IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 13, 2021, accepted January 21, 2021, date of publication February 9, 2021, date of current version February 24, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3056677

A Multi-View Co-Training Clustering Algorithm Based on Global and Local Structure Preserving

WEILING CAI, HONGHAN ZHOU, AND LE XU
Department of Computer Science and Technology, Nanjing Normal University, Nanjing 210097, China
Corresponding author: Weiling Cai (caiwl@nuaa.edu.cn)

This work was supported in part by the Natural Science Foundation of Jiangsu Province under Grant BK20161560, Grant BK20171479, and Grant BK20161020; and in part by the National Science Foundation of China under Grant 61876087 and Grant 61603193.

ABSTRACT Multi-view clustering which integrates the complementary information from different views for better clustering, is a fundamental and important topic in machine learning. In this paper, we present a multi-view co-training clustering algorithm based on global and local structure preserving. Here the global structure is referred to the integration of the *within-cluster compactness* and *between-cluster separation*; the local structure is referred to the neighborhood information. Our algorithm at first preserves both the global and local structure to the subspace in each view. And then, this algorithm obtains the clustering result in the subspace of each view, and utilizes the clustering labels of one view to guide the subspace clustering in another view. In this way, the differences and compatibilities among the multiple views are fused together to form the final cluster partition. Therefore, the clustering result takes full account of the global and local structure information of the multi-view data, which is helpful to improve the of clustering accuracy. Experimental results on the multi-view text datasets and image datasets demonstrate the effectiveness and correctness of the proposed algorithm.

INDEX TERMS Multi-view clustering, subspace learning, co-training.

I. INTRODUCTION

With the popularization of Internet and the coming of big data era, there is a great deal of data generated in the network every day. A great deal of data is generated in the form of multi-view [1]–[3]. For example, the pictures in the web page and the description text of these pictures, videos and subtitles of videos, and so on. Although these data can be clustered and classified in the single view, many references [4]–[6] have proved that the multi-view learning methods can explore information from different views to improve the learning performance. Therefore, multi-view learning is attracting more and more attention. The goal of multi-view learning is to employ the multi-view information to achieve better performance than the single view. So far, multi-view learning has been widely researched and applied in data mining, computer vision and natural language processing.

In real life, data usually exist in the form without class labels. In order to determine the labels, a lot of experienced experts are needed, and a lot of resources are consumed. Therefore, it is particularly important to automatically generate the labels by means of clustering [7]–[9]. In 2018, a subspace clustering method [7] is proposed by introducing structured autoencoder which map input data points into nonlinear latent spaces. In 2020, an end-to-end clustering method [9] is developed by minimizing the discrepancy between pairwise sample assignments for each data point.

When data comes from multiple sources, multi-view clustering utilizes the information from all the views to explore the data structure, and meanwhile maximizes the consistency among the different views. In general, multi-view clustering can be divided into three categories: multi-view clustering based on ensemble [10]–[13], multi-view clustering based on subspace learning [14]–[20], and multi-view clustering based on co-training [21]–[24].

Multi-view clustering based on ensemble [10]–[13] integrates the clustering results obtained in each view to form the final clustering results. In [10], a novel ensemble clustering approach is proposed based on fast propagation of cluster-wise similarities via random walks. In [11], another ensemble clustering approach is developed based on ensemble-driven

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

cluster uncertainty estimation and local weighting strategy. In [12], a new bipartite graph is constructed between objects and base clusters and then efficiently partitioned to achieve the consensus clustering result.

Multi-view clustering based on subspace learning [14]–[20] calculates a unified low-dimensional embedding for multi-view data and then utilizes the clustering algorithm in the low-dimensional space. Kumar and Rai [14] proposed co-regularized multi-view spectral clustering which uses the normalized cut to get the first k eigenvectors of the unified Laplace matrix. Although this method is easy to understand, simple and feasible, the objective function of the method is non-convex and therefore the method often can't get the optimal solution. Because of this drawback, Lu *et al.* [15] proposed a convex sparse spectral clustering algorithm, which not only transforms the original non-convex problem into convex problem, but also improves the clustering accuracy effectively. Yin *et al.* [16] proposed a multi-view clustering method using pairwise sparse subspace representation, which performs sparse representation of each view and maximizes the correlation between different views. Kan *et al.* [17] proposed a supervised multi-view dimensionality reduction algorithm MvDA which aims to find a discriminant common space by jointly learning the specific transformation for multiple views.

Multi-view clustering based on co-training [21]–[25] integrates clustering information obtained from multiple views by using the idea of co-training. The co-training framework has been widely used in the semi-supervised learning. Recently, it becomes more and more popular in multi-view learning. Kumar and Iii [21] proposed a co-training approach for multi-view spectral clustering which uses spectral clustering on individual graphs to get the discriminative eigenvectors in each view, and then uses these eigenvectors to modify the graph structure in the other view. Zhao *et al.* [22] presented an unsupervised multi-view dimensionality reduction algorithm via subspace structure agreement which applies locality preserving projections and co-training strategy to yield the multi-view subspace. Lu *et al.* [24] proposed auto-encoder based co-training multi-view representation learning. It at first trains auto-encoder of each view and then trains a supervised network, and these two stages share the weights partly and assist each other by co-training process.

In recent years, many scholars have considered the fact that multi-view data often has high dimensionality, so a series of algorithms combining dimensionality reduction and multi-view co-training clustering have emerged. A subspace co-training framework for multi-view clustering [23] is one of the typical examples. The hypothesis of this algorithm is that the samples belonging to the same class in different views should be clustered into the same cluster, while the samples from different classes should be grouped into the different clusters. This algorithm performs the clustering method in the subspaces of each view to get the labels, and then exchanges the labels between the different views until the algorithm converges.

In general, the above algorithms effectively utilize the information from multi-view data to achieve the goal of performance promotion. However, most of the multi-view clustering algorithms often focus on the global structure of data distribution, and ignore the local structure information. To overcome this shortcoming, this paper develops a multi-view co-training clustering algorithm based on global and local structure preserving (MCGLP). This algorithm aims to preserve the structural information of the original space to the corresponding subspace, in which the structural information contains the global structure (represented by within-class compactness and between-class separation) and the local structure (represented by the neighborhood information). And then, the clustering analysis is performed respectively in the subspace of each view, and the co-training approach is applied to fuse the clustering information among the multiple views. The co-training process in our algorithm is to take full account of differences and compatibilities in multi-view data to formulate the final cluster partition. Compared with the existing algorithms, the proposed algorithm takes full account of the global structure information and local structure information of the data, which is helpful to improve the accuracy of clustering. Since the algorithm is performed in subspace instead of the original space, the computation is reduced to a certain extent. The effectiveness and correctness of the proposed algorithm are demonstrated by experiments on the multi-view text datasets and image datasets.

II. RELATED WORK

Many algorithms have been developed to solve the multi-view clustering problem. Among them, a subspace co-training framework for multi-view clustering (SCMVC) was proposed by Zhao *et al.*[23]. The algorithm combines the Linear Discriminant Analysis (LDA) and K-means into the co-training framework. The objective function of LDA is:

$$\max_{p} \frac{Tr(P^T S_b P)}{Tr(P^T S_w P)} \tag{1}$$

where S_b represents the between-class scatter, S_w represents the within-class scatter, P represents the projection matrix. The purpose of this objective function is to find the projection matrix which maximizes the ratio of $Tr(P^T S_b P)$ to $Tr(P^T S_w P)$. The objective function of K-means is:

$$\min_{H} \sum_{k} \sum_{i \in C_k} ||X_i - m_k||^2 \tag{2}$$

where H represents the cluster labels, K represents the number of clusters, X_i represents the *ith* sample and m_k represents the *kth* cluster center. Compared with LDA, the objection function of K-means is to find a cluster partition making the samples falling into the same cluster more compact.

SCMVC has the following assumptions: (1) conditional dependence among different views; (2) although each sample has different representations in different views, they should be assigned to the same cluster. The objection function of

co-training subspace multi-view clustering algorithm can be expressed as:

$$\max_{P^{(1)},P^{(2)}} CAI(H^{(1)}, H^{(2)}) \tag{3}$$

where

$$CAI(H^{(1)}, H^{(2)}) = \frac{1}{n}\sum_{i=1}^{n}\delta(h_i^{(1)}, map(h_i^{(2)})) \tag{4}$$

$$H^{(v)} = \arg\min_{H^{(v)}} \sum_{k=1}^{K}\sum_{h_i^{(v)}=k} ||P^{(v)^T}x_i - P^{(v)^T}m_k||^2 \tag{5}$$

n represents the number of samples, K represents the number of clusters, $H^{(v)}$ represents the cluster labels of the vth view, $P^{(v)}$ represents the projection matrix of the vth view, mk represents the cluster center of kth class, $\delta(a, b)$ is a function equal to unity if $a = b$ and zero otherwise, the $map()$ function return an optimal mapping between cluster labels in different views. The goal of the formula (3) is to maximize the consistency of cluster labels in different views. In SCMVC, LDA is firstly utilized to map the samples into the subspace in each view and then K-means is performed in the subspace to get the cluster labels of different views. After that, the obtained cluster labels are exchanged among the different views. The above steps are repeated until the cluster labels in different views tend to be the consistent.

III. THE PROPOSED ALGORITHM

A. ALGORITHM DESCRIPTION

In this paper, a multi-view co-training clustering based on global and local structure preserving (MCGLP) is proposed. This algorithm performs the multi-view clustering by fusing the information obtained from the subspace of each view. These subspaces are resulted by taking into account not only the global structure information but also the local structure information. In MCGLP, the structure information is described as follows: the global structure information is represented by within-class compactness and between-class separation; the local structure information is represented by the neighborhood information. Based on this, the objective function is established to preserve both the global and local structure information to the subspace of each view. After that, K-means clustering method is utilized in subspace clustering, and other suitable clustering algorithm can also be chosen according to the specific circumstance. In addition, MCGLP algorithm integrates the idea of collaborative learning, and employs the subspace clustering results obtained from the current view to guide the description of the global and local structure in other views. In this way, the information among the multiple views is fused to improve the correctness of the final clustering.

1) GLOBAL STRUCTURE INFORMATION

Suppose that input data has V views, the vth view has $d^{(v)}$ dimension. The sample set on the vth view is $X^{(v)} = \{x_1, x_2, \ldots, x_n\}$ and consists of two clusters marked as $C1$ and $C2$. $C1$ is composed of samples $X_1 = \{x_1^1, x_2^1, \ldots, x_{n_1}^1\}$ and $C2$ is composed of samples $X_2 = \{x_1^2, x_2^2, \ldots, x_{n_2}^2\}$. The cluster center is given by:

$$m_i = \frac{1}{n_i}\sum_{x_j \in X_i} x_j, i = 1, 2 \tag{6}$$

The within-cluster scatter is:

$$R_i = \sum_{x_j \in X_i}(x_j - m_i)(x_j - m_i)^T, i = 1, 2 \tag{7}$$

Total within-cluster scatter is:

$$R_w = R_1 + R_2 \tag{8}$$

The between-cluster scatter is:

$$R_b = (m_1 - m_2)(m_1 - m_2)^T \tag{9}$$

In MCGLP, the subspace obtained in each view aims to separate the different clusters as much as possible, meanwhile keep inner clusters compact. To preserve such global information into subspace, the corresponding objective function can be defined as:

$$\max_U \frac{Tr\left(U^T R_b U\right)}{Tr\left(U^T R_w U\right)} \tag{10}$$

where U is the projection matrix and maps the sample into the subspace. The trace of the between-cluster scatter matrix is in the numerator of the formula (10) and the trace of the total within-cluster scatter is in its denominator. In order to maximize this objective function, the numerator should be as large as possible, and the denominator should be as small as possible. To achieve this goal, the data of each subspace view should be compact within each cluster and separate as far as possible between different clusters.

2) LOCAL STRUCTURE INFORMATION

Consider the sample set on the vth view $X^{(v)} = \{x_1, x_2, \ldots, x_n\}$, the local structure information in MCGLP is described by the relationship among these samples, and such relationship is preserved to the subspace by establishing an objective function. The above process is concretely described as the following steps:

1) Constructing the adjacency graph G with n nodes. If sample x_i and x_j are 'close', we should put an edge between node i and j. If x_i and x_j belong to the same cluster and x_i is in k nearest neighbors of j or x_i is in k nearest neighbor of i, the node i and j should be connected by an edge.
2) Choosing the weight W. W is a sparse $n \times n$ symmetric matrix. W_{ij} denotes the weight of edge between node i and j. If there is no edge between node i and j, $W_{ij} = 0$. If node i and j are connected, there are two methods to calculate the weight:

a) Heat kernel method. If node i and j are connected, the weight between node i and j is calculated by $W_{ij} = e^{-||x_i - x_j||^2/t}$.
b) Zero-one method. If there is an edge between node i and j, $W_{ij} = 1$; otherwise, $W_{ij} = 0$.

3) Constructing the objective function to preserve the local structure information:

$$\min_U \frac{Tr\left(U^T X L X^T U\right)}{Tr\left(U^T X D X^T U\right)} \tag{11}$$

where D is a diagonal matrix obtained by $D = \sum_j W_{ij}$. $L = D - W$ is Laplacian matrix.

3) GLOBAL AND LOCAL STRUCTURE INFORMATION PRESERVING

In MCGLP, the objective function to preserve both the global and local structure information in the vth view is further defined as follows:

$$\max_U \frac{Tr\left(U^T (R_b + X D X^T) U\right)}{Tr\left(U^T (R_w + X L X^T) U\right)} \tag{12}$$

It is observed from the formula (12) that the adjustment of the amplitude of U does not affect the direction of U and the value of (12). By setting the denominator of (12) to non-zero constant and maximizing the numerator, the optimization problem of (12) is transformed into:

$$\begin{aligned} &\max\ Tr\left(U^T (R_b + X D X^T) U\right) \\ &\text{s.t. } U^T (R_w + X L X^T) U = I \end{aligned} \tag{13}$$

Employing the method of Lagrange multiplier, the formula of (13) has been changed into:

$$\begin{aligned} L(U, \lambda) = Tr\left(U^T (R_b + X D X^T) U\right) \\ -\lambda_i (U^T (R_w + X L X^T) U - I) \end{aligned} \tag{14}$$

At the extreme point of (14), it should be satisfied:

$$\frac{\partial L(U, \lambda)}{\partial U} = 0 \tag{15}$$

Thus, the value of U should be satisfied:

$$(R_b + X D X^T) U = \lambda_i (R_w + X L X^T) U \tag{16}$$

Getting the eigenvalues $(\lambda_1 > \lambda_2 > \cdots > \lambda_p)$ and the corresponding eigenvectors $(u_1, u_2, \ldots, u_p)$ from:

$$(R_w + X L X^T)^{-1} (R_b + X D X^T) \tag{17}$$

The first $f^{(v)}$ eigenvectors are used to compose the projection matrix U where $f^{(v)}$ is the dimensionality of the subspace in the vth view.

4) FLOWCHART OF PROPOSED ALGORITHM

In order to integrate the global and local clustering information among multiple views, MCGLP uses the idea of collaborative learning. The information exchanging among different views is achieved by exchanging the cluster labels among different views. For illustrative purposes, we only present the execution flow of MCGLP algorithm under two views in Fig. 1.

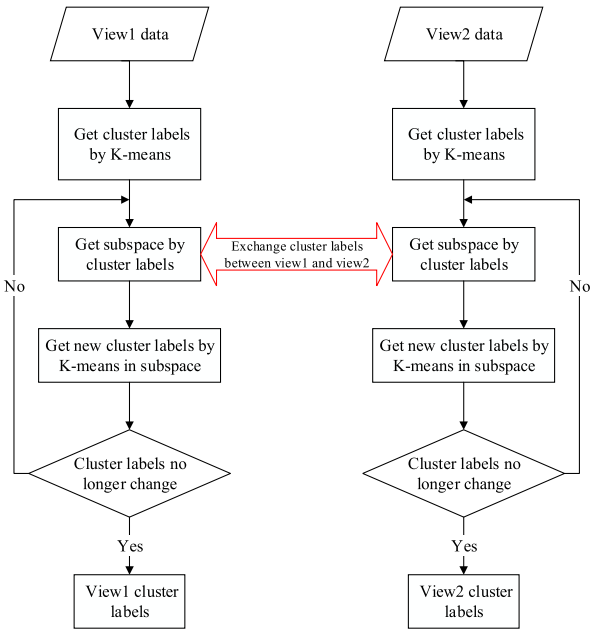

FIGURE 1. Flowchart of proposed algorithm.

The specific steps of the algorithm are as follows: Firstly, get the projection matrix from each view. Secondly, utilize projection matrix to map the original space to its subspace and get cluster labels in subspace by using K-means. Thirdly, exchange the cluster labels among the different views and return to the first step. Solve iteratively until the algorithm converges.

B. ALGORITHM ANALYSIS

1) ALGORITHM PROCEDURE

2) TIME COMPLEXITY

Suppose that the dataset has s views, n samples, the dimensionality of sample in each view is $d^{(v)}$, the dimensionality of the subspace is $f^{(v)}$. k is the number of clusters, g is the iteration number of K-means, and t is the iteration number of MCGLP. The time complexity of modeling global structure can be roughly divided into the calculation of the within-cluster scatter and between-cluster scatter matrix. Its time complexity is $O(nd^{(v)}m^{(v)})$, where $m^{(i)} = \min(n, d^{(v)})$. At the same time, the time complexity of constructing local structure depends on matrix multiplication whose time complexity is $O((d^{(v)})^2 n)$. According to (16), the time complexity of computing the eigenvectors and eigenvalues is $O((m^{(v)})^3)$. Comparing the complexity of the global structure information

Algorithm 1

Input: multi-view dataset $X = \{X^{(v)}|v = 1, 2, \ldots, s\}, X^{(v)} = \{x_1^{(v)} \ldots .x_n^{(v)}\}$, where the dataset X consists of s views and n samples; the number k of clusters;
Output: cluster labels of each view $H = \{H^{(v)}|v = 1, 2, \ldots, s\}, H^{(v)} = \{h_1^{(v)} \ldots .h_n^{(v)}\}$;
Initialize:
1. normalize and centralize the feature vectors in each view;
2. use K-means clustering to get cluster labels $H^{(v)}$ in each view;
3. select k samples closest to each cluster to form $S^{(v)} = \{s_1^{(v)} \ldots .s_k^{(v)}\}$;
for $t = 1$ **to** iter **do**
  for $v = 1$ **to** s **do**
4. exchange the cluster labels between different views, get projection matrix by (17) in each view and project the samples into subspace;
5. use K-means clustering with initial points $S^{(v)}$ to obtain the new cluster labels in each view;
6. choose the samples closest to the clusters to update $S^{(v)}$;
  end for
end for

and the local structure information, it can be found that the complexity of calculating local structure is greater than that of global structure. Therefore, the global and local structure construction has the time complexity of $O((d^{(v)})^2 n)$. In general, the time complexity of K-means clustering is related to the number of samples, the dimensionality of the data, the number of clusters and the number of iterations, which can be expressed as $O(knf^{(v)}g)$. Finally, the time complexity of MCGLP algorithm can be expressed as $O((knf_{\max}g + d_{\max}^2 n)st)$, where $f_{\max} = \max(f^{(v)})$ and $d_{\max} = \max(d^{(v)})$, v=1, 2, 3, $\cdots$, s.

3) CHOOSE THE INITIAL POINTS OF K-MEANS CLUSTER

It is generally known that K-means clustering is affected by the initial points, which may lead to the final clusters being trapped in the local optimum rather than the global optimum. Here we perform the experiment on Iris dataset to reflect this phenomenon. Iris dataset is collected by Fisher, and contains 150 samples, 4 attributes, 3 classes. Fig. 2 displays the original dataset, and marks the samples from the different classes with different colors. Red, green, and blue color symbols represent the samples from the first, second and third class, respectively.

Fig. 3 shows the clustering result obtained by K-means when trapped in local optimum. In this figure, the three black stars denote the initial cluster centers, and the three purple diamonds denote the cluster centers obtained by K-means. The red, green and blue symbols represent the clustering results obtained by K-means. From the clustering results of this figure, it can be seen that when two initial cluster centers fall into the same class, K-means will split the original single cluster into two clusters, and merge the original second and third cluster into one cluster. From this result, we can see that the K-means algorithm is unstable, and does not always guarantee the global optimum.

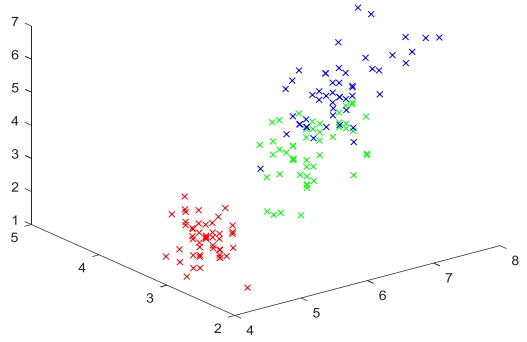

FIGURE 2. Iris dataset with three clusters.

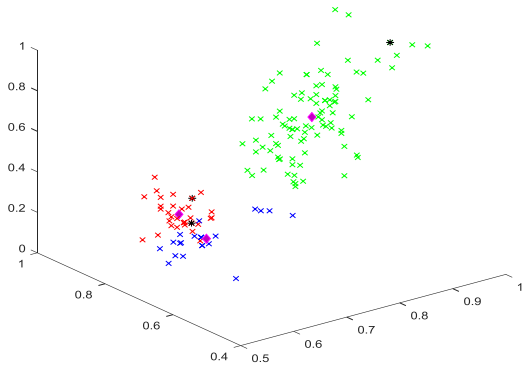

FIGURE 3. Local optimum of the K-means on Iris dataset.

In order to avoid the convergence of the algorithm to the local optimal solution, a multiple-attempt strategy in MCGLP is adopted to select the final cluster centers. Specifically, MCGLP applies K-means multiple times to yield the different clustering centers, and selects the clustering result that makes the objective function minimum, thus avoiding the problem of the local minimum point.

IV. EXPERIMENTS

To verify the correctness and effectiveness of the proposed algorithm, the experiments are performed on the synthetic dataset, the text dataset and the image dataset.

A. EVALUATION METRICS

In this paper, two clustering evaluation indexes are used respectively: Cluster Accuracy (CA) [26] and Normalized Mutual Information (NMI) [27].

Clustering accuracy (CA) is calculated by matching the cluster labels obtained by cluster algorithm to the correct class labels, and getting the best matching numbers divided

by the total number of the samples.

$$CA = \frac{map(H, S)}{n} \tag{18}$$

where H is the cluster labels obtained by the algorithm, S is the correct class labels, $map()$ function is the best mapping of H and S, and n is the total number of samples.

Mutual information (MI) is a method to measure the mutual dependence between two random variables in probability theory and information theory. Suppose two random variables X and Y. MI determines how similar the joint distribution $p(x, y)$ is to the products of factored marginal distribution $p(x)$, $p(y)$ and can be defined as:

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log\left(\frac{p(x, y)}{p(x)p(y)}\right) \tag{19}$$

Normalized mutual information (NMI) lies between 0 and 1 and larger values indicate more accurate clustering labels. NMI used in this paper is:

$$U(X, Y) = 2R = 2\frac{I(X; Y)}{H(X) + H(Y)} \tag{20}$$

$$\begin{aligned} H(X) &= \sum_{i=1}^{n} p(x_i) I(x_i) = \sum_{i=1}^{n} p(x_i) \log_2 \frac{1}{p(x_i)} \\ &= -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) \end{aligned} \tag{21}$$

where $H(X)$ and $H(Y)$ are entropy of X and Y, respectively.

B. SYNTHETIC MULTI-VIEW DATASET

For the convenience of visual observation, we apply MCGLP on the reconstructed wine dataset to verify the correctness of this algorithm. The wine dataset is a data collection from three different breeds of wine produced in the same area of Italy. The dataset contains 178 samples, 13 attributes. The first, second and third class contains 59, 71, and 48 samples, respectively. Here we split the 13 attributes into two views, the first 6 attributes form the first view, and the last 7 attributes form the second view. The following experiment on the reconstructed wine dataset proves that MCGLP can iteratively adjust the projection matrix, so that it can get better clustering performance in the subspace.

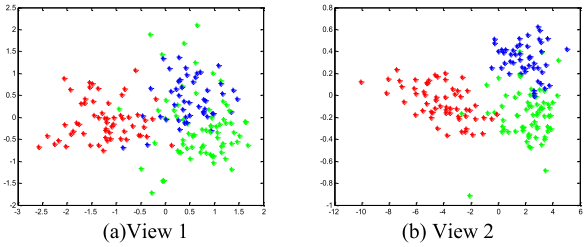

FIGURE 4. Subspaces of two views on wine dataset in the first iteration of MCGLP.

Fig. 4 shows the distribution of the wine dataset in the two views' subspaces when the MCGLP algorithm is first iterated. Fig. 4 (a) and (b) represent the subspaces of View1 and View2, respectively. Red, blue and green symbols represent the samples from three different categories. From Fig. 4, it can be observed that in the early iteration stage of MCGLP, the samples of different categories are mixed in the subspace, especially in View1, and thus it is difficult to distinguish between the green labeled class and the blue labeled class.

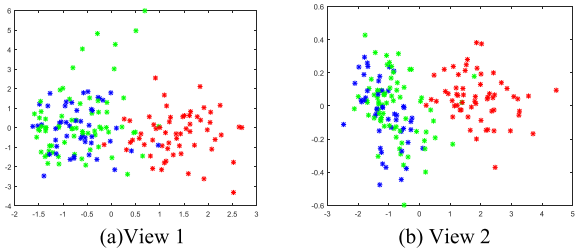

FIGURE 5. Subspaces of two views on wine dataset in the third iteration of MCGLP.

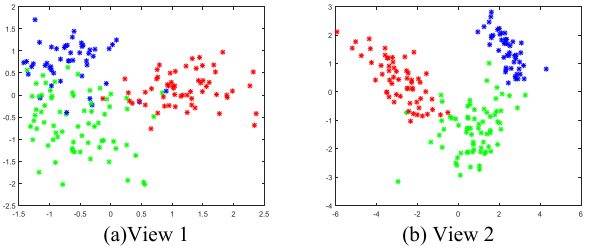

FIGURE 6. Subspaces of two views on wine dataset in the fifth iteration of MCGLP.

Fig. 5 and Fig. 6 show the subspaces of two views for the wine dataset in the third and fifth iteration of MCGLP, respectively. From the result, it can be seen that as the number of iterations increases, the discrimination between the different classes becomes more apparent.

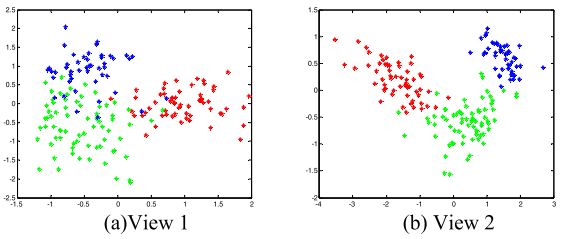

FIGURE 7. Subspaces of two views on wine dataset in the first iteration of MCGLP.

Fig. 7 shows the subspaces of two views on wine dataset in last iteration of MCGLP. Compared with Fig. 5 and Fig. 6, it is found that at the later iteration stage of MCGLP, the samples from different categories can be well differentiated in the two views, and the clustering accuracies of 92.13% and 87.07% are finally obtained in two views, respectively. These experimental results show that MCGLP can employ the clustering information of different views to continuously adjust the subspaces, so that better clustering results are achieved during the iteration.

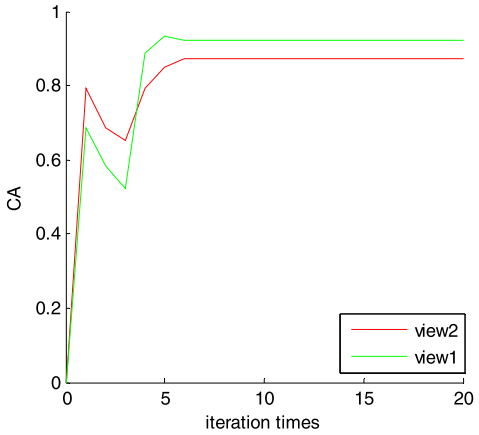

FIGURE 8. CA curves on wine dataset changed as the iteration increases.

Fig. 8 shows the clustering accuracy (CA) curves obtained by MCGLP in two views on the wine dataset as the number of iteration increases. From Fig. 8, we can see that the clustering accuracy does not monotonically increase with the number of iterations, but the final clustering result is better than the initial clustering result.

From the above experimental results on the wine dataset, it can be observed that MCGLP exchanges the clustering information among the different views and continuously adjusts the projection matrix U to make the samples from the different classes in the subspace more easily distinguishable, thereby achieving better clustering results. The above experiment can prove the feasibility and effectiveness of the proposed MCGLP algorithm.

C. REAL-LIFE MULTI-VIEW DATASET

To examine the effectiveness of MCGLP, we select two text datasets and one image dataset. Table 1 shows the basic information of each dataset.

TABLE 1. Basic information of datasets.

	Number of samples	Number of classes	Number of views	Attribute Number of each view
3sources	169	6	3	3560, 3631, 3068
ADS	3279	2	3	457, 495, 472
Mfeat	2000	10	6	216, 76, 64, 6, 240, 47

We choose three algorithms to compare with the proposed algorithm MCGLP. Three algorithms are as follows: Principal component analysis (PCA) plus K-means clustering, Locality Preserving Projection (LPP) plus K-means clustering and subspace co-training multi-view clustering (CoKmLDA). For the sake of fairness, the dimensionality of subspace in all algorithms is set to K-1 where K is the number of clusters.

1) 3SOURCES DATASET

We select 3sources dataset, which is composed of news reported by BBC, Reuters and Guardian. In this dataset, 168 news are reported in three agencies, 194 news are reported in two agencies, and 53 new are reported in single agencies. We choose the news reported in all three agencies to construct a new dataset. There are 6 classes in this dataset which are business, entertainment, health, politics, sport and technology. Each class contains about 28 news. Table 2 and Table 3 show the CA and NMI on the 3sources dataset, respectively.

TABLE 2. CA of four algorithms on 3sources dataset.

	View1	View2	View3	Average
PCA+Kmeans	**0.4852**	0.3964	0.4083	0.43
LPP+Kmeans	0.3491	0.3550	0.3550	0.3531
CoKmLDA	0.3609	0.3609	0.3609	0.3609
MCGLP	0.4497	**0.4497**	**0.4497**	**0.4497**

TABLE 3. NMI of four algorithms on 3sources dataset.

	View1	View2	View3	Average
PCA+K-means	0.2873	0.1928	0.2828	0.2543
LPP+K-means	0.1589	0.1240	0.1089	0.1306
CoKmLDA	0.2813	0.2813	0.2813	0.2813
MCGLP	**0.3218**	**0.3218**	**0.3218**	**0.3218**

TABLE 4. CA of four algorithms on ads dataset.

	View1	View2	View3	Average
PCA+K-means	0.8548	0.8390	0.8591	0.851
LPP+K-means	0.8509	0.7304	**0.8676**	0.8163
CoKmLDA	0.8551	**0.8551**	0.8551	0.8551
MCGLP	**0.8557**	0.8539	0.8567	**0.8554**

The above results in Table 2 and 3 show that the clustering results obtained by MCGLP in 3sources dataset are significantly better than the other three algorithms. Compared with the single-view algorithm (PCA+K-means and LPP+K-means), the multi-view clustering algorithms (CoKmLDA and MCGLP) can achieve better performance than the single-view clustering algorithms; compared with CoKmLDA, MCGLP can obtain better clustering performance due to the incorporation of global and local structure information in dimension reduction.

2) ADS DATASET

This dataset represents a set of possible advertisements on Internet pages. The features encode the image's URL and alt text, the anchor text, words near the anchor text and so on. The task is to predict whether an image is an advertisement or not. The ads dataset consists of 3 views, 3279 samples

TABLE 5. NMI of four algorithms on ads dataset.

	View1	View2	View3	Average
PCA+K-means	0.0072	0.0006	**0.0087**	0.0055
LPP+K-means	0.0003	0.0026	0.0068	0.0032
CoKmLDA	0.0061	**0.0075**	0.0053	0.0063
MCGLP	**0.0120**	**0.0075**	0.0053	**0.0083**

TABLE 6. CA of four algorithms on mfeat dataset.

	View1	View2	View3	View4	View5	View6	Average
PCA+Kmeans	0.5435	0.713	0.3905	0.7325	0.6725	0.5635	0.6026
LPP+Kmeans	0.6465	0.6885	0.3905	0.652	0.6495	**0.622**	0.6082
CoKmLDA	0.5665	0.597	0.543	0.7015	**0.752**	0.6065	0.6278
MCGLP	**0.704**	**0.7555**	**0.593**	**0.7585**	0.745	0.5495	**0.6843**

TABLE 7. NMI of four algorithms on mfeat dataset.

	View1	View2	View3	View4	View5	View6	Average
PCA+Kmeans	0.4865	**0.8124**	0.4726	0.6818	0.6393	0.6007	0.6155
LPP+Kmeans	0.5696	0.7390	0.4780	**0.7908**	0.6493	**0.5998**	0.6378
CoKmLDA	0.5887	0.6330	0.6048	0.6763	0.7207	0.5971	0.6368
MCGLP	**0.6691**	0.7242	**0.6093**	0.7295	**0.7374**	0.5711	**0.6734**

and 2 classes. Table 4 and Table 5 show the CA and NMI obtained by four algorithms on the ads dataset, respectively.

From the results in Table 4 and 5, it can be seen that MCGLP is slightly better than the other three algorithms. At the same time, compared with the single-view algorithm, the multi-view clustering algorithm, whether it is CoKmLDA or MCGLP, has a relatively small difference in CA between different views. These results show that the multi-view clustering algorithm can make full use of the information between different views, making the final clustering results tend to be consistent. Compared with CoKmLDA, MCGLP is slightly better than the CoKMLDA in the average accuracy rate due to the use of local structure information in MCGLP.

3) MFEAT DATASET

Mfeat dataset consists of features of handwritten numeral ('0'-'9') extracted from a set of Dutch utility maps. Every sample has been digitized in binary images. These digits are represented in terms of the following six feature sets: Fourier coefficients of the character shapes, profile correlations, Karhunen-Love coefficients, pixel averages, Zernike moments, and morphological features. This dataset contains 6 views, 2000 samples, 10 classes ('0'-'9'), and each class has 200 samples. Table 6 and Table 7 show the CA and NMI on the Mfeat dataset, respectively.

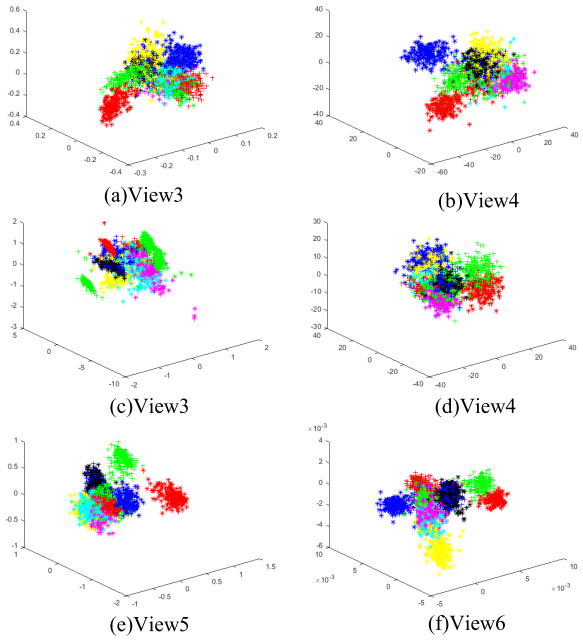

FIGURE 9. Subspaces of the Mfeat dataset obtained by MCGLP.

From the results in Table 6 and 7, it can be seen that MCGLP does not achieve better clustering results than the other algorithms in individual view, but the average accuracy is higher than the other algorithms, indicating that MCGLP is superior to the other three algorithms in clustering performance. Compared with the multi-view algorithms CoKmLDA and MCGLP, although the single-view algorithm can achieve better results in individual views, it does not employ this useful information to improve the clustering performance in other views, so the overall performance is not as good as that of the multi-view clustering algorithm.

Furthermore, since MCGLP uses the idea of cooperative learning to guide the subspace formation, it can preserve the clustering structure of the whole data into the subspaces of each view. To validate this advantage of MCGLP, we take Mfeat dataset as an example and use the visual method to examine the effectiveness of subspaces. Fig. 9 is the final subspaces of each view obtained on Mfeat dataset. In Fig. 9, the symbols with the same color represent the samples from the same class. From the Fig. 9, we can see that samples belongs to the same class are clustered in a shape of 'ball'. The results show that the proposed algorithm keeps global structure information and local structure information, which is helpful to improve the cluster.

4) DISCUSSION OF SUBSPACE DIMENSIONALITY

How to determine the dimension number p of the subspace is very important in dimension reduction. For the clustering-based dimensionality reduction, some researchers [28]–[30] have already done the theoretical analysis and experimental studies on this problem. They have pointed out that in general, mapping the original data into $K-1$ space is a good choice

TABLE 8. NMI of four algorithms on mfeat dataset.

	View1	View2	View3	View4	View5	View6	Ave-rage
1	0.2665	0.2985	0.4535	0.3025	0.3615	0.3280	0.3351
2	0.3265	0.3405	0.5765	0.3955	0.4260	0.3055	0.3951
3	0.3245	0.3575	0.6080	0.4940	0.5280	0.3680	0.4467
4	0.3895	0.4875	0.5735	0.5460	0.5590	0.3495	0.4842
5	0.4990	0.5815	0.5625	0.6500	0.7025	0.6285	0.4993
6	0.5540	0.6440	0.6095	0.6950	0.7060	0.6000	0.6348
7	0.5880	0.5885	0.5905	0.7120	0.7215	0.4810	0.6136
8	0.5945	0.6830	0.6185	0.7605	0.7505	0.4075	0.6357
9	**0.6691**	**0.7241**	0.6092	0.7295	0.7374	0.5711	0.6734
10	0.5035	0.6735	**0.6595**	**0.9020**	**0.7655**	0.6510	**0.6925**
11	0.5590	0.7125	0.5560	0.7630	0.7190	0.5980	0.6513
12	0.5895	0.7240	0.5415	0.8090	0.7370	**0.6640**	0.6775

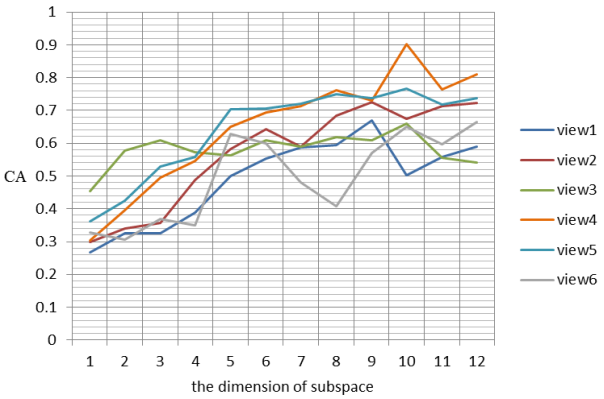

FIGURE 10. CA under different subspace dimensions.

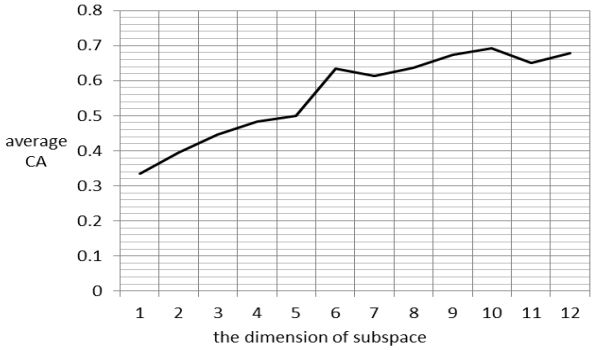

FIGURE 11. Average CA under different subspace dimensions.

where K is the number of clusters. However, $p = K$ is also a good choice in many cases. The additional basis vector can be chosen for a particular emphasis that could help to search for broader configuration space and avoid being prematurely trapped by a local minimum.

In order to verify the above idea, the experiments are carried out on the Mfeat dataset. Specifically, the dimension number p of subspace in MCGLP is set from 1 to 12 and the corresponding clustering accuracies are recorded. Table 8 gives the clustering accuracies of MCGLP obtained on different views under different subspace dimensions. Fig. 10 shows the curves of clustering accuracy on each view as p increases. Fig. 11 shows the changing curve of average clustering accuracy as p increases.

By observing Table 8 and Fig. 10, it can be seen that when the dimension of the subspace is 9, the clustering accuracy of MCGLP on View 1 and View 2 is higher; when the dimension of the subspace is 10, the clustering accuracy of MCGLP on View 3, View 4 and View 5 is higher. From Fig. 11, it is observed that when the dimension of the subspace is set to 10, the average accuracy of MCGLP is the highest; when the dimension of the subspace is greater than 10, the average accuracy does not increase obviously. From this result, we can see that when the dimension of subspace is set to K or $K-1$, MCGLP can achieve better clustering results. The experimental results again support the conclusion of the subspace dimension obtained from the existing literature. In summary, K and $K-1$ are both good choices of the subspace dimension.

V. CONCLUSION

In this paper, we propose a multi-view co-training clustering algorithm based on subspace structure preserving. By preserving the clustering structure into subspace, this algorithm obtains subspace clustering results in a single view and then uses this result to guide subspace clustering in another view. In this approach, the differences and complementarities among the multiple views are fused together to form the final clustering partition. Compared with the existing algorithms, this algorithm considers both the global

structure and local structure together, which is helpful to improve the accuracy of clustering. The results on artificial dataset and multi-view datasets show that the proposed algorithm is effective and can achieve better cluster result than other algorithms. At the same time, we discussed the choice of the dimensions of the subspace, and explained that the dimension of the subspace is suitable for selecting K and K-1, where K is the number of clusters.

REFERENCES

[1] Z. Deng, R. Liu, P. Xu, K.-S. Choi, W. Zhang, X. Tian, T. Zhang, L. Liang, B. Qin, and S. Wang, "Multi-view clustering with the cooperation of visible and hidden views," *IEEE Trans. Knowl. Data Eng.*, early access, Mar. 31, 2020, doi: 10.1109/TKDE.2020.2983366.
[2] F. Nie, G. Cai, J. Li, and X. Li, "Auto-weighted multi-view learning for image clustering and semi-supervised classification," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1501–1511, Mar. 2018.
[3] G.-Y. Zhang, C.-D. Wang, D. Huang, W.-S. Zheng, and Y.-R. Zhou, "TW-Co-K-means: Two-level weighted collaborative K-means for multi-view clustering," *Knowl.-Based Syst.*, vol. 150, pp. 127–138, Jun. 2018.
[4] Q. Wang, Y. Dou, X. Liu, Q. Lv, and S. Li, "Multi-view clustering with extreme learning machine," *Neurocomputing*, vol. 214, pp. 483–494, Nov. 2016.
[5] N. F. Chikhi, "Multi-view clustering via spectral partitioning and local refinement," *Inf. Process. Manage.*, vol. 52, no. 4, pp. 618–627, Jul. 2016.
[6] B. Jiang, F. Qiu, L. Wang, and Z. Zhang, "Bi-level weighted multi-view clustering via hybrid particle swarm optimization," *Inf. Process. Manage.*, vol. 52, no. 3, pp. 387–398, May 2016.
[7] X. Peng, J. Feng, S. Xiao, W.-Y. Yau, J. T. Zhou, and S. Yang, "Structured AutoEncoders for subspace clustering," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 5076–5086, Oct. 2018.
[8] R. Wang, F. Nie, Z. Wang, H. Hu, and X. Li, "Parameter-free weighted multi-view projected clustering with structured graph learning," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 10, pp. 2014–2025, Oct. 2020.
[9] X. Peng, H. Zhu, J. Feng, C. Shen, H. Zhang, and J. T. Zhou, "Deep clustering with sample-assignment invariance prior," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4857–4868, Nov. 2020.
[10] D. Huang, C.-D. Wang, J.-S. Wu, J.-H. Lai, and C.-K. Kwoh, "Ultra-scalable spectral clustering and ensemble clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 6, pp. 1212–1226, Jun. 2020.
[11] D. Huang, C.-D. Wang, H. Peng, J. Lai, and C.-K. Kwoh, "Enhanced ensemble clustering via fast propagation of cluster-wise similarities," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 51, no. 1, pp. 508–520, Jan. 2021.
[12] D. Huang, C.-D. Wang, and J.-H. Lai, "Locally weighted ensemble clustering," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1460–1473, May 2018.
[13] H. Liu and Y. Fu, "Consensus guided multi-view clustering," *ACM Trans. Knowl. Discovery Data*, vol. 12, no. 4, pp. 1–42, 2018.
[14] A. Kumar and P. Rai, "Co-regularized multi-view spectral clustering," in *Proc. Int. Conf. Neural Inf. Process. Syst.* Red Hook, NY, USA: Curran Associates, 2011, pp. 1413–1421.
[15] C. Lu, S. Yan, and Z. Lin, "Convex sparse spectral clustering: Single-view to multi-view," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2833–2843, Jun. 2016.
[16] Q. Yin, S. Wu, R. He, and L. Wang, "Multi-view clustering via pairwise sparse subspace representation," *Neurocomputing*, vol. 156, pp. 12–21, May 2015.
[17] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 188–194, Jan. 2016.
[18] J. Xu, J. Han, F. Nie, and X. Li, "Re-weighted discriminatively embedded *K*-means for multi-view clustering," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 3016–3027, Jun. 2017.
[19] R. Xia, Y. Pan, L. Du, and J. Yin, "Robust multi-view spectral clustering via low-rank and sparse decomposition," in *Proc. AAAI Conf. Artif. Intell.*, 2014, pp. 2149–2155.
[20] X. Zhang, L. Zhao, L. Zong, X. Liu, and H. Yu, "Multi-view clustering via multi-manifold regularized nonnegative matrix factorization," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2014, pp. 1103–1108.
[21] A. Kumar and H. D. Iii, "A co-training approach for multi-view spectral clustering," in *Proc. Int. Conf. Mach. Learn.* Omnipress, 2011, pp. 393–400.
[22] X. Zhao, X. Wang, and H. Wang, "Multi-view dimensionality reduction via subspace structure agreement," *Multimedia Tools Appl.*, vol. 76, no. 16, pp. 17437–17460, Aug. 2017.
[23] X. Zhao, N. Evans, and J.-L. Dugelay, "A subspace co-training framework for multi-view clustering," *Pattern Recognit. Lett.*, vol. 41, pp. 73–82, May 2014.
[24] R. K. Lu, J. W. Liu, Y. F. Wang, H. J. Xie, and X. Zuo, "Auto-encoder based co-training multi-view representation learning," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*. Cham, Switzerland: Springer, 2019, pp. 119–130
[25] C.-K. Lee and T.-L. Liu, "Guided co-training for multi-view spectral clustering," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 4042–4046.
[26] H. Zhao, Z. Ding, and Y. Fu, "Multi-view clustering via deep matrix factorization," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 2921–2927.
[27] X. Cao, C. Zhang, H. Fu, S. Liu, and H. Zhang, "Diversity-induced multi-view subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 586–594.
[28] C. Ding, X. He, H. Zha, and H. D. Simon, "Adaptive dimension reduction for clustering high dimensional data," in *Proc. IEEE Int. Conf. Data Mining*. Washington, DC, USA: IEEE Computer Society, Dec. 2002, pp. 147–154.
[29] C. Ding and T. Li, "Adaptive dimension reduction using discriminant analysis and K-means clustering," in *Proc. 24th Int. Conf. Mach. Learn. (ICML)*, 2007, pp. 521–528.
[30] K. Sim, V. Gopalkrishnan, A. Zimek, and G. Cong, "A survey on enhanced subspace clustering," *Data Mining Knowl. Discovery*, vol. 26, no. 2, pp. 332–397, Mar. 2013.

WEILING CAI received the B.Sc. and Ph.D. degrees in computer science from the Nanjing University of Aeronautics and Astronautics, China, in 2003 and 2008, respectively. She is currently an Associate Professor with the Department of Computer Science and Technology, Nanjing Normal University. Her research interests include machine learning, pattern recognition, data mining, and image processing. In these areas she has published more than 20 technical articles in refereed international journals. She was awarded the Honorable Mention by the Pattern Recognition Magazine.

HONGHAN ZHOU received the B.S. degree from the College of Flight, Shanghai University of Engineering Science, in 2018. He is currently pursuing the master's degree with the Department of Computer Science and Technology, Nanjing Normal University. His research interests include machine learning and multi-view learning.

LE XU received the B.S. degree from the College of Electronics and Computer, Southeast University Chengxian College, in 2018. He is currently pursuing the master's degree with the Department of Computer Science and Technology, Nanjing Normal University. His research interests include machine learning, deep learning, and computer vision.