Bipartite Graph-based Discriminative Feature Learning for Multi-View Clustering

Weiqing Yan
School of Computer and Control Engineering, Yantai University
Yantai, China
wqyan@tju.edu.cn

Jindong Xu
School of Computer and Control Engineering, Yantai University
Yantai, China
xujindong1980@163.com

Jinglei Liu
School of Computer and Control Engineering, Yantai University
Yantai, China
jinglei_liu@sina.com

Guanghui Yue
School of Biomedical Engineering, Health Science Center, Shenzhen University, China
yueguanghui@szu.edu.cn

Chang Tang*
School of Computer Science, China University of Geosciences, Wuhan, China
tangchang@cug.edu.cn

ABSTRACT

Multi-view clustering is an important technique in machine learning research. Existing methods have improved in clustering performance, most of them learn graph structure depending on all samples, which are high complexity. Bipartite graph-based multi-view clustering can obtain clustering result by establishing the relationship between the sample points and small anchor points, which improve the efficiency of clustering. Most bipartite graph-based clustering methods only focus on topological graph structure learning depending on sample nodes, ignore the influence of node features. In this paper, we propose bipartite graph-based discriminative feature learning for multi-view clustering, which combines bipartite graph learning and discriminative feature learning to a unified framework. Specifically, the bipartite graph learning is proposed via multi-view subspace representation with manifold regularization terms. Meanwhile, our feature learning utilizes data pseudo-labels obtained by fused bipartite graph to seek projection direction, which make the same label be closer and make data points with different labels be far away from each other. At last, the proposed manifold regularization terms establish the relationship between constructed bipartite graph and new data representation. By leveraging the interactions between structure learning and discriminative feature learning, we are able to select more informative features and capture more accurate structure of data for clustering. Extensive experimental results on different scale datasets demonstrate our method achieves better or comparable clustering performance than the results of state-of-the-art methods.

*Corresponding author.

MM '22, October 10–14, 2022, Lisboa, Portugal

ACM ISBN 978-1-4503-9203-7/22/10…$15.00
https://doi.org/10.1145/3503161.3548144

CCS CONCEPTS

• **Computing methodologies** → **Feature selection**; • **Information systems** → **Clustering**.

KEYWORDS

Multi-view clustering, Subspace Clustering, Feature Selection

ACM Reference Format:
Weiqing Yan, Jindong Xu, Jinglei Liu, Guanghui Yue, and Chang Tang. 2022. Bipartite Graph-based Discriminative Feature Learning for Multi-View Clustering. In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22), October 10–14, 2022, Lisboa, Portugal.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3503161.3548144

1 INTRODUCTION

Clustering task is one of the most fundamental and important tasks in machine learning and data mining fields, which has been studied for many years. Cluster analysis is a representative unsupervised learning approach which can perform clustering operations on data with unknown label information. It is the process of dividing data samples into multiple clusters according to the relationship between them. It can divide the samples that are close to each other into the same cluster, and divide the samples that are far away into different clusters [34].

With the increasing requirements for data informatization, data information is often collected by multiple sensors or different social media or views to satisfy people requirement[22]. For example, in autonomous driving field, Lidar data and binocular image data are often collected for more accurate 3D object detection task [30]; in scene understanding field, images, text and voice signals are often fused for scene description; in the classification task, an image can be characterized according to different types of features, e.g., SIFT, GIST, LBP, Garbor, and HoG. Such data is known as multi-view data. In multi-view data, each instance object is described by much richer information. For clustering tasks, it is still a challenge that how to improve clustering performance with rich multi-view information [2, 6, 45].

A lots of multi-view clustering methods have been proposed, for example, graph-based multi-view clustering [23, 24, 28, 38, 40, 42, 43, 48, 49], subspace-based multi-view clustering [3, 5, 7, 9, 11,

14, 19, 27, 32, 37, 41, 46, 50, 51, 53, 54]. In general, these methods first compute a common graph similar matrix or self-representation matrix based on each view, and then use graph cutting algorithms or spectral clustering methods on the fused graph matrix or self-presentation matrix to obtain the clustering results [26, 34, 45]. In previous methods, a graph or self-representation similarity matrix is learnt depending on all samples for each view, which often affect the quality of each view similarity matrix due to the involved biases [18]. In addition, the high computation cost is a trouble for the computation of all samples. Moreover, it would take more time if the spectral clustering algorithm was applied to fusion graph or new representation, which involves eigen-decomposition of the Laplacian matrix.

Some researchers [13, 15, 18, 20, 21, 25, 33] are devoted to reducing the algorithm complexity. In these methods, they computed similarity relationship depending on bipartite graph between samples and anchors (landmarks) instead of computing traditional graph between all samples. Hence, these method achieved higher efficiency than those traditional graph-based (all samples) methods by reducing the data scale computation. For example, Kang et al. [13] proposed a structured bipartite graph based multi-view subspace clustering method for large-scale data. Sun et al. [33] proposed unified anchor learning and graph construction learning combination multi-view subspace clustering method. However, these methods paid much attention to structure information learning between samples for clustering, ignored the instinct feature information of each sample to the influence for clustering due to noisy or redundant feature of each sample. In addition, high dimension feature computation will affect efficient of clustering. To reduce redundant features and select informative features, some feature selection methods for multi-view data representation were proposed, for example [8, 31, 35, 36, 47, 52]. These methods fully exploited the underlying correlations between feature and structure information of all samples. However, these methods focused on feature selection based on all samples, it was not suitable to large-scale multi-view feature selection or clustering task.

To address these issues, we propose a novel BIpartite Graph-based Discriminative Feature Learning method for multi-view clustering, termed as BIG-DFL. Different to previous methods, we combine the feature learning and bipartite graph structure learning to a framework, where the both learning can be iteratively boosted by using the result of the other one. Specifically, we learn bipartite graph structure via multi-view subspace representation with manifold regularization terms. Meanwhile, the discriminative feature is learnt via analyzing data intrinsic distribution with pseudo-labels obtained by learnt bipartite graph. At last, the manifold regularization terms establish the relationship between constructed bipartite graph and data in discriminative feature space . In this paper, our main contributions are shown in the following:

1) We propose a novel bipartite graph-based feature Learning method for multi-view clustering, which combines bipartite graph learning and discriminative feature learning to a unified framework. By the interactions between the two ways, we are able to select more informative features and capture more accurate graph structure for clustering.

2) Different previous methods, our feature learning utilizes data pseudo-labels obtained by fused bipartite graph to seek projection directions, which make the same label be closer and make data points with different labels be far away from each other. Meanwhile, the learnt discriminative data can affect back on the construction of bipartite graph by manifold regularization terms in multi-view subspace representation. The both learning can be iteratively boosted by taking advantage of the result of the other one.

3) In optimization process, an efficient optimization strategy is employed to avoid to solving the eigen-problem for discriminative feature learning. The Alternating Direction Method of Multipliers (ADMM) strategy guarantees the convergence of the proposed method. Extensive experimental results demonstrate the effectiveness of our method and the superiority than the state- of-the-art methods.

The rest of this paper is organized as follows. Section 2 gives a brief review of the most related work. In section 3, we present the details of the proposed method and optimization algorithm. Experimental results and analysis are shown in Section 4. In Section 5, we provide a conclusion of this paper.

2 RELATED WORK

2.1 Multi-view subspace clustering

For a data set $X^i \in R^{d_i \times n}$, $i = 1, ..., v$, the subspace representation Z^i for the i-th view can be obtained by using the subspace learning method on each single view. Z^i depicts the relationship between samples. The nonzero elements in Z^i denotes the corresponding data points are from the same subspace. In general, MVSC can be described by the following problem [**?** **?**]:

$$\min_{Z^i} \left\| X^i - X^i Z^i \right\|_p + \alpha f(Z^i) \tag{1}$$

With different forms of f, Eq.(1) gives solutions with different properties. Gao et al. [**?**] proposed a consensus self-expression matrix to guarantee the consistence among different views. In addition of constructing a consistent representation, Brbi'c et al. [3] considered the low-rankness and sparsity of Z^i for clustering. Wang et al. [41] proposed latent intactness-aware similarity representation for multi view subspace clustering; By considering the influence of noise and outlier for clustering, Tang et al. [37] constructed a joint affinity graph with the constraints of diversity regularization and low rank. Zhang et al. [50] proposed a jointly learning method for the latent representation and subspace representations to improve the generalization. These methods achieve good performance in terms of clustering accuracy. Nevertheless, their self-expression matrix is calculated by sample to sample, thus the complexity is high, and they are not suitable to subspace clustering task from large-scale data sets.

2.2 Anchor graph-based multi-view clustering

To deal with large scale data, some anchor-based methods were proposed. Li et al. [21] combined anchor graphs of different views by a local manifold fusion method, and then they employ classical spectral clustering to obtain clustering labels. Its efficient was higher than that of traditional all samples-based multi-view spectral clustering. Liu et al. [25] constructed an augmented view to

balance the disagreement of different views by left singular matrix of anchor graph of each view. This informative augmented view improves the structure of original anchor graph, thus, it is more efficient compared with [21]. Li et al. [18] learned the consensus anchor graph, the common anchor, and anchor graph of each view in an interactively reinforcing way. Kang et al. [15] proposed an anchor-based multi-view subspace clustering method. It reduced the computational complexity in linear time and improved the clustering performance. Due to ignoring the graph structure, they further proposed a multi-view structured graph learning (MSGL) for scalable multi-view subspace clustering [13]. To select optimal anchor points, sun et al. [33] combined anchor points learning and graph structure learning into a unified framework. These approaches provided more efficient results than other multi-view clustering methods based on graph or self-representation of all samples. Inspired by these methods, we propose bipartite graph-based discriminative feature learning method to improve feature selection efficiency.

2.3 Unsupervised multi-view feature selection

Feature selection is a process concerning selecting informative features, which could largely reduce negative impacts from noise or irrelevant features [52]. Unsupervised feature selection for multi-view data often excavates the irrelevant and complementary information to improve the performance of feature selection. Tang et al. [35] proposed a Consensus learning Guided Multi-View Unsupervised Feature Selection(CGMV-UFS), which constructs a consensus clustering indicator matrix by non-negative matrix factorization (NMF) method with local structure preservation. Dong et al. [8] proposed an Adaptive Collaborative Similarity Learning (ACSL), which learned the common similarity structure and selected feature for each view by dynamically learning between each other. Wan et al. [47] proposed an adaptive similarity embedding method, which consider inter-view local structure and inter-samples global structure by a common embedding projection. Tang et al. [36] proposed a cross-view local structure preserved diversity and consensus semantic representation model(CRV-DCS), which captured the common and distinguishing information of different views by embedding projection with local structure preservation. However, these methods focused on all-samples-based feature selection, it was not suitable to large-scale multi-view feature selection or clustering task.

3 PROPOSED METHOD

In this section, we present our proposed method in detail, i.e., Adaptive Structured Discriminative Learning for Scalable Multi-View Subspace Clustering. To facilitate an accurate understanding, we first introduce some important notations and variables. Given a matrix X, $\|X\|_F$ denotes the Frobenius norm of X, $\|X\|_{1,2} = \sum_{j=1}^{m}\sum_{i=1}^{n}\sqrt{x_{ij}^2}$, $Tr(X) = \sum_{i=1}^{m} x_{ii}$ denotes the trace of X. A multi-view data set with V views and N samples is denoted as $X = \{X^1, X^2, ..., X^V\}$, X^i denotes data matrix of the ith view, its dimension is $f_i \times n$. $X_s = \{X_s^1, X_s^2, ..., X_s^V\}$ denotes the anchor point set, where $X_s^i \in R^{f_i \times n_s}$. The affinity matrix in the i-th view between the anchor X_s^i and X^i is denoted as $Z^i \in R^{n \times n_s}$.

3.1 Formulation

For multi-view data, to address the scalability problem, the structure information Z^i between sample points and anchor points can be obtained by multi-view subspace learning [**?**] [17] as follows:

$$\min_{Z^i} \sum_{i=1}^{V} \left\| X^i - X_s^i Z^{iT} \right\|_F^2 + \alpha \left\| Z^i \right\|_F^2 \qquad s.t.\ \ Z^i \geq 0, Z^i \mathbf{1} = \mathbf{1} \tag{2}$$

For high-dimensional feature data, there exist inevitably some redundant and noisy features, which can affect the results of clustering. Hence, in this paper, we hope to use the structure information of samples to analysis feature information of each sample, moreover, reduce dimension of feature. Linear Discriminant Analysis (LDA)[10] used clustering labels to seek projection directions, which made data points with the same label be closer and made data points with different labels be far away from each other. It is a popular supervised feature extraction method. In this paper, we explore LDA to scalable multi-view unsupervised scenarios by the structure Z^i in Eq.(2). In the above section, data structure Z^i can be obtained, moreover the clustering pseudo-labels T of samples can be obtained by fused bipartite graph structure Z. Then, our discriminative feature analysis is denoted as:

$$P^i = \arg\min_{P^i} \frac{tr(P^i S_t P^{iT})}{tr(P^i S_b P^{iT})} + R\left(P^i\right) \tag{3}$$

S_t, S_b denote total scatter matrix and between-class scatter matrix respectively, which can be calculated based on the fused data structure Z. $R(P^i)$ is a constraint term of P^i. To reduce noisy and redundancy features, improve the efficient for clustering, we hope to find important features and discard unimportant features for clustering. Based on this consideration, we use the term $R(P^i)$ to achieve the feature selection functionality. We hope that each column of P^i in each view could weigh the importance of original feature, i.e., larger values of P^i correspond to the discriminative features, smaller or 0 values of P^i correspond to the unimportant features. Hence, we employ $l_{1,2}$ -norm regularization on P to make the column sparsity of P^i. $R(P^i)$ is denoted as: $R(P^i) = \|P^i\|_{1,2}$. By this equation, we can carry out feature selection according to the l_2-norm value of each column of P^i.

During feature selection process, we hope the selected features can influence backward on the graph structure learning process by the local manifold geometrical structure preservation. i.e., when the distance between the sample point $P^i x_p^i$ and the anchor point $P^i x_q^i$ is small, the similarity z_{pq} should be large. The objective is:

$$\begin{aligned} &\min_{Z^i, P^i} \sum_{i=1}^{V}\sum_{p=1}^{N}\sum_{q=1}^{N_s} z_{pq}^i \left(\frac{1}{2}\|P^i x_p^i - P^i x_q^i\|^2\right) \\ &= \min_{Z^i, P^i} \sum_{i=1}^{v} Z^i \odot D(P^i) \end{aligned} \tag{4}$$

where $x_p^i \in X^i, x_q^i \in X_s^i$. $\odot$ indicates the hadamard product. $D(P^i) = [d_{pq}^i] \in R^{n \times n_S}$, $d_{pq}^i = \frac{1}{2}\|P^i x_p^i - P^i x_q^i\|^2$ is the distance between sample point $P^i x_p^i$ and anchor point $P^i x_q^i$.

Combining subspace term (2), discriminative feature term (3), graph regularization term (4), the unified optimization is proposed

as follows:

$$\begin{aligned}&\min_{P^i,Z^i}\sum_{i=1}^{V}\left\|X^i-X_s^iZ^{iT}\right\|_F^2+\alpha\left\|Z^i\right\|_F^2+\frac{tr(P^iS_tP^{iT})}{tr(P^iS_bP^{iT})}\\&+\|P^i\|_{1,2}+\beta Z^i\odot D(P^i)\\&s.t.\ \ Z^i\geq 0,Z^i\mathbf{1}=\mathbf{1}\end{aligned}\tag{5}$$

In this equation, S_t and S_b is total scatter matrix and between-class scatter matrix respectively. They can be calculated by data pseudo labels. We use the rich graph structure information Z_i contained in all views to get pseudo labels. Traditional clustering methods based on bipartite graph [39] implement spectral clustering to the following adjacency matrix $\overline{S}$:

$$\overline{S}=\frac{1}{V}\sum_{i=1}^{V}S^i=\frac{1}{V}\sum_{i=1}^{V}\widehat{Z^i}\widehat{Z^i}^T,\ \ where\ \ \widehat{Z^i}=Z^i\Lambda^{-1/2}\tag{6}$$

where $\Lambda=diag(\sum_{p=1}^{N}Z_{pq}^i)$. However, the computation of $\overline{S}$ and eigen-decomposition of $\overline{S}$ need to spend at least $O(n^2m+n^2k)$ time. Depending on the theorem proposed by [1], the eigenvectors of $\overline{S}$ is equal to the left singular vectors of $\overline{Z}$, which is denoted as:

$$\overline{Z}=[\widehat{Z^1},\widehat{Z^2},...,\widehat{Z^V}]\tag{7}$$

where $\overline{Z}\in R^{n\times n_SV}$, its dimension is less than the dimension of $\overline{S}$. Hence, our pseudo-labels can be obtained by performing SVD on $\overline{Z}$ and k-means instead of performing eigen-decomposition on $\overline{S}$, which reduces computational cost.

3.2 Optimal Solution

In this section, we introduce the optimization steps of the proposed method in detail. An alternating direction method of multipliers(ADMM) algorithm is employed to optimize the problem(5). Next, we show the optimal process in detail.

Z-subproblem:

$$\begin{aligned}&\min_{Z^i}\sum_{i=1}^{V}\left\|X^i-X_s^iZ^{iT}\right\|_F^2+\alpha\left\|Z^i\right\|_F^2+\beta Z^i\odot D(P^i)\\&s.t.Z^i\geq 0,Z^i\mathbf{1}=\mathbf{1}\end{aligned}\tag{8}$$

where $\mathbf{1}$ denotes a column vector with all elements 1. To solve this problem, we can convert this equation by each sample representation, i.e.,

$$\begin{aligned}&\min_{Z^i}\sum_{i=1}^{V}\left(\sum_{p=1}^{n}\left\|x_p^i-X_s^iz_{p:}^T\right\|^2+\sum_{p=1}^{n}\sum_{q=1}^{n_s}\alpha z_{pq}^iz_{pq}^i+\beta z_{pq}^id_{pq}^i\right)\\&s.t.z_{pq}^i\geq 0,\sum_{q=1}^{n_s}z_{pq}^i=1\end{aligned}\tag{9}$$

where x_p^i is the p-th sample of X^i. From this equation, we can observe that the optimization of each row Z^i is uncorrelated. Thus, we can optimize Z^i by rows:

$$\begin{aligned}&\min_{z_{p:}^i}\left\|x_p^i-X_s^iz_{p:}^{iT}\right\|^2+\alpha z_{p:}^iz_{p:}^{iT}+\beta z_{p:}^id_{p:}^{iT}\\&s.t.z_{p:}^i\geq 0,z_{p:}^i\mathbf{1}=\mathbf{1}\end{aligned}\tag{10}$$

where the subscript p : denotes the p-th row of corresponding matrix. This problem, which is a convex quadratic optimization problem, can be easily solved.

P-subproblem:

$$\min_{P^i}\sum_{i=1}^{V}\beta Z^i\odot D(P^i)\ +\frac{tr(P^iS_wP^{iT})}{tr(P^iS_bP^{iT})}+\|P^i\|_{1,2}\tag{11}$$

The first term in the Eq.(11) is convex optimization, its derivative is given by:

$$P^iM^i=P^i\sum_{p=1}^{n}\sum_{q=1}^{n_s}w_{pq}^iw_{pq}^{iT}z_{pq}^i\tag{12}$$

where $w_{pq}^i=x_p^i-x_q^i$.

For the second term of Eq.(11), the scatter matrices S_w, S_b can be calculated based on the pseudo-label T obtained by Eq.(10) and Eq.(7). The optimization of this item requires the eigen-decomposition of matrix, which can take a lot of time and memory. Hence, we resort to the Spectral Regression Discriminant Analysis (SRDA) method [4], which can convert the constraint term to an equivalent regression term. Hence, the second term in the Eq.(11) is converted into:

$$\begin{aligned}&\min_{P^i}\sum_{i=1}^{V}\frac{tr(P^iS_tP^{iT})}{tr(P^iS_bP^{iT})}\\&=\min_{P^i}\sum_{i=1}^{V}\|P^i\bar{X}^i-Y\|_F^2\end{aligned}\tag{13}$$

where, $\bar{X}^i=[\bar{X}_1^i,\bar{X}_2^i,\cdots,\bar{X}_k^i,\cdots,\bar{X}_K^i]\in R^{f^i\times n}$, which is the centered data matrix of the i-th view. $\bar{X}_k^i=[x_1^i-\bar{x}^i,x_2^i-\bar{x}^i,\cdots,x_{m_k}^i-\bar{x}^i]\in R^{f^i\times m_k}$, which denotes the centered data matrix of k-th class in the i-th view. $\bar{x}^i$ denotes the total sample mean vector. m_k denotes the number of data points in k-th class. The row vectors in Y are eigenvectors of the problem $\widetilde{W}y=\lambda y$, where $\widetilde{W}=diag(W_k)$ is a block diagonal matrix. All the elements in $W_k\in R^{m_k\times m_k}$ is equal to $1/m_k$. Next, the Y can be obtained by SRDA algorithm.

SRDA makes the problem (13) easier and more efficient, since it need not the eigen-decomposition of matrix to obtain Y, which is a large save of both time and memory. Depending on the characteristic of block diagonal matrix $\widetilde{W}$, the $K+1$ eigenvectors of $\widetilde{W}$ can be directly obtained:

$$\begin{aligned}&U=[\mathbf{1}^T;u^1;u^2;\cdots;u^k;\cdots;u^K]\\&u^k=[0,\cdots,0,1,\cdots,1,0,\cdots,0]\in R^n\end{aligned}\tag{14}$$

where the number of 1 is m_k in u^k. In [4], the authors had shown the $K-1$ orthogonal eigenvectors are sufficient to represent a K class problem. Depending on U, we can employ the Gram-Schmidt orthogonalization algorithm to obtain the $K-1$ orthogonal eigenvectors $\{y_k\}_{k=1}^{K-1}$. Moreover, the Y in Eq. (13) is

$$Y=[y_1;y_2;\cdots;y_{K-1}]\in R^{(K-1)\times n}\tag{15}$$

For the third term of Eq.(11), inspired by the recent optimization progress on $l_{1,2}$ norm, the term can be efficiently solved by an iterative approach [29]. The minimization of $\|P^i\|_{1,2}$ is a equivalent

problem with the following problem:

$$P^{i(t+1)} = \arg\min_{P^{i(t)}} Tr\left(P^{i(t)}\Lambda^{i(t)}P^{i(t)T}\right) \tag{16}$$

$$\Lambda^{i(t)} = diag(\Lambda^{i(t)}_{j,j}), \Lambda_{j,j} = \frac{1}{\left\|P^{i(t)}_{:j}\right\|} = \frac{1}{\sqrt{P^{iT}_{:j}P^{i}_{:j}}} \tag{17}$$

Overall, the problem (11) is converted as:

$$\begin{aligned} P^{i(t+1)} &= \arg\min_{P^i} \beta Z^i \odot D(P^i) + \\ &\|P^i\bar{X}^i - Y\|_F^2 + Tr\left(P^i\Lambda^i P^{iT}\right) \end{aligned} \tag{18}$$

The optimal solution of (18) can be given as

$$P^{i(t+1)} = Y\bar{X}^{iT}(\bar{X}^i\bar{X}^{iT} + \beta M^i + \Lambda^{i(t)})^{-1} \tag{19}$$

3.3 Convergence and Complexity Analysis

Convergence analysis: As the iteration proceeds, Z sub-problem is strictly convex. For P sub-problem, we take the optimization method in [29] with convergence guarantee. For the overall problem, we use ADMM to optimize it. Therefore, it is easy to achieve convergence of our optimization algorithm. The process of the overall optimization are listed in the following Algorithm 1. In the experimental section, we also show the value variation of objective function with iteration times to further validate the convergency of our algorithm.

Computational complexity: Z_i is solved by the matlab function quadprog, which need $O(nn_s^2fv + nn_s^3v)$. n_s denotes the number of anchors, f is the total number of features, v is the number of views. The computational cost of P_i is $O(kn^2fv)$, k is the number of clustering. To obtain pesodu labels, we need additional time $O(n_s^3v^3 + nk^2t)$ for clustering. In addition, we need $O(nn_stf)$ for the K-means at the process of anchor selection, t is the number of iterations.

Algorithm 1 BIG-DFL

Input: Data matrix X^i, anchors X^i_s, parameter α, β
Output: Data matrix P^i, clustering results
Initialization: P^i
1: **repeat**
2: Update Z^i by solving Eq. (10);
3: Obtain pseudo-labels by clustering depending on $\overline{Z}$ as (7);
4: Obtain Y by Eq.(15);
5: Update P^i as follows:
6: **repeat**
7: compute Λ^i as Eq. (17);
8: update P^i as Eq.(19);
9: **until** convergence;
10: **until** convergence;
11: Select the features by $\|P^i\|_2$;
12: Compute the representation matrix $\overline{Z}$ by Eq.(2) and Eq.(7);
13: Obtain the clustering results by performing SVD on $\overline{Z}$ and k-means.

4 EXPERIMENT

In this section, some experiments are conducted to demonstrate the effectiveness on 9 widely real datasets (Table 1) by comparing with 7 state-of-art methods.

Table 1: Benchmark datasets

Datasets	Samples	Clusters	Views	Feature numbers
3Source[1]	169	6	3	3560,3631,3068
NGs[2]	500	5	3	2000,2000,2000
BBCSport [44]	544	5	2	3183,3203
HW2sources	2000	10	2	784,256
UCI-digitt[3]	2000	10	3	64,76,216
Scene-15 [16]	4485	15	3	20,59,40
Caltech101-all	9144	102	6	48,40,254,1984,512,928
NUSWIDEOBJ	30000	31	5	65,226,145,74,129
YouTubeFace	101499	31	5	64,512,64,647,838

The proposed method combines bipartite graph structure learning of multi-view samples and discriminative feature learning of each sample into a unified framework. Therefore, we compare these two kinds of multi-view clustering methods with the proposed method to show the superiority of our method. These methods focused on feature selection include LS [12], CGMV-UFS[35], CRV-CDS[36], ACSL[8]. These methods focused on anchor graph structure learning include LMVSC[15], MSGL[13], SMVSC [33]. In order to compare with previous anchor graph structure learning approaches fairly, we employ the same anchor selection method with the paper [15] to obtain the X_S in our method.

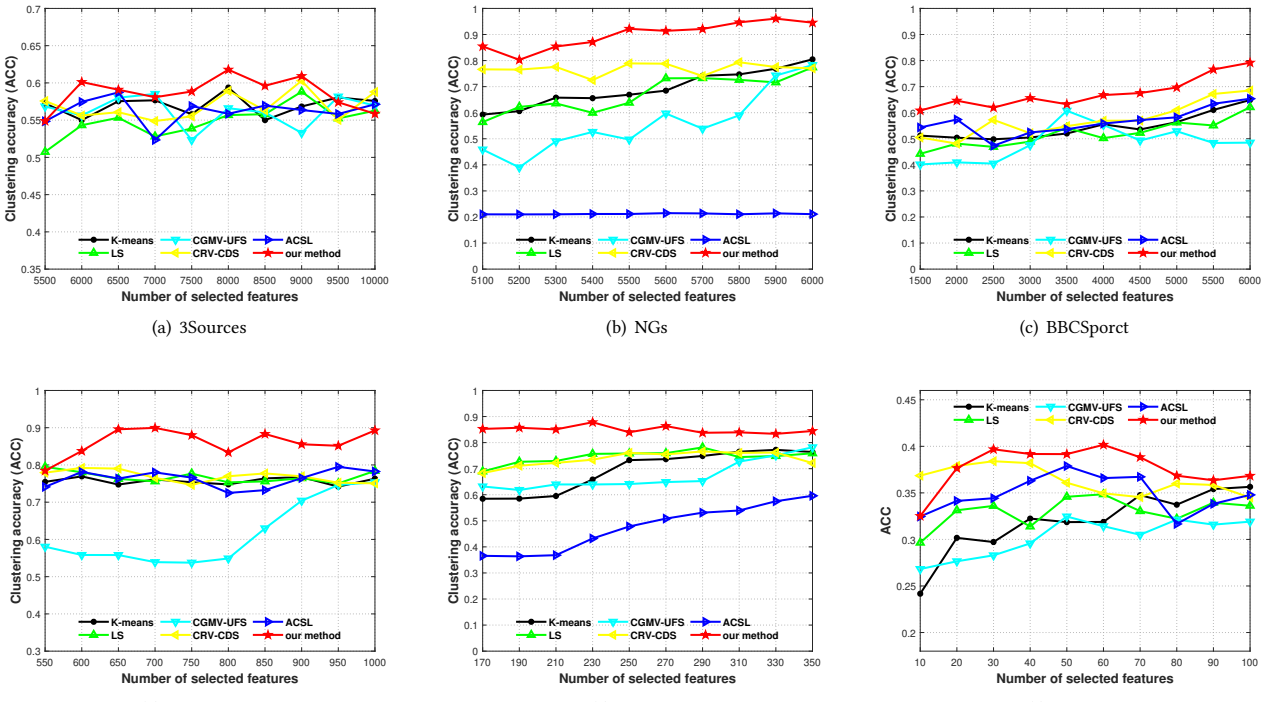
4.1 Experimental results

In this paper, we use accuracy (ACC), normalized mutual information (NMI), and purity to comprehensively evaluate the performance. Higher value indicates better performance. As to each dataset, the best clustering results are obtained from the optimal parameters for all the methods.

Table 2 shows the comparison results of different feature selection methods for multi-view clustering and the proposed method on six datasets. From this table, we can observe that the proposed method obtains better results than those of other methods, whatever in ACC, NMI and Purity. The proposed method exceeds the second best CRV-CDS method the average 8.72%, 6.35%, and 7.52% in terms of ACC, NMI, and Purity, respectively. For NGs dataset, our method exceeds the second best CRV-CDS method 16.7%, 16.5%, and 14.2% in terms of ACC, NMI, and Purity. In addition, we compare the running time of different methods in Table 3. Although LS and CGMU-UFS is more efficient than the proposed method, the ACC, NMI and Purity are lower than the results of our method in all datasets. Compared with CRV-CDS, it can be see that our running time is lower than that of CRV-CDS with increasing with the increase of the number of samples (HW2sources, UCI-Digit, and Scene-15). This reason is our feature selection method is achieved by bipartite graph structure with small anchor points. From this term, the proposed method not only improves the clustering performance, but also reduces computational cost. To further illustrate the effectiveness of our feature selection to clustering, we compare the performance for different methods with different selected numbers of features in Fig. 1. From this figure, it is observed that the proposed method can obtain higher accuracy value than other methods in most case, when the number of selected feature is same.

Table 2: Clustering performance comparison of different feature selection methods.

Metrics	Methods	3Scource	NGS	BBCSport	HW2sources	UCI-Digit	Scene-15
ACC	LS [12]	0.593 ±0.077	0.733 ±0.180	0.628 ±0.122	0.779 ±0.072	0.782 ±0.082	0.336 ±0.014
	CGMV-UFS[35]	0.572 ±0.079	0.791 ±0.144	0.546 ±0.077	0.740 ±0.093	0.782 ±0.078	0.333 ±0.014
	ACSL [8]	0.598 ±0.074	0.218 ±0.016	0.673 ±0.129	0.801 ±0.074	0.600 ±0.051	0.374 ±0.014
	CRV-CDS[36]	0.609 ±0.070	0.794 ±0.165	0.685 ±0.125	0.805 ±0.058	0.785 ±0.087	0.384 ±0.019
	Our method	**0.618 ±0.089**	**0.961 ±0.014**	**0.792 ±0.084**	**0.900 ±0.054**	**0.878 ±0.049**	**0.436 ±0.020**
NMI	LS [12]	0.489 ±0.073	0.633 ±0.204	0.481 ±0.141	0.772 ±0.045	0.771 ±0.046	0.342 ±0.006
	CGMV-UFS [35]	0.473 ±0.045	0.668 ±0.146	0.340 ±0.121	0.757 ±0.054	0.774 ±0.043	0.321 ±0.010
	ACSL [8]	0.498 ±0.059	0.021 ±0.015	0.562 ±0.136	0.767 ±0.042	0.594 ±0.024	0.386 ±0.007
	CRV-CDS [36]	0.508 ±0.075	0.723 ±0.165	0.608 ±0.122	0.795 ±0.045	0.773 ±0.049	0.386 ±0.005
	Our method	**0.543 ±0.054**	**0.888 ±0.030**	**0.671 ±0.086**	**0.847 ±0.026**	**0.821 ±0.027**	**0.404 ±0.005**
Purity	LS [12]	0.700 ±0.054	0.749 ±0.169	0.672 ±0.100	0.812 ±0.056	0.806 ±0.067	0.376 ±0.013
	CGMV-UFS [35]	0.704 ±0.041	0.803 ±0.129	0.596 ±0.089	0.779 ±0.072	0.812 ±0.063	0.375 ±0.012
	ACSL [8]	0.707 ±0.044	0.221 ±0.016	0.719 ±0.104	0.823 ±0.057	0.654 ±0.036	0.415 ±0.014
	CRV-CDS [36]	0.703 ±0.053	0.819 ±0.142	0.754 ±0.096	0.821 ±0.050	0.809 ±0.072	0.416 ±0.014
	Our method	**0.742 ±0.038**	**0.961 ±0.014**	**0.816 ±0.067**	**0.906 ±0.042**	**0.881 ±0.042**	**0.467 ±0.013**



(a) 3Sources (b) NGs (c) BBCSporct

(d) HW2source (e) UCIdigit (f) Scene-15

Figure 1: ACC of different methods with different selected features on different datasets.

Table 4 shows the comparison results of different anchor graph-based learning multi-view clustering methods. It can be seen that these values obtained by the proposed method are higher than those of other methods in terms of ACC, NMI, and purity. In particular, compared with the second best SMVSC method, our gain is 19.55%, 15.38%, 20.35% in average in terms of ACC, NMI, and Purity,respectively. Compared with these methods, the proposed method combines anchor graph structure learning and discriminative feature learning of each sample into a unified framework. By discriminative feature learning, the proposed method can reduce redundant and noisy feature, moreover, capture more accurate structure of data for clustering.

Table 3: The clustering time (s) comparison of different feature selection methods.

Methods	3Scource	NGS	BBCSport	HW2sources	UCI-Digit	Scene-15
LS [12]	0.20	0.06	0.07	0.22	0.22	1.51
CGMV-UFS [35]	0.79	1.13	1.00	2.77	3.68	19.49
ACSL[8]	–	7096.07	199.68	314.84	51.28	691.74
CRV-CDS [36]	42.60	12.52	31.21	66.87	62.59	443.62
Our method	7.21	12.42	13.30	14.84	8.78	19.58

Table 4: Clustering performance comparison of different anchor-based methods.

Metrics	Methods	3Scource	NGS	BBCSport	HW2sources	UCI-Digit	Scene-15
ACC	LMVSC [15]	0.430 ±0.075	0.253 ±0.022	0.487 ±0.084	0.594 ±0.027	0.811 ±0.063	0.330 ±0.013
	MSGL [13]	0.425 ±0.049	0.199 ±0.120	0.434 ±0.075	0.475 ±0.037	0.665 ±0.006	0.331 ±0.016
	SMVSC [33]	0.379	0.706	0.465	0.724	0.808	0.330
	Our method	**0.618 ±0.089**	**0.961 ±0.014**	**0.792 ±0.084**	**0.900 ±0.054**	**0.878 ±0.049**	**0.436 ±0.020**
NMI	LMVSC[15]	0.185 ±0.074	0.069 ±0.022	0.259 ±0.107	0.567 ±0.020	0.760 ±0.029	0.336 ±0.006
	MSGL [13]	0.209 ±0.083	0.068 ±0.050	0.163 ±0.083	0.418 ±0.030	0.641 ±0.005	0.283 ±0.013
	SMVSC [33]	0.476	0.561	0.453	0.675	0.747	0.339
	Our method	**0.543 ±0.054**	**0.888 ±0.030**	**0.671 ±0.086**	**0.847 ±0.026**	**0.821 ±0.027**	**0.404 ±0.005**
Purity	LMVSC [15]	0.690 ±0.096	0.673 ±0.040	0.782 ±0.094	0.620 ±0.022	0.840 ±0.028	0.362 ±0.011
	MSGL [13]	0.715 ±0.092	0.670 ±0.040	0.813 ±0.100	0.503 ±0.035	0.754 ±0.006	0.386 ±0.025
	SMVSC [33]	0.450	0.712	0.509	0.730	0.808	0.343
	Our method	**0.742 ±0.038**	**0.961 ±0.014**	**0.816 ±0.067**	**0.906 ±0.042**	**0.881 ±0.042**	**0.467 ±0.013**

In addition of the above dataset, the proposed method can process the large-scale dataset (Caltech101-all, NUS, YouTubeFace). The time complexity of our method is higher than those of other anchor-based multi-view clustering methods (LMSV, MSGL, SMVSC), if we use all samples to learn bipartite graph structure. Hence, for large-scale datasets, we use small samples as training dataset to obtain discriminative feature and data labels, instead of using all sample data, and then predict left data labels by k-Nearest neighbor(k-NN) based on the selected feature and obtained small sample labels. Specifically, we select small samples (1/2, 1/4, 1/10 of the total number of samples for Caltech101-all, NUS, and YouTubeFace) as training data, and the left data as test data from large scale dataset. After applying the proposed method to training data points, we can obtain the labels of training data points and selected discriminative feature. For test data points, we concatenate all selected features from multi-view test data point, and then employ 1NN to the anchor points to predict their labels. Moreover, we compare with anchor-based clustering method [13, 15, 33]. The results are shown in Table 5. From this table, we can see that the proposed method outperforms the results of [15] in terms of ACC, NMI, Purity. Compared with [33], the proposed method takes less time, although the precision is close in some cases. This reason is that the proposed method takes small number of samples to select feature, and then employs k-NN algorithm to data with selected feature, which improve time efficient.

Overall, compared with anchor-based multi-view clustering methods in Table 4 and Table 5, the proposed method method outperforms the results of other methods in small-scale datasets, and also achieves better or comparable clustering performance in large-scale datasets.

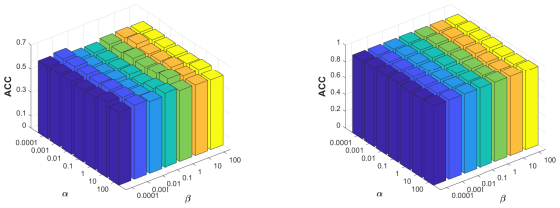
Table 5: Clustering performance comparison on large-scale datasets. The highest and the second highest values under each metric are bolded and underlined, respectively.

Metrics	Method	Caltech 101-all	NUS	YouTube Face
ACC	LMVSC[15]	0.114	0.148	0.139
	MSGL[13]	0.145	0.137	0.146
	SMVSC[33]	**0.275**	<u>0.190</u>	**0.249**
	Our method	<u>0.264</u>	**0.225**	<u>0.243</u>
NMI	LMVSC[15]	0.251	0.126	0.110
	MSGL[13]	0.294	<u>0.133</u>	0.173
	SMVSC[33]	<u>0.351</u>	0.121	**0.228**
	Our method	**0.441**	**0.210**	<u>0.215</u>
Purity	LMVSC[15]	0.189	0.191	0.201
	MSGL[13]	0.228	0.214	0.275
	SMVSC[33]	<u>0.340</u>	**0.233**	**0.337**
	Our method	**0.410**	<u>0.230</u>	<u>0.319</u>
Time	LMVSC[15]	**42.797**	**46.154**	<u>385.115</u>
	MSGL[13]	87.781	65.015	494.243
	SMVSC[33]	75.282	164.953	1378.953
	Our method	<u>44.369</u>	<u>61.718</u>	**178.637**

4.2 Parameter sensitivity and Convergence study

In this section, we show the experiment in parameter sensitivity and convergence study. Due to the limitation of page numbers, we only show a few of them.

In the proposed method, there are two parameters, i.e., α and β. In this subsection, we analyze the sensitive of α and β. The ACC values with different α and β on different datasets are shown in Fig. 2. As can be seen, the result variations of the different parameter on these datasets are very similar. Our method is not very sensitive to parameters α and β. For different datasets, good results can be expected when α= 0.001 and β= 0.0001 in most cases.



(a) 3scource (b) NGS

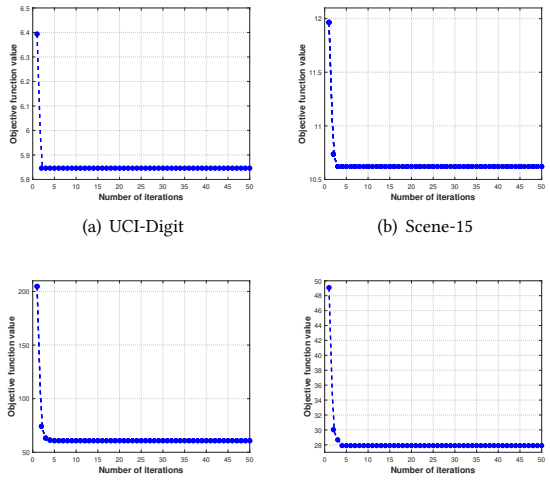
(c) BBCSport (d) HW2sources

Figure 2: ACC values with different α and β on different datasets.

As we analysed in section 3.3, the proposed method converges theoretically. Next, a convergence experimental study for the proposed method is conducted on different datasets. Fig. 3 shows the convergence curves of the proposed method on different datasets. From this figure, we can see that the objective function values converge rapidly to stable value within 5 iterations on these datasets.

5 CONCLUSION

In this paper, we propose a bipartite graph-based discriminative feature learning method for multi-view clustering, which simultaneously consider bipartite graph structure learning and discriminative feature learning to a unified framework. Different previous feature learning based on all samples, we establish bipartite graph structure to learn discriminative feature, which improve the efficient of feature selection. Meanwhile, the learnt discriminative data is formulated as regularization terms for our bipartite graph learning. The joint both learning can be iteratively boosted by using the result of the other one. In discriminative feature optimization process,



(a) UCI-Digit (b) Scene-15

(c) Caltech101-all (d) YouTubeFace

Figure 3: Convergence curves of the proposed method on different datasets.

we employ an efficient optimization strategy to avoid to solving the eigen-problem, which reduce computation cost. In experiment, we compared state-of-the-art multi-view feature selection and anchor-based multi-view clustering methods with the proposed method. Extensive experimental results on different scale datasets show the proposed method obtain better performance.

ACKNOWLEDGMENTS

This work was supported by the Natural Science Foundation of China (No.61801414, 62076228, 62072391, 62001302, 62066013), Guangdong Basic and Applied Basic Research Foundation under Grant No. 2019A1515111205, Natural Science Foundation of Shandong Province No. ZR2021LZH001.

REFERENCES

[1] Séverine Affeldt, Lazhar Labiod, and Mohamed Nadif. 2020. Spectral clustering via ensemble deep autoencoder learning (SC-EDAE). *Pattern Recognition* 108 (2020), 107522.
[2] Steffen Bickel and Tobias Scheffer. 2004. Multi-view clustering. In *ICDM*, Vol. 4. 19–26.
[3] Maria Brbić and Ivica Kopriva. 2018. Multi-view low-rank sparse subspace clustering. *Pattern Recognition* 73 (2018), 247–258.
[4] Deng Cai, Xiaofei He, and Jiawei Han. 2007. SRDA: An efficient algorithm for large-scale discriminant analysis. *IEEE transactions on knowledge and data engineering* 20, 1 (2007), 1–12.
[5] Xiaochun Cao, Changqing Zhang, Huazhu Fu, Si Liu, and Hua Zhang. 2015. Diversity-induced multi-view subspace clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, 586–594.
[6] Kamalika Chaudhuri, Sham M Kakade, Karen Livescu, and Karthik Sridharan. 2009. Multi-view clustering via canonical correlation analysis. In *Proceedings of international conference on machine learning*. 129–136.
[7] Yongyong Chen, Xiaolin Xiao, and Yicong Zhou. 2020. Multi-view subspace clustering via simultaneously learning the representation tensor and affinity matrix. *Pattern Recognition* 106 (2020), 107441.
[8] Xiao Dong, Lei Zhu, Xuemeng Song, Jingjing Li, and Zhiyong Cheng. 2019. Adaptive Collaborative Similarity Learning for Unsupervised Multi-view Feature Selection. In *Proceedings of International Joint Conference on Artificial Intelligence*. IEEE, 2064–2070.
[9] Ruidong Fan, Tingjin Luo, Wenzhang Zhuge, Sheng Qiang, and Chenping Hou. 2020. Multi-view subspace learning via bidirectional sparsity. *Pattern Recognition*

108 (2020), 107524.
[10] Menahem Friedman and Abraham Kandel. 1999. *Introduction to pattern recognition: statistical, structural, neural and fuzzy logic approaches*. Vol. 32. World Scientific Publishing Company.
[11] Hongchang Gao, Feiping Nie, Xuelong Li, and Heng Huang. 2016. Multi-view Subspace Clustering. In *2015 IEEE International Conference on Computer Vision (ICCV)*. IEEE.
[12] Xiaofei He, Deng Cai, and Partha Niyogi. 2005. Laplacian score for feature selection. *Advances in neural information processing systems* 18 (2005).
[13] Zhao Kang, Zhiping Lin, Xiaofeng Zhu, and Wenbo Xu. 2021. Structured Graph Learning for Scalable Subspace Clustering: From Single View to Multiview. *IEEE Transactions on Cybernetics* (2021).
[14] Zhao Kang, Xinjia Zhao, Chong Peng, Hongyuan Zhu, Joey Tianyi Zhou, Xi Peng, Wenyu Chen, and Zenglin Xu. 2020. Partition level multiview subspace clustering. *Neural Networks* 122 (2020), 279–288.
[15] Zhao Kang, Wangtao Zhou, Zhitong Zhao, Junming Shao, Meng Han, and Zenglin Xu. 2020. Large-scale multi-view subspace clustering in linear time. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 4412–4419.
[16] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. 2006. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*. IEEE.
[17] Chun-Guang Li and Rene Vidal. 2015. Structured sparse subspace clustering: A unified optimization framework. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 277–286.
[18] Lusi Li and Haibo He. 2020. Bipartite graph based multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering* (2020).
[19] Ruihuang Li, Changqing Zhang, Qinghua Hu, Pengfei Zhu, and Zheng Wang. 2019. Flexible Multi-View Representation Learning for Subspace Clustering.. In *Proceedings of International Joint Conference on Artificial Intelligence*. 2916–2922.
[20] Xuelong Li, Han Zhang, Rong Wang, and Feiping Nie. 2020. Multi-view clustering: A scalable and parameter-free bipartite graph fusion method. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020).
[21] Yeqing Li, Feiping Nie, Heng Huang, and Junzhou Huang. 2015. Large-scale multi-view spectral clustering via bipartite graph. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.
[22] Yingming Li, Ming Yang, and Zhongfei Zhang. 2018. A survey of multi-view representation learning. *IEEE transactions on knowledge and data engineering* 31, 10 (2018), 1863–1883.
[23] Zhenglai Li, Chang Tang, Xinwang Liu, Xiao Zheng, Guanghui Yue, Wei Zhang, and En Zhu. 2021. Consensus Graph Learning for Multi-view Clustering. *IEEE Transactions on Multimedia* (2021).
[24] Youwei Liang, Dong Huang, and Chang-Dong Wang. 2019. Consistency meets inconsistency: A unified graph learning framework for multi-view clustering. In *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, IEEE, 1204–1209.
[25] Tyng-Luh Liu. 2017. Guided co-training for large-scale multi-view spectral clustering. *arXiv preprint arXiv:1707.09866* (2017).
[26] Yong Liu. 2021. Refined Learning Bounds for Kernel and Approximate k-Means. *Advances in neural information processing systems* (2021).
[27] Shirui Luo, Changqing Zhang, Wei Zhang, and Xiaochun Cao. 2018. Consistent and specific multi-view subspace clustering. In *Thirty-second AAAI conference on artificial intelligence*.
[28] Yanying Mei, Zhenwen Ren, Bin Wu, Yanhua Shao, and Tao Yang. 2021. Robust graph-based multi-view clustering in latent embedding space. *International Journal of Machine Learning and Cybernetics* (2021), 1–12.
[29] Feiping Nie, Heng Huang, Xiao Cai, and Chris Ding. 2010. Efficient and robust feature selection via joint l2, 1-norms minimization. *Advances in neural information processing systems* 23 (2010).
[30] Aditya Prakash, Kashyap Chitta, and Andreas Geiger. 2021. Multi-Modal Fusion Transformer for End-to-End Autonomous Driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 7077–7087.
[31] Weixiang Shao, Lifang He, Chun-Ta Lu, Xiaokai Wei, and S Yu Philip. 2016. Online unsupervised multi-view feature selection. In *Proceedings of IEEE international conference on data mining (ICDM)*. IEEE, 1203–1208.
[32] Xiaomeng Si, Qiyue Yin, Xiaojie Zhao, and Li Yao. 2022. Consistent and diverse multi-View subspace clustering with structure constraint. *Pattern Recognition* 121 (2022), 108196.
[33] Mengjing Sun, Pei Zhang, Siwei Wang, Sihang Zhou, Wenxuan Tu, Xinwang Liu, En Zhu, and Changjian Wang. 2021. Scalable Multi-view Subspace Clustering with Unified Anchors. In *Proceedings of ACM International Conference on Multimedia*. ACM, 3528–3536.
[34] Shiliang Sun. 2013. A survey of multi-view machine learning. *Neural computing and applications* 23, 7 (2013), 2031–2038.
[35] Chang Tang, Jiajia Chen, Xinwang Liu, Miaomiao Li, Pichao Wang, Minhui Wang, and Peng Lu. 2018. Consensus learning guided multi-view unsupervised feature selection. *Knowledge-Based Systems* 160 (2018), 49–60.
[36] Chang Tang, Xiao Zheng, Xinwang Liu, Wei Zhang, Jing Zhang, Jian Xiong, and Lizhe Wang. 2021. Cross-view locality preserved diversity and consensus learning for multi-view unsupervised feature selection. *IEEE Transactions on Knowledge and Data Engineering*. (2021).
[37] Chang Tang, Xinzhong Zhu, Xinwang Liu, Miaomiao Li, Pichao Wang, Changqing Zhang, and Lizhe Wang. 2018. Learning a joint affinity graph for multiview subspace clustering. *IEEE Transactions on Multimedia* 21, 7 (2018), 1724–1736.
[38] Hao Wang, Yan Yang, and Bing Liu. 2019. GMC: Graph-based multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering* 32, 6 (2019), 1116–1129.
[39] Meng Wang, Weijie Fu, Shijie Hao, Dacheng Tao, and Xindong Wu. 2016. Scalable semi-supervised learning by efficient anchor graph regularization. *IEEE Transactions on Knowledge and Data Engineering* 28, 7 (2016), 1864–1877.
[40] Rong Wang, Feiping Nie, Zhen Wang, Haojie Hu, and Xuelong Li. 2019. Parameter-free weighted multi-view projected clustering with structured graph learning. *IEEE Transactions on Knowledge and Data Engineering* 32, 10 (2019), 2014–2025.
[41] Xiaobo Wang, Zhen Lei, Xiaojie Guo, Changqing Zhang, Hailin Shi, and Stan Z Li. 2019. Multi-view subspace clustering with intactness-aware similarity. *Pattern Recognition* 88 (2019), 50–63.
[42] Yiming Wang, Dongxia Chang, Zhiqiang Fu, and Yao Zhao. 2021. Consistent Multiple Graph Embedding for Multi-View Clustering. *arXiv preprint arXiv:2105.04880* (2021).
[43] Yang Wang, Lin Wu, Xuemin Lin, and Junbin Gao. 2018. Multiview spectral clustering via structured low-rank matrix factorization. *IEEE transactions on neural networks and learning systems* 29, 10 (2018), 4833–4843.
[44] Rongkai Xia, Yan Pan, Lei Du, and Jian Yin. 2014. Robust Multi-View Spectral Clustering via Low-Rank and Sparse Decomposition. In *AAAI*. 2149–2155.
[45] Yan Yang and Hao Wang. 2018. Multi-view clustering: A survey. *Big Data Mining and Analytics* 1, 2 (2018), 83–107.
[46] Zhiyong Yang, Qianqian Xu, Weigang Zhang, Xiaochun Cao, and Qingming Huang. 2019. Split multiplicative multi-view subspace clustering. *IEEE Transactions on Image Processing* 28, 10 (2019), 5147–5160.
[47] Wan Yuan, Sun Shengzi, and Cheng Zeng. 2021. Adaptive Similarity Embedding for Unsupervised Multi-View Feature Selection. *IEEE Transactions on Knowledge and Data Engineering* 33, 10 (oct 2021), 3338–3350.
[48] Kun Zhan, Chaoxi Niu, Changlu Chen, Feiping Nie, Changqing Zhang, and Yi Yang. 2018. Graph structure fusion for multiview clustering. *IEEE Transactions on Knowledge and Data Engineering* 31, 10 (2018), 1984–1993.
[49] Kun Zhan, Changqing Zhang, Junpeng Guan, and Junsheng Wang. 2017. Graph learning for multiview clustering. *IEEE transactions on cybernetics* 48, 10 (2017), 2887–2895.
[50] Changqing Zhang, Huazhu Fu, Qinghua Hu, Xiaochun Cao, Yuan Xie, Dacheng Tao, and Dong Xu. 2018. Generalized latent multi-view subspace clustering. *IEEE transactions on pattern analysis and machine intelligence* 42, 1 (2018), 86–99.
[51] Guangyu Zhang, Yuren Zhou, Changdong Wang, Dong Huang, and Xiaoyu He. 2021. Joint representation learning for multi-view subspace clustering. *Expert Systems with Applications* 166 (2021), 113913.
[52] Rui Zhang, Feiping Nie, Xuelong Li, and Xian Wei. 2019. Feature selection with multi-view data: A survey. *Information Fusion* 50 (2019), 158–167.
[53] Qinghai Zheng, Jihua Zhu, Yuanyuan Ma, Zhongyu Li, and Zhiqiang Tian. 2021. Multi-view subspace clustering networks with local and global graph information. *Neurocomputing* 449 (2021), 15–23.
[54] Wencheng Zhu, Jiwen Lu, and Jie Zhou. 2019. Structured general and specific multi-view subspace clustering. *Pattern Recognition* 93 (2019), 392–403.