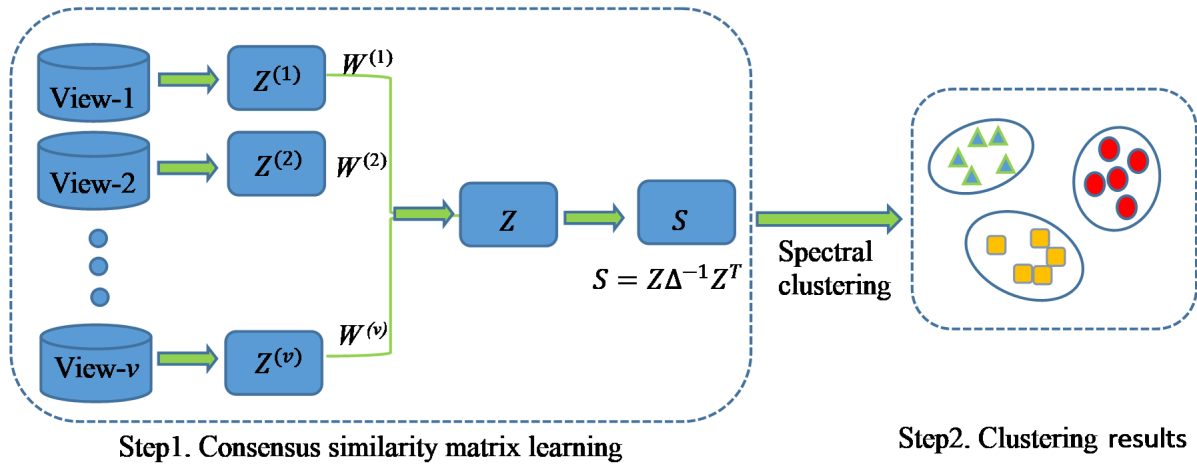


FIGURE 2. The framework of the proposed method.

process of the two stages. The consensus similarity matrix is constructed at the first stage and spectral clustering is performed at the following stage. The method we proposed in this paper aims to learn a consensus representation S from multiple views for clustering, and the solution process of the consensus similarity matrix is shown in Fig 3. We still use two views to illustrate.

B. ANCHOR-BASED ADAPTIVE SIMILARITY CONSTRUCTION

The similarity matrix is learned by assigning a local connection-based adaptive optimal neighbor to each sample. The similarity $z_{ij}^{(1)}$ between the samples $x_i^{(1)}$ and the anchor $u_j^{(1)}$ is expressed as the probability that $u_j^{(1)}$ is the neighbor of $x_i^{(1)}$. A closer distance corresponds to a greater probability and similarity. Therefore, the neighbor probability of obtaining the i -th sample in view-1 needs to solve the following problems:

$$z_i^{(1)T} \min_{\mathbf{1}=1, z_{ij}^{(1)} \geq 0} \sum_{j=1}^l \|x_i^{(1)} - u_j^{(1)}\|_2^2 z_{ij}^{(1)} + \gamma \sum_{j=1}^l z_{ij}^{(1)2}, \tag{1}$$

where $z_i^{(1)T}$ represents the i -th row of $Z^{(1)}$, $z_{ij}^{(1)}$ is the j -th value of $z_i^{(1)T}$, $\mathbf{1}$ denotes a column vector with all elements equal to 1, l is the number of anchors and γ is the regularization parameter. Let $d_{ij}^{(1)} = \|x_i^{(1)} - u_j^{(1)}\|_2^2$, and (1) can be rewritten in vector form as

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

Given the equality and inequality constraints in (2), we use the Lagrangian function with KKT condition [26] to solve. The Lagrangian function of (2) is

$$L(z_i^{(1)}, \eta, \beta_i) = \frac{1}{2} \|z_i^{(1)} + \frac{d_i^{(1)}}{2\gamma_i}\|_2^2 - \eta(z_i^{(1)T} - 1) - \beta_i^T z_i^{(1)}, \tag{3}$$

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, to solve the above optimization problem, the following conditions must be met:

$$\begin{cases} \frac{\partial L}{\partial z_i^{(1)}} \Big|_{z_i^{(1)}=z_i'} = 0 \\ \eta \neq 0 \\ \beta_i \geq 0 \\ \beta_i z_i' = 0 \\ z_i'^T \mathbf{1} - 1 = 0 \\ z_i' \geq 0, \end{cases} \tag{4}$$

where z_i' is the optimal solution, and γ can be set as $\gamma = \frac{k}{2} d_{i,k+1}^{(1)} - \frac{1}{2} \sum_{j=1}^k d_{ij}^{(1)}$ [25], such that the optimal solution to (3) is

$$z_{ij}^{(1)} = \begin{cases} \frac{d_{i,k+1}^{(1)} - d_{ij}^{(1)}}{kd_{i,k+1}^{(1)} - \sum_{j=1}^k d_{ij}^{(1)}} & j \leq k \\ 0 & j > k. \end{cases} \tag{5}$$

The number of neighbors k is much easier to adjust than the regularization parameter γ , because k is an integer with an explicit meaning. For each sample, we can use (5) to assign its neighbors and obtain a sparse $z_i^{(1)}$ that has exactly k nonzero values.

C. INTEGRATING INTRA-VIEW AND INTER-VIEW SIMILARITIES

After getting the intra-view similarity of two views respectively, $Z^{(1)}$ and $Z^{(2)}$, we use $Z = W^{(1)}Z_{ij}^{(1)} + W^{(2)}Z_{ij}^{(2)}$, $Z \in R^{n \times l}$ to represent the similarity of all samples and anchors in the two views in order to make full use of the information. In Fig 2, matrices $W^{(1)}$ and $W^{(2)}$ can help to realize this inter-view fusion. $W^{(1)}$ and $W^{(2)}$ represent the fusion weights of view-1 and view-2 respectively. $W_i^{(1)} = 1$ if the i -th sample

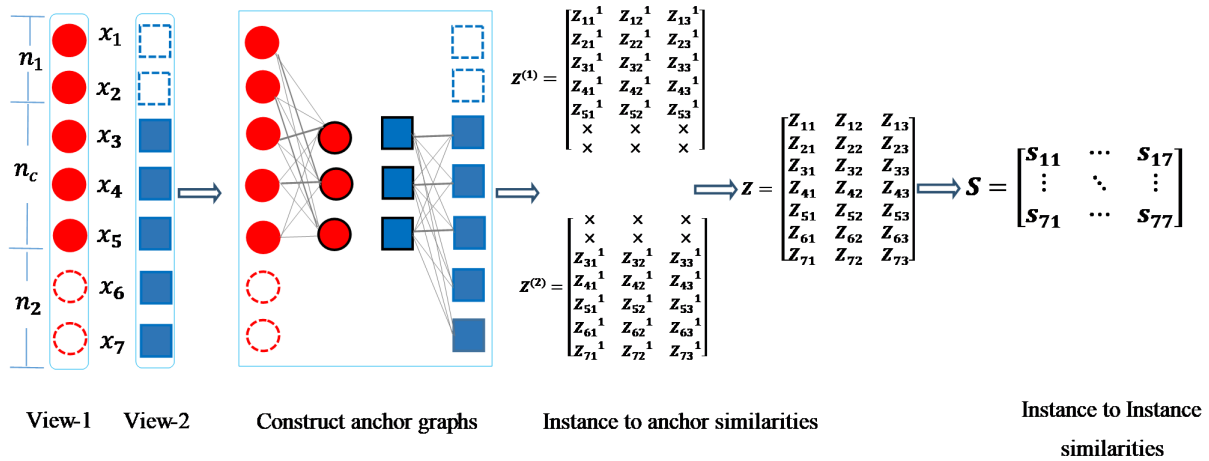


FIGURE 3. The solution process of the consensus similarity matrix.

only appears in view-1, and $W_i^{(1)} = 0$ if the i -th sample only appears in view-2.

The non-missing samples in a view represent the information that we can use. The more information a view provides, the greater the weight of this view. Moreover, the view with more missing samples should be assigned a smaller weight to help achieve a highly reliable and consistent representation and to reduce the negative impact of incomplete views. If a large difference is observed in the sample missing rates of multiple views and if the weights are equally assigned, then those views with too many missing samples may provide too much inaccurate information and affect the final clustering results. Therefore, in our method, when the i -th sample appears in both views, $W_i^{(1)}$ can be expressed in the following formula.

$$W_i^{(1)} = (n_c + n_1)/(n_c + n_1 + n_c + n_2) = (n_c + n_1)/(n_c + n) \quad (6)$$

In a similar way, $W^{(2)}$ can be computed. $W_i^{(1)} + W_i^{(2)} = 1$. The fusion similarity matrix $S \in R^{n \times n}$ among all samples in the two views can be obtained by [27].

$$S = Z\Lambda^{-1}Z^T, \quad \Lambda \in R^{l \times l}, \quad \Lambda = \text{diag}(Z^T 1)$$

S combines intra-view and inter-view similarities.

D. SPECTRAL CLUSTERING

After obtaining the fused similarity matrix S , we perform spectral clustering to obtain the final clustering result.

Spectral clustering learns a low-dimensional representation $F \in R^{n \times c}$ for clustering according to the consensus matrix S . Spectral clustering minimizes the problem (7) through feature decomposition on L , obtains the corresponding c minimum feature vectors, and then performs k -means clustering to obtain the clustering result.

$$\min_{F^T F = I} \text{Tr}(F^T L F), \quad (7)$$

where $\text{Tr}(\cdot)$ is the trace of the matrix, and c is the number of clusters of the cluster. $L = D - S$ is a Laplacian matrix [28], $D \in R^{n \times n}$ is a diagonal matrix with $D_{ii} \in \sum_{j=1}^n S_{ij}$ and I is the identity matrix.

Algorithm 1 AAPMC

Input: Partial multiview data $\{X^{(1,2)}, X^{(1)}, X^{(2)}\}$; the number of nearest anchors k ; the number of clusters c ;

Output: The cluster indicators;

- 1: Select the common part as anchors, the number of anchors is $l = n_c$.
 - 2: Use formula (5) to calculate intra-view similarities $Z^{(1)}$ and $Z^{(2)}$.
 - 3: Weighted fusion $Z^{(1)}, Z^{(2)}$, get $Z \in R^{n \times l}$.
 - 4: Calculate the fusion similarity matrix $S = Z\Lambda^{-1}Z^T$.
 - 5: Perform spectral clustering to obtain the results.
-

E. EXTENSION FOR MULTIPLE VIEWS

The AAPMC method we proposed can not only handle partial multi-view clustering of two views, but also can be easily extended to more than two partial views. Taking three views as an example, the entire extension process is shown in Fig 4.

Fig 4 considers all possible situations in the incomplete three-view data, which contains three types of examples: missing no view, missing one view, and missing two views. n_c represents the number of instances that appear in all three views. n_{12} represents the number of instances shared by view 1 and view 2. n_{13} and n_{23} have similar meanings. n_v represents the number of instances that only exist in the v -th view.

We first divide this partial three-view case into three two-view subcases. To adjust each subcase so that we can directly conduct two-view anchor-based similarity construction, we rearrange the instances according to their types. After dividing the partial three-view case, we then construct

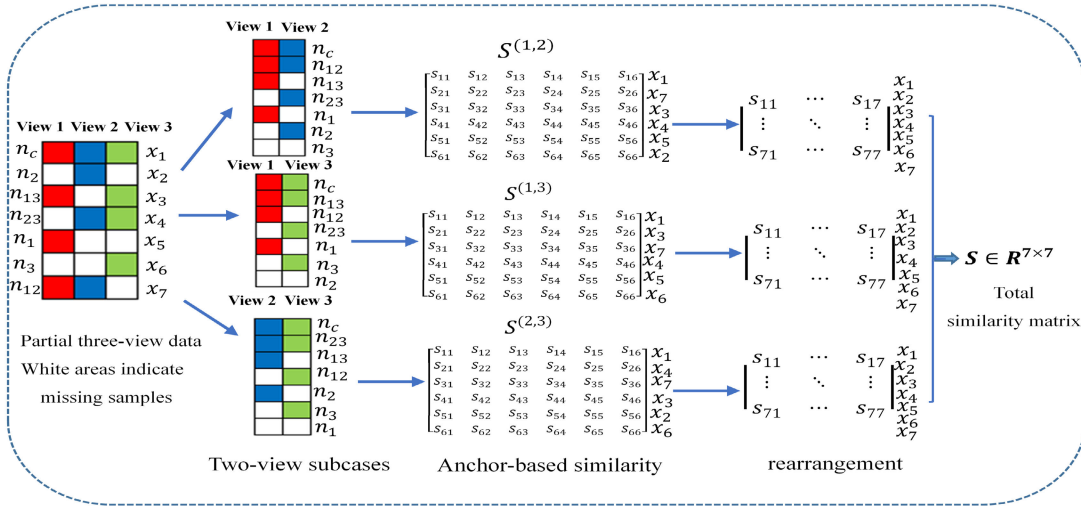


FIGURE 4. Anchor-based similarity reconstruction for partial three-view data.

a similarity matrix for each subcase. The anchor-based similarity reconstruction method can be parallelly applied here. For each partial two-view subcase, we select the common instances present in both views as anchors. Next, we compute a truncated similarity matrix and the corresponding similarity matrix for each subcase. To further fuse the above similarity matrices in three partial two-view subcases, we rearrange them into aligned similarity matrices whose rows and columns follow the original order of instances. Finally, we perform spectral clustering on the consensus matrix S to obtain the clustering results.

F. COMPUTATIONAL COMPLEXITY ANALYSIS

At the first stage of the similarity reconstruction, the time complexity of generating the similarity matrix S is $O(nl \sum_v d_v)$, where n is the total number of samples, l is the total number of anchors and d_v is the feature dimension of the v -th view.

At the second stage of spectral clustering on the fused similarity matrix S , benefiting from the properties of the similarity matrix S , by executing SVD, the time complexity becomes $O(\min\{nl^2, n^2l\})$. If we only need the c largest singular values, the time complexity can be reduced to $O(nc^2)$ [29].

IV. EXPERIMENTS

To verify the effectiveness of AAPMC proposed in this paper, we compare it with eight advanced methods on five datasets.

A. DATASETS

Oxford Flowers Dataset (Flowers17) [30] is made up of 17 flower classes, each with 80 images which are described by color, shape, and textures. Following [11],[20], we take the X^2 distance matrix of color and shape features as the two views.

TABLE 2. Description of the used datasets.

Datasets	Samples	Features	Class	View
Flowers17	1360	1360/1360	17	2
Digit	2000	216/76	10	2
USPS-MNIST	500	784/256	10	2
Synthetic	200	2/2	2	2
3Sources	416	3560/3631/3068	6	3

Multiple Features Handwritten Dataset (Digit) [31] has six feature sets of ten classes of digits and each class holds 200 instances, summing up to 2,000 instances. Following [20], we set view-1 as 76 Fourier coefficients of the character shapes, and view-2 as 216 profile correlations.

USPS-MNIST Dataset merges two famous handwritten datasets: USPS [32] and MNIST [33]. USPS includes 9, 298 digit images with the size of 16×16 in ten classes, while there are 70,000 digit images with the size of 28×28 in the MNIST. The same digits in two datasets can be considered as described in two different views. We follow [20] and randomly select 50 images per digit class from each dataset. Consequently, each view comprises 500 instances.

Synthetic Dataset [34] consists of two views. For each view, we select 200 data points from a two-component Gaussian mixture model as instances at random. There are two clusters (i.e., cluster 1 and 2). Specifically, the cluster means are $U_1^{(1)} = [1, 1]$ and $U_2^{(1)} = [4, 2]$ in view-1, $U_1^{(2)} = [1, 3]$ and $U_2^{(2)} = [3, 1]$ in view-2. The corresponding covariances are

$$\sum_1^{(1)} = \begin{bmatrix} 0.3 & 0 \\ 0 & 0.4 \end{bmatrix} \quad \sum_2^{(1)} = \begin{bmatrix} 0.2 & 0.15 \\ 0.15 & 0.35 \end{bmatrix}$$

$$\sum_1^{(2)} = \begin{bmatrix} 0.25 & -0.05 \\ -0.050 & 0.2 \end{bmatrix} \quad \sum_2^{(2)} = \begin{bmatrix} 0.4 & 0.1 \\ 0.1 & 0.3 \end{bmatrix}$$

3Sources Dataset [35] is collected from three online news sources: BBC, Reuters, and The Guardian. In all, there are

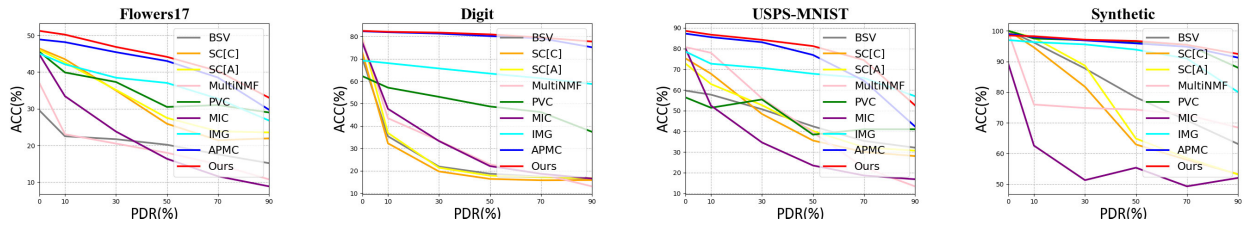


FIGURE 5. The ACC indicator of two-view datasets.

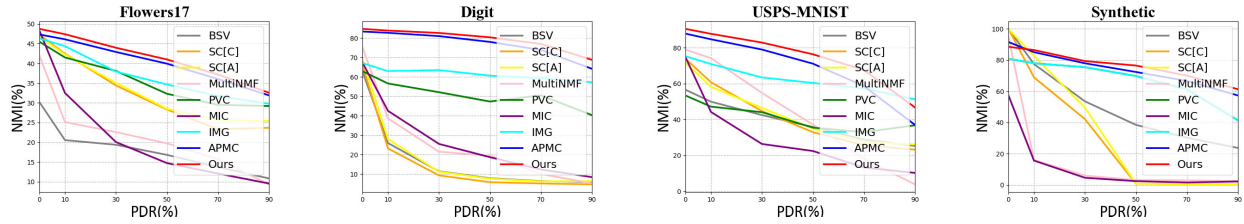


FIGURE 6. The NMI indicator of two-view datasets.

948 news articles including 416 distinct news stories of six topic classes from the period February to April 2009. Among these distinct stories, 169 are reported in all three sources, 194 are in two sources, and 53 appear in a single news source.

B. COMPARISON METHODS

1) BSV (BEST SINGLE VIEW)

As a fundamental method, the best single view (BSV) method firstly fills in the missing values in each feature with the average value of the feature for each view, and then performs clustering on each view and reports the best result.

2) SC[C]

After preprocessing, we concatenate each instance’s features from different views into a single feature vector. Then, we obtain an instance-to-instance similarity matrix and perform spectral clustering.

3) SC[A]

After preprocessing, we first compute an instance to instance similarity matrix for each view. Then, we fuse these similarity matrices by equal-weighted average and perform spectral clustering.

4) MultiNMF [12]

In multiview NMF, a structure sparsity based unsupervised feature selection method is proposed to seek a common latent subspace for multiview clustering. It is designed for complete multiview learning, meanwhile it is the foundation of the following partial multiview algorithms. Therefore, we impute all the missing values as BSV, and then perform the multiview NMF.

5) PVC [10]

The partial multiview clustering is the first work in dealing with partial multi-modal data based on subspace mapping.

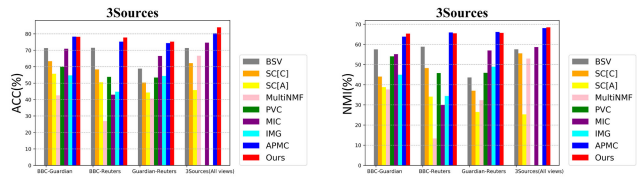


FIGURE 7. The ACC and NMI indicator of three-view dataset.

It is proposed based on NMF to establish a latent sub space, in which the instances described a same sample in different views are close to each other and similar instances in the same view should be well grouped.

6) MIC [11]

Extends MultiNMF via weighted NMF with $L_{2,1}$ regularization.

7) IMG [15]

Incomplete MultiModality Grouping which integrate global structure of data to subspace learning.

8) APMC [20]

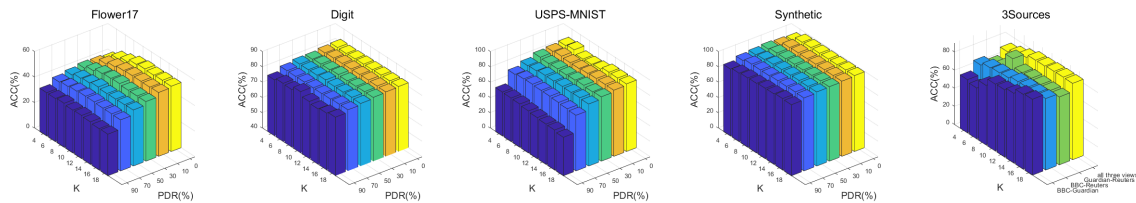
Anchor based Partial multiview Clustering method, which utilizes anchors to reconstruct instance-to-instance relationships for clustering.

C. EXPERIMENTAL SETTINGS

3Sources is a partial multiview dataset, while other datasets are complete. We set Partial Data Ratio (PDR) from 10% to 90% with 20% as interval. 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. Same to APMC, each method was operated 20 times, and the

TABLE 3. The running time (seconds) of different methods on different datasets at 50% PDR.

Dataset	BSV	SC[C]	SC[A]	MultiNMF	PVC	MIC	IMG	APMC	OURS
Digit	3.859	2.125	2.258	0.891	14.407	133.953	704.575	0.744	0.645
Flowers17	17.482	1.998	1.886	2.878	87.931	143.14	756.87	0.807	0.762
USPS-MNIST	0.727	0.465	0.489	0.602	2.98	7.05	9.525	0.29	0.244
Synthetic	0.339	0.302	0.279	0.2	0.561	0.877	1.22	0.206	0.179
3Sources	0.93	1.381	0.737	2.491	–	10.105	–	0.589	0.574

**FIGURE 8.** Influence of the number of nearest anchors k with different PDR settings on ACC.

average was calculated to eliminate the uncertainty caused by randomness.

In order to evaluate the performance of the clustering results, two classic clustering evaluation indicators are adopted in this paper: clustering accuracy (ACC) and standardized mutual information (NMI). These two evaluation indicators measure the clustering algorithm performance by the similarity between the clustering results and the reference results. The values of these two indicators are between 0 and 1, and the larger the value is, the better performance of clustering algorithm will be.

D. RESULTS AND ANALYSIS

Fig 5 and Fig 6 report the results of ACC and MIN values on two-view datasets with different PDR settings, respectively. Fig 7 gives the result of ACC and MIN values on three-view dataset with different PDR settings. Table 3 shows the running time of different methods on different datasets at 50% PDR. It has the same trend under other PDRs. Based on these results, the following observations and discussions are made.

When the dataset is complete, all methods show relatively high performance. As PDR increases, the performance of all algorithms drops, the improvement of our AAPMC over BSV, SC[C], SC[A], MultiNMF, PVC, IMG and APMC becomes larger.

Over two-view data sets, according to the trends of ACC and NMI with varied PDR, the above algorithms can be roughly divided into 3 classes: BSV, SC[C], SC[A], MultiNMF, MIC, and PVC, IMG, APMC, respectively. The first class, SC[C] and BSV perform poorly in most cases, which demonstrates that concatenating all views into a long single view is not a good approach in dealing with the multiview clustering tasks. This is mainly because the differences between views in feature scales and distribution are ignored and the complementary information between different views cannot be utilized. For the second class, as the PDR increases, the value of ACC and NMI of the

SC[A], MIC, and MultiNMF methods decreases significantly, especially on the Digit dataset and the Synthetic dataset. This shows that filling the missing samples with corresponding average samples is not a good way to solve the partial multiview clustering problem. As for the third class, we found that the PVC, IMG, and APMC methods can obtain relatively acceptable performance, which proves that using the complementary information of the view to solve the partial multiview clustering problem is an effective method. Compared with all the above methods, our AAPMC mostly shows better performance. In the case of two views, only in the complete Synthetic dataset, can't our method achieve the best performance, but the gap between our result and the optimal result is very small.

Over 3Sources dataset, our proposed AAPMC method consistently outperforms other competitors in terms of ACC indicator in three two-view cases and one three-view case. In terms of NMI indicator, the three-view case is better than all competitors, but on the two two-view data sets, the APMC algorithm has a better NMI value. When the number of views is increased from two to three, AAPMC produces better results, which indicates that it can be extended to more than two views.

Compared with the previous method, the running time of APMC has been greatly reduced. But our method has the least running time on all data sets, and it still has a certain reduction compared to APMC. Therefore, our method improves the clustering accuracy while further reducing the running time.

E. PARAMETER SENSITIVITY

To further explore the performance of our method, here we also give a parameter study. Our method AAPMC only has one parameter k to be fine-tuned. We set PDR from 0% to 90% as aforementioned, and explore the clustering performance of AAPMC by ranging k within {4,6,...,18}. As shown in Fig 8 and Fig 9, our method is insensitive to k in a relatively wide range.

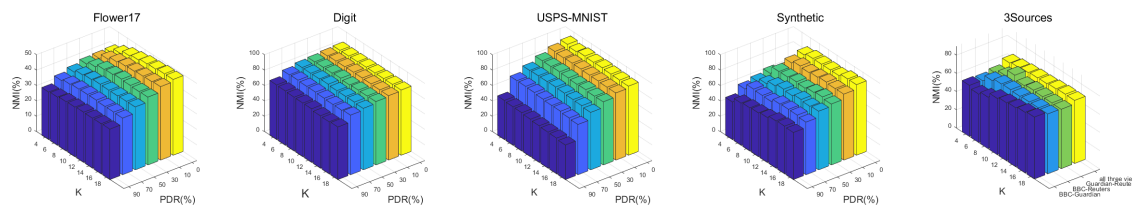


FIGURE 9. Influence of the number of nearest anchors k with different PDR settings on NMI.

V. CONCLUSION

In recent years, the partial multiview clustering problem has been widely studied, and many clustering algorithms have been proposed. However, the existing methods still have the following shortcomings, such as high time complexity, or too many parameters involved, or cannot be extended to more than two views.

In this paper, a partial multiview clustering method AAPMC based on adaptive anchor strategy is proposed, which can solve the shortcomings of previous methods. The experimental results fully validate its superiority.

As for future work, it will be interesting to extend the adaptive anchor-based method to partial multi-view clustering problem where there are no sample among multiple views contains all the view features. In addition, most of the existing incomplete multi-view clustering methods (including our method) require prior of the number of clusters. We plan to refer to COMIC [36] and adjust our method so that clustering can be performed without prior of cluster size.

REFERENCES

- [1] S. Sun, "A survey of multi-view machine learning," *Neural Comput. Appl.*, vol. 23, nos. 7–8, pp. 2031–2038, Dec. 2013.
- [2] C. Xu, D. Tao, and C. Xu, "A survey on multi-view learning," 2013, *arXiv:1304.5634*. [Online]. Available: <http://arxiv.org/abs/1304.5634>
- [3] Q. Yin, S. Wu, and L. Wang, "Unified subspace learning for incomplete and unlabeled multi-view data," *Pattern Recognit.*, vol. 67, pp. 313–327, Jul. 2017.
- [4] J. Salvador and J. R. Casas, "Multi-view video representation based on fast Monte Carlo surface reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3342–3352, Sep. 2013.
- [5] Z. Ding, M. Shao, and Y. Fu, "Robust multi-view representation: A unified perspective from multi-view learning to domain adaption," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 5434–5440.
- [6] X. Yao, J. Han, D. Zhang, and F. Nie, "Revisiting co-saliency detection: A novel approach based on two-stage multi-view spectral rotation co-clustering," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3196–3209, Jul. 2017.
- [7] P. Zhou, Y.-D. Shen, L. Du, F. Ye, and X. Li, "Incremental multi-view spectral clustering," *Knowl.-Based Syst.*, vol. 174, pp. 73–86, Jun. 2019, doi: [10.1016/j.knsys.2019.02.036](https://doi.org/10.1016/j.knsys.2019.02.036).
- [8] S. Wang, X. Liu, E. Zhu, C. Tang, J. Liu, J. Hu, J. Xia, and J. Yin, "Multi-view clustering via late fusion alignment maximization," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 3778–3784.
- [9] N. Rai, S. Negi, S. Chaudhury, and O. Deshmukh, "Partial multi-view clustering using graph regularized NMF," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 2192–2197.
- [10] S. Y. Li, Y. Jiang, and Z. H. Zhou, "Partial multi-view clustering," in *Proc. AAAI*, 2014, pp. 1968–1974.
- [11] W. Shao, L. He, and P. S. Yu, "Multiple incomplete views clustering via weighted nonnegative matrix factorization with $L_{2,1}$ regularization," in *Proc. ECML*, 2015, pp. 318–334.
- [12] J. Liu, C. Wang, J. Gao, and J. Han, "Multi-view clustering via joint nonnegative matrix factorization," in *Proc. SIAM Int. Conf. Data Mining*, May 2013, 252–260.
- [13] M. Hu and S. Chen, "Doubly aligned incomplete multi-view clustering," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, 2262–2268.
- [14] A. Trivedi, P. Rai, H. Daume, and S. L. DuVall, "Multiview clustering with incomplete views," in *Proc. Adv. Neural Inf. Process. Syst. Workshop*, 2010, pp. 1–8.
- [15] H. Zhao, H. Liu, and Y. Fu, "Incomplete multi-modal visual data grouping," in *Proc. IJCAI*, 2016, pp. 2392–2398.
- [16] B. Qian, X. Shen, Y. Gu, Z. Tang, and Y. Ding, "Double constrained NMF for partial multi-view clustering," in *Proc. Int. Conf. Digit. Image Comput., Techn. Appl. (DICTA)*, Nov. 2016, pp. 1–7.
- [17] H. Gao, Y. Peng, and S. Jian, "Incomplete multi-view clustering," in *Proc. Int. Conf. Intell. Inf. Process.* Cham, Switzerland: Springer, 2016, pp. 245–255.
- [18] X. Liu, M. Li, L. Wang, Y. Dou, Y. Dou, J. Yin, J. Yin, and E. Zhu, "Multiple kernel k-means with incomplete kernels," in *Proc. AAAI*, 2017, pp. 2259–2265.
- [19] X. Liu, X. Zhu, M. Li, L. Wang, C. Tang, J. Yin, D. Shen, H. Wang, and W. Gao, "Late fusion incomplete multi-view clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 10, pp. 2410–2423, Oct. 2019.
- [20] J. Guo and J. Ye, "Anchors bring ease: An embarrassingly simple approach to partial multiview clustering," in *Proc. AAAI*, Jan. 2019, pp. 118–125.
- [21] V. R. de Sa, P. W. Gallagher, J. M. Lewis, and V. L. Malave, "Multi-view kernel construction," *Mach. Learn.*, vol. 79, nos. 1–2, pp. 47–71, May 2010.
- [22] W. Liu, J. Wang, S. Kumar, and S. F. Chang, "Hashing with Graphs," in *Proc. ICML*, 2011, pp. 1–8.
- [23] S. Xiang, F. Nie, C. Zhang, and C. Zhang, "Nonlinear dimensionality reduction with local spline embedding," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1285–1298, Sep. 2009.
- [24] M. Ren, Q. Zhang, and J. Zhang, "An introductory survey of probability density function control," *Syst. Sci. Control Eng.*, vol. 7, no. 1, pp. 158–170, Jan. 2019.
- [25] F. Nie, X. Wang, and H. Huang, "Clustering and projected clustering with adaptive neighbors," in *Proc. Knowl. Discovery Data Mining*, 2014, pp. 977–986.
- [26] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [27] W. Liu, J. He, and S. F. Chang, "Large graph construction for scalable semi-supervised learning," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 679–686.
- [28] L. Hagen and A. B. Kahng, "New spectral methods for ratio cut partitioning and clustering," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 11, no. 9, pp. 1074–1085, 1992.
- [29] M. Holmes, A. Gray, and C. Isbell, "Fast SVD for large-scale matrices," in *Proc. NIPS Workshop*, vol. 28, 2007, 249–252.
- [30] M.-E. Nilsback and A. Zisserman, "A visual vocabulary for flower classification," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2006, pp. 1447–1454.
- [31] A. K. Jain, R. P. W. Duin, and J. C. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, Jan. 2000.
- [32] J. J. Hull, "A database for handwritten text recognition research," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 5, pp. 550–554, May 1994.
- [33] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

- [34] J. Guo and W. Zhu, "Partial multi-view outlier detection based on collective learning," in *Proc. AAAI*, 2018, pp. 774–783.
- [35] D. Greene and P. Cunningham, "A matrix factorization approach for integrating multiple data views," in *Proc. ECML/PKDD*, 2009, pp. 423–438.
- [36] X. Peng, Z. Huang, J. Lv, H. Zhu, and J. Zhou, "COMIC: Multi-view clustering without parameter selection," in *Proc. ICML*, 2019, pp. 5092–5101.



XIA JI was born in November 1982. She received the Ph.D. degree. She is currently a Lecturer with the School of Computer Science and Technology, Anhui University. Her research interests include machine learning, data mining, and granular computing.



LEI YANG was born in June 1993. He is currently pursuing the master's degree with the School of Computer Science and Technology, Anhui University. His current research interests include partial multiview clustering and information fusion.



SHENG YAO was born in October 1979. She received the Ph.D. degree. She is currently a Lecturer with the School of Computer Science and Technology, Anhui University. Her research interests include rough set theory and granular computing.

...