Efficient Multi-view Unsupervised Feature Selection with Adaptive Structure Learning and Inference

Chenglong Zhang[1] , **Yang Fang**[2] , **Xinyan Liang**[3] , **Han Zhang**[4] , **Peng Zhou**[5] , **Xingyu Wu**[6] , **Jie Yang**[7] , **Bingbing Jiang**[1,*] and **Weiguo Sheng**[1]
[1]Hangzhou Normal University, Hangzhou, China
[2]Chongqing University of Posts and Telecommunications, Chongqing, China
[3]Shanxi University, Taiyuan, China
[4]Northwestern Polytechnical University, Xi'an, China
[5]Anhui University, Hefei, China
[6]Hong Kong Polytechnic University, Hong Kong SAR, China
[7]University of Technology Sydney, NSW, Australia
{clzhang123,liangxinyan48}@163.com, fangyang@cqupt.edu.cn, zhanghan9937@gmail.com, zhoupeng@ahu.edu.cn, xingy.wu@polyu.edu.hk, jie.yang-1@uts.edu.au, jiangbb@hznu.edu.cn, w.sheng@ieee.org

Abstract

As data with diverse representations become high-dimensional, multi-view unsupervised feature selection has been an important learning paradigm. Generally, existing methods encounter the following challenges: (i) traditional solutions either concatenate different views or introduce extra parameters to weight them, affecting the performance and applicability; (ii) emphasis is typically placed on graph construction, yet disregarding the clustering information of data; (iii) exploring the similarity structure of all samples from the original features is suboptimal and extremely time-consuming. To solve this dilemma, we propose an efficient multi-view unsupervised feature selection (EMUFS) to construct bipartite graphs between samples and anchors. Specifically, a parameter-free manner is devised to collaboratively fuse the membership matrices and graphs to learn the compatible structure information across all views, naturally balancing different views. Moreover, EMUFS leverages the similarity relations of data in the feature subspace induced by $l_{2,0}$-norm to dynamically update the graph. Accordingly, the cluster information of anchors can be accurately propagated to samples via the graph structure and further guide feature selection, enhancing the quality of selected features and the computational costs in solution processes. A convergent optimization is developed to solve the formulated problem, and experiments demonstrate the effectiveness and efficiency of EMUFS.

1 Introduction

As information technology develops rapidly, data collected from heterogeneous sources often contains multiple representations [Hu *et al.*, 2019; Zhong and Pun, 2020; Liang *et al.*, 2022b; Liang *et al.*, 2022a; Zhao *et al.*, 2023; Peng *et al.*, 2023]. To process this kind of data, multi-view learning has been proposed in recent years [Hu *et al.*, 2021; Liu *et al.*, 2022; Yang *et al.*, 2024; Xu *et al.*, 2024; Liang *et al.*, 2024]. As a special case, multi-view feature selection that aims to select representative features from the original feature space has become a fundamental task. Depending on the availability of data class labels, current methods can be realized in supervised, semi-supervised, and unsupervised ways [Bai *et al.*, 2021; Jiang *et al.*, 2022; Zhang *et al.*, 2023b; Li *et al.*, 2024b; Zhao *et al.*, 2024]. Considering the expensive cost of manually labeling data, several researchers have devoted to the multi-view unsupervised feature selection, which utilizes the intrinsic data structure to select informative features without the guidance of label information.

To select a feature subset from multi-view data, existing methods can be mainly categorized into two manners. The first kind of method directly concatenates the features from different views and then invokes single-view models on the concatenated features. Typical methods include feature selection with graph learning [Nie *et al.*, 2016; Chen *et al.*, 2023; Tang *et al.*, 2023] and spectral feature selection [Zhao and Liu, 2007; Li and Tang, 2015; Zhou *et al.*, 2023]. Considering that different views have specific properties and contribute variously to final models, this kind of method treats different views equally and neglects the difference between them, causing performance deterioration in practical domains [Tang *et al.*, 2019; Zhong and Pun, 2021]. Instead of simply concatenating different views, another kind of method exploits the underlying correlations among views and introduces weights to balance the contributions of different views. Representative methods include feature selection with adap-

*Corresponding author.

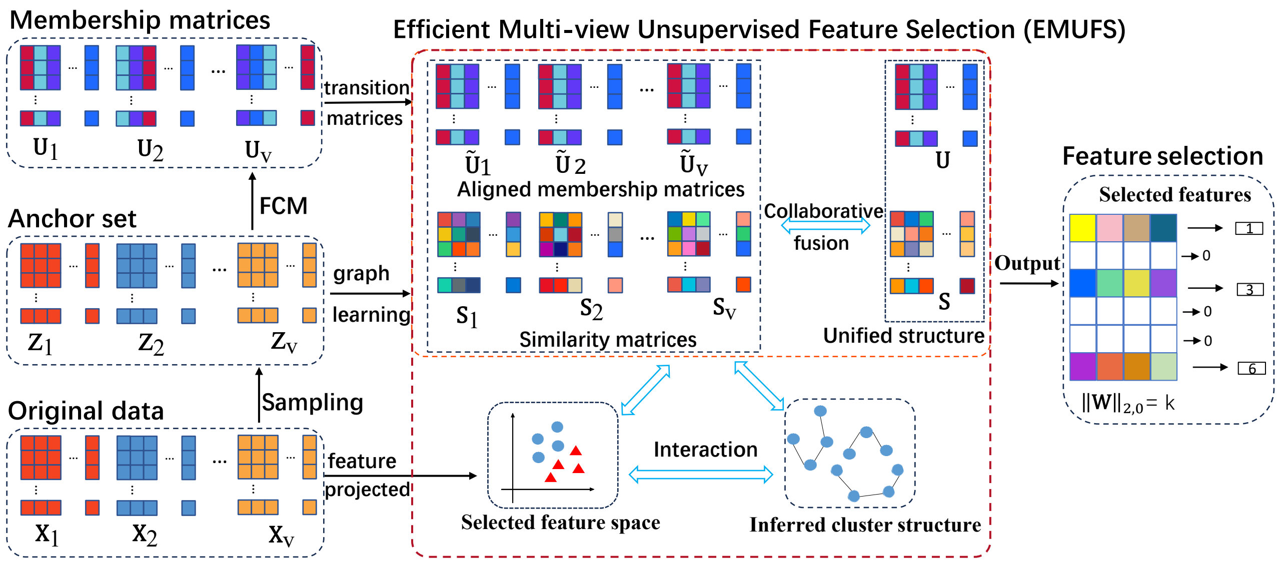


Figure 1: Schematic illustration of EMUFS. Concretely, EMUFS generates anchors and utilizes Fuzzy C-Means (FCM) clustering to mine the view-specific cluster structures of anchors by membership matrices. Afterward, the aligned membership matrices and the bipartite graphs between samples and anchors are collaboratively fused to effectively explore the cluster and similarity structure compatible across all views. Moreover, the cluster information of anchors can be propagated to samples via the bipartite graph, so as to guide the feature selection process and reduce computational costs. Finally, the $l_{2,0}$-norm constraint is imposed on the feature projection matrix to identify the top-k features.

tive similarity and view weight (ASVW) [Hou *et al.*, 2017], multilevel projections with adaptive neighbor graph for multi-view feature selection (MAMFS) [Zhang *et al.*, 2021], and robust feature selection via multi-group adaptive graph representation (MGAGR) [You *et al.*, 2023]. Commonly, these methods follow two separate steps, i.e., constructing a graph on each view to explore view-specific structures and performing feature selection based on the weighted combination of multiple graphs. To avoid the trivial solution of view weights, an extra weight-related parameter that needs to be manually determined is introduced, weakening their applicability [Li *et al.*, 2022]. Moreover, they suffer from a higher computation due to constructing graphs on all samples. Recently, [Chen *et al.*, 2023] proposed to learn a bipartite graph between samples and anchors to reduce computational costs. However, they focus on learning graphs to characterize similarity structures, yet are incapable of capturing the inherent cluster structure of data that can positively guide feature selection.

To explore the cluster structure information, [Feng *et al.*, 2013] proposed to learn the feature projections and the pseudo cluster labels of data (i.e., the cluster indicator matrix) by a linear regression model. Inheriting from [Feng *et al.*, 2013], many variants have been developed in recent years. For example, [Dong *et al.*, 2018] incorporated graph learning into the framework of [Feng *et al.*, 2013], such that the similarity graph can be updated in the feature selection process. Instead of the orthogonal constraint, [Shi *et al.*, 2023] imposed a binary hash constraint on the cluster indicator matrix to improve the discrimination. Despite achieving some progress, these methods still face the following limitations: i) the similarity graph directly derived from the original features is susceptible to poor-quality features since the original data usually contains irrelative features, impairing the graph reliability and finally affecting the effectiveness of selected features; ii) their focus lies in learning the cluster indicator matrix of all samples, such that the optimization procedure involves the decomposition or inverse operations of high-order matrices, leading to expensive computational complexity.

Motivated by the aforementioned issues, we propose an efficient multi-view unsupervised feature selection with adaptive structure learning and inference (EMUFS). Fig. 1 illustrates the basic framework of EMUFS. The main contributions of this paper are summarized as follows:

- We propose an efficient multi-view unsupervised feature selection method that simultaneously leverages membership matrices and bipartite graphs to capture the cluster information of anchors and the similarity structure between samples and anchors, facilitating the ultimate feature selection and reducing computational costs.
- We design a collaborative fusion manner for membership matrices and graphs to learn the compatible structures, so that the cluster information of samples can be inferred via the similarity between samples and anchors, balancing different views without extra parameters.
- We utilize the neighbor relations of data in the selected feature subspace to adaptive learn a unified bipartite graph, reducing the impact of poor-quality features.

2 Our Proposed Methodology

2.1 Notations

Throughout the paper, vectors are written in bold lowercase letters, and matrices are written in bold uppercase letters. Besides, $\mathbf{m}_i$ represents the i-th row of a given matrix $\mathbf{M}$, $\mathrm{Tr}(\mathbf{M})$

Notation	Description
n , m	The numbers of samples and anchors, respec
c, V	The number of classes and views, respective
k	The number of selected features
d_v	The dimension of v-th view
$d = \sum_{v=1}^{V} d_v$	The total dimension of V views
$\mathbf{x}_i^v \in \mathbb{R}^{d_v \times 1}$	The i-th sample in v-th view
$\mathbf{z}_j^v \in \mathbb{R}^{d_v \times 1}$	The j-th anchor in v-th view
$\mathbf{x}_i = [\mathbf{x}_i^1; ...; \mathbf{x}_i^V] \in \mathbb{R}^{d \times 1}$	The i-th sample
$\mathbf{z}_j = [\mathbf{z}_j^1; ...; \mathbf{z}_j^V] \in \mathbb{R}^{d \times 1}$	The j-th anchor
$\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_n] \in \mathbb{R}^{d \times n}$	The concatenated feature matrix of samples
$\mathbf{Z} = [\mathbf{z}_1, ..., \mathbf{z}_m] \in \mathbb{R}^{d \times m}$	The concatenated feature matrix of anchors
$\{\mathbf{U}_v\}_{v=1}^{V} \in \mathbb{R}^{m \times c}$	The initialized membership matrices of anch
$\{\widetilde{\mathbf{U}}_v\}_{v=1}^{V} \in \mathbb{R}^{m \times c}$	The aligned membership matrices of anchor:
$\mathbf{W} \in \mathbb{R}^{d \times c}$	The feature projection/selection matrix
$\mathbf{1} \in \mathbb{R}^{c \times 1}$	The all-one vector

Table 1: Description of Notations

denotes the trace of $\mathbf{M}$ and $\|\mathbf{M}\|_F = \sqrt{\mathrm{Tr}(\mathbf{M}^T\mathbf{M})}$ denotes the Frobenius norm. $\|\mathbf{M}\|_{2,1} = \sum_i \|\mathbf{m}_i\|_2$ and $\|\mathbf{M}\|_{2,0} = \sum_i \|\mathbf{m}_i\|_2^0$ denote the $l_{2,1}$-norm and $l_{2,0}$-norm of $\mathbf{M}$, respectively. Table 1 lists the frequently used notations.

2.2 Similarity Structure Learning and Fusion

Most methods construct n-order graphs to mine the similarity relations among samples, leading to high computational costs [Hu *et al.*, 2022; Li *et al.*, 2024a]. To this end, our proposed EMUFS learns bipartite graphs between samples and anchors to improve the efficiency of graph construction. Specifically, the alternate sampling strategy proposed by [Li *et al.*, 2022] is used to generate a consistent anchor set on different views. Then, the view-specific similarity structures between samples and generated anchors can be captured by solving:

$$\min_{\mathbf{S}_v\mathbf{1}=\mathbf{1},\mathbf{S}_v\geq 0} \sum_{i=1}^{n}\sum_{j=1}^{m} \|\mathbf{x}_i^v - \mathbf{z}_j^v\|_2^2 s_{ij}^v + \tau\|\mathbf{S}_v\|_F^2, \tag{1}$$

where $\mathbf{S}_v \in \mathbb{R}^{n\times m}$ is the bipartite graph on the v-th view, and s_{ij}^v measures the similarity between sample $\mathbf{x}_i$ and anchor $\mathbf{z}_j$ in v-th view. Eq.(1) can be directly solved by the adaptive neighbor strategy [Nie *et al.*, 2014].

To learn a consensus similarity graph, the view-specific graphs can be coalesced in the following fusion manner:

$$\min_{\mathbf{S},\boldsymbol{\phi}} \sum_{v=1}^{V} \phi_v^\eta \|\mathbf{S}-\mathbf{S}_v\|_F^2 \;\; \text{s.t.} \;\boldsymbol{\phi}\geq 0, \boldsymbol{\phi}^T\mathbf{1}=1, \mathbf{S1}=\mathbf{1}, \mathbf{S}\geq 0, \tag{2}$$

where $\mathbf{S}$ denotes a unified graph that compatibly crosses multiple views, and the exponential parameter η controls the distribution of the view weights $\{\phi_v\}_{v=1}^V$. To further reveal the underlying relation between $\mathbf{S}$ and $\{\mathbf{S}_v\}_{v=1}^V$, we fix $\boldsymbol{\phi}$ and set the derivation of Eq. (2) w.r.t. $\mathbf{S}$ to zero:

$$\sum_{v=1}^{V} \phi_v^\eta(\mathbf{S}-\mathbf{S}_v)=0 \Longrightarrow \mathbf{S}=\sum_{v=1}^{V}\alpha_v\mathbf{S}_v; \alpha_v=\phi_v^\eta/\sum_{v=1}^{V}\phi_v^\eta, \tag{3}$$

where α_v can be regarded as the weight since $\alpha_v \geq 0$ and $\sum_{v=1}^{V}\alpha_v = 1$. In Eq. (3), the equality relation of $\mathbf{S}$ and $\sum_{v=1}^{V}\alpha_v\mathbf{S}_v$ can be further relaxed by introducing a flexible regression residue (i.e., $\mathbf{S} - \sum_{v=1}^{V}\alpha_v\mathbf{S}_v$) to measure the mismatch between them, achieving a new fusion paradigm as:

$$\min_{\mathbf{S1}=\mathbf{1},\mathbf{S}\geq 0,\boldsymbol{\alpha}\geq 0,\boldsymbol{\alpha}^T\mathbf{1}=1} \|\mathbf{S} - \sum_{v=1}^{V}\alpha_v\mathbf{S}_v\|_F^2. \tag{4}$$

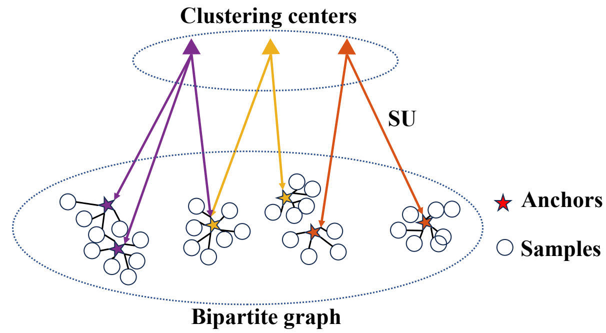


Figure 2: Information propagation illustration, where colored arrows represent the clustering information propagating from anchors to samples.

Eq. (4) can fuse multiple graphs without the exponential parameter η. Further, considering that the poor-quality features in the original space can undermine graph structures, we propose to dynamically update the graph $\mathbf{S}$ according to the relations of samples and anchors in the selected feature subspace:

$$\min_{\mathbf{W},\mathbf{S},\boldsymbol{\alpha}} \sum_{i=1}^{n}\sum_{j=1}^{m} \|\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{z}_j\|_F^2 s_{ij} + \lambda\|\mathbf{S} - \sum_{v=1}^{V}\alpha_v\mathbf{S}_v\|_F^2$$
$$\text{s.t.} \;\; \mathbf{S1}=\mathbf{1}, \mathbf{S}\geq 0, \boldsymbol{\alpha}\geq 0, \boldsymbol{\alpha}^T\mathbf{1}=1, \|\mathbf{W}\|_{2,0}=k. \tag{5}$$

In contrary to previous methods [Nie *et al.*, 2016; Hou *et al.*, 2017] that applied an $l_{2,1}$-norm on the feature projection matrix $\mathbf{W}$, the $l_{2,0}$-norm constraint is imposed on $\mathbf{W}$ to automatically select salient features without sorting features in advance. Moreover, the graph learned from the selected features can accurately explore the similarity structure and alleviate the impacts of low-quality features.

2.3 Cluster Information Inference on Graph

A membership matrix that contains the cluster structure of data is widely used since it can be regarded as the latent discriminative information in unsupervised scenarios [Zhang *et al.*, 2023a; Lu *et al.*, 2023]. For convenience, existing models often generate membership matrices on each view independently and coalesce them for further analyses. Nevertheless, the clusters preserved in the membership matrices might be mismatched between different views, which means that different membership matrices are not aligned column-widely [Wang *et al.*, 2022]. As a result, the direct fusion of membership matrices might suffer from inconsistent cluster information across views. To solve this limitation, the membership matrices on different views are first generated by the FCM [Bezdek *et al.*, 1984] to mine the view-specific cluster structures, then different membership matrices can be aligned by solving the following problem:

$$\min_{\mathbf{T}_v} \|\mathbf{U}_1 - \mathbf{U}_v\mathbf{T}_v\|_F^2 \;\; \text{s.t.} \;\; \mathbf{T}_v^T\mathbf{T}_v = \mathbf{T}_v\mathbf{T}_v^T = \mathbf{I}, \tag{6}$$

where the permutation matrix $\mathbf{T}_v$ makes $\mathbf{U}_1$ and $\mathbf{U}_v$ consistent at the level of clusters and can be directly solved by performing the SVD on $\mathbf{U}_1^T\mathbf{U}_v$. We exploit membership matrices on anchors instead of all samples to explore the latent cluster structure of data, which avoids the decomposition or inverse operations of high-order matrices in the subsequent solution process. To capture the consensus cluster structure of anchors, the same fusion approach as Eq.(4) can be employed to merge the aligned membership matrices:

$$\min_{\mathbf{U1}=\mathbf{1},\mathbf{U}\geq 0,\boldsymbol{\alpha}\geq 0,\boldsymbol{\alpha}^T\mathbf{1}=1} \|\mathbf{U}-\sum_{v=1}^{V}\alpha_v\widetilde{\mathbf{U}}_v\|_F^2, \tag{7}$$

where $\widetilde{\mathbf{U}}_v$ (i.e., $\mathbf{U}_v\mathbf{T}_v$) denotes the aligned membership matrix, serving as the label guidance for feature selection. With the learned bipartite graph $\mathbf{S}$, the label guidance (i.e., cluster information) of samples can be directly inferred from anchors, as shown in Fig. 2. Therefore, the final optimization objective of the proposed EMUFS is obtained as:

$$\begin{aligned}&\min_{\|\mathbf{W}\|_{2,0}=k,\mathbf{F}\geq 0,\mathbf{S}\geq 0,\mathbf{U}\geq 0,\boldsymbol{\alpha}\geq 0} \|\mathbf{X}^T\mathbf{W}-\mathbf{F}\|_F^2+\lambda\|\mathbf{F}-\mathbf{SU}\|_F^2\\&+\beta(\|\mathbf{S}-\sum_{v=1}^{V}\alpha_v\mathbf{S}_v\|_F^2+\|\mathbf{U}-\sum_{v=1}^{V}\alpha_v\mathbf{U}_v\|_F^2)\\&+\gamma\sum_{i=1}^{n}\sum_{j=1}^{m}\|\mathbf{W}^T\mathbf{x}_i-\mathbf{W}^T\mathbf{z}_j\|_2^2 s_{ij}\\&\text{s.t. } \mathbf{F1}=\mathbf{1},\mathbf{S1}=\mathbf{1},\mathbf{U1}=\mathbf{1},\boldsymbol{\alpha}^T\mathbf{1}=1,\end{aligned} \tag{8}$$

where $\mathbf{F}$ denotes the cluster labels of samples and simultaneously leverages the similarity and cluster information. In Eq. (8), the membership matrix and the bipartite graph are collaboratively fused to facilitate the cluster information inference on the graph. Moreover, the similarity structure is also guided by the relations between samples and anchors in the selected feature subspace, enhancing the graph quality. Consequently, similarity structure learning, cluster information inference and feature selection are incorporated into a unified framework and can benefit from each other, such that the discriminative features can be selected under the guidance of the learned structure compatible across all views.

3 Optimization and Analyses

To solve the problem in Eq. (8), an iterative optimization is devised to solve one variable by fixing other variables. Considering that the $l_{2,0}$-norm optimization problem is complicated, an auxiliary variable $\mathbf{E}=\mathbf{W}$ is introduced to transform Eq. (8) into the following equivalent problem:

$$\begin{aligned}&\min_{\|\mathbf{E}\|_{2,0}=k,\mathbf{W},\mathbf{F}\geq 0,\mathbf{S}\geq 0,\mathbf{U}\geq 0,\boldsymbol{\alpha}\geq 0} \|\mathbf{X}^T\mathbf{W}-\mathbf{F}\|_F^2+\lambda\|\mathbf{F}-\mathbf{SU}\|_F^2\\&+\beta(\|\mathbf{S}-\sum_{v=1}^{V}\alpha_v\mathbf{S}_v\|_F^2+\|\mathbf{U}-\sum_{v=1}^{V}\alpha_v\mathbf{U}_v\|_F^2)\\&+\gamma\sum_{i=1}^{n}\sum_{j=1}^{m}\|\mathbf{W}^T\mathbf{x}_i-\mathbf{W}^T\mathbf{z}_j\|_2^2 s_{ij}+\frac{\mu}{2}\|\mathbf{E}-\mathbf{W}+\frac{\mathbf{\Pi}}{\mu}\|_F^2\\&\text{s.t. } \mathbf{F1}=\mathbf{1},\mathbf{S1}=\mathbf{1},\mathbf{U1}=\mathbf{1},\boldsymbol{\alpha}^T\mathbf{1}=1,\end{aligned} \tag{9}$$

where $\mu\in\mathbb{R}^{1\times 1}$ and $\mathbf{\Pi}\in\mathbb{R}^{n\times c}$ denote the penalty parameter and Lagrange multipliers, respectively. The solution procedures are given as follows:

• **U subproblem:** When fixing other variables, the problem of Eq. (9) is simplified into:

$$\min_{\mathbf{U1}=\mathbf{1},\mathbf{U}\geq 0} \lambda\|\mathbf{F}-\mathbf{SU}\|_F^2+\beta\|\mathbf{U}-\sum_{v=1}^{V}\alpha_v\widetilde{\mathbf{U}}_v\|_F^2. \tag{10}$$

Algorithm 1 Optimization procedures for EMUFS

Input: Multi-view data $\mathbf{X}=[\mathbf{X}_1,\cdots,\mathbf{X}_V]^T$, the cluster number c, and the parameters λ, β and γ;
1: Initialize the view-weights $\alpha_v=1/V$, a random feature selection matrix $\mathbf{W}$, $\mathbf{E}=\mathbf{\Pi}=\mathbf{0}$, $\mu=1$ and $\rho=1.1$;
2: Generate the anchor set $\{\mathbf{Z}_v\}_{v=1}^V$, and initialize the membership matrices $\{\mathbf{U}_v\}_{v=1}^V$ by FCM;
3: Learn the permutation matrix $\mathbf{T}_v$ by solving Eq. (6);
4: **repeat**
5: Update $\mathbf{U}$ by Eq. (10);
6: Update $\boldsymbol{\alpha}$ by Eq. (13);
7: Update $\mathbf{W}$ by Eq. (15);
8: Update $\mathbf{E}$ by Eq. (16);
9: Update $\mathbf{S}$ by Eq. (19);
10: Update $\mathbf{F}$ by Eq. (19);
11: Update $\mathbf{\Pi}$ and μ by Eq. (20);
12: **until** Eq. (9) converges;
Output: Feature selection matrix $\mathbf{W}$ with k nonzero rows.

For computational efficiency, we first ignore the constraints to calculate the latent solution $\mathbf{U}^*$ and then project $\mathbf{U}^*$ into the constrained space. Specifically, setting the derivative of Eq. (10) w.r.t. $\mathbf{U}$ to zero, the latent solution of $\mathbf{U}$ is obtained as: $\mathbf{U}^*=(\lambda\mathbf{S}^T\mathbf{S}+\beta\mathbf{I})^{-1}(\lambda\mathbf{S}^T\mathbf{F}+\beta\sum_{v=1}^{V}\alpha_v\widetilde{\mathbf{U}}_v)$. Subsequently, the optimal solution of $\mathbf{U}$ can be derived by solving:

$$\min_{\mathbf{U1}=\mathbf{1},\mathbf{U}\geq 0}\|\mathbf{U}-\mathbf{U}^*\|_F^2. \tag{11}$$

Eq. (11) can be solved with a closed-form solution [Huang *et al.*, 2015]. Unlike previous methods that directly learn the cluster information of all samples and need the $\mathcal{O}(n^2)$ computational complexity at least, EMUFS explores the cluster structure of anchors first and then uses the learned bipartite graph to propagate the cluster information, avoiding the decomposition or inverse operations of n-order dense matrices.

• **$\boldsymbol{\alpha}$ subproblem:** By fixing the other variables, the optimization subproblem w.r.t. $\boldsymbol{\alpha}$ is:

$$\min_{\boldsymbol{\alpha}^T\mathbf{1}=1,\boldsymbol{\alpha}\geq 0}\|\mathbf{S}-\sum_{v=1}^{V}\alpha_v\mathbf{S}_v\|_F^2+\|\mathbf{U}-\sum_{v=1}^{V}\alpha_v\widetilde{\mathbf{U}}_v\|_F^2. \tag{12}$$

Setting $\boldsymbol{\psi}$ = vec($\mathbf{S}$) $\in\mathbb{R}^{nm\times 1}$, $\boldsymbol{\omega}$ = vec($\mathbf{U}$) $\in\mathbb{R}^{mc\times 1}$, $\mathbf{\Psi}$ = [vec($\mathbf{S}_1$),$\cdots$,vec($\mathbf{S}_V$)] $\in\mathbb{R}^{nm\times V}$ and $\mathbf{\Omega}$ = [vec($\widetilde{\mathbf{U}}_1$),$\cdots$,vec($\widetilde{\mathbf{U}}_V$)] $\in\mathbb{R}^{mc\times V}$), thus Eq. (12) becomes:

$$\min_{\boldsymbol{\alpha}\geq 0,\boldsymbol{\alpha}^T\mathbf{1}=1}\boldsymbol{\alpha}^T\mathbf{Q}\boldsymbol{\alpha}-2\boldsymbol{\alpha}^T\mathbf{p}, \tag{13}$$

where $\mathbf{Q}=\mathbf{\Psi}^T\mathbf{\Psi}+\mathbf{\Omega}^T\mathbf{\Omega}$, $\mathbf{p}=\mathbf{\Psi}^T\boldsymbol{\psi}+\mathbf{\Omega}^T\boldsymbol{\omega}$. Since $\mathbf{Q}$ is semi-definite, Eq. (13) is a quadratic convex problem and can be solved efficiently [Jiang *et al.*, 2023].

• **W subproblem:** By fixing other variables except for $\mathbf{W}$, Eq. (9) is simplified into:

$$\begin{aligned}&\min_{\mathbf{W}}\|\mathbf{X}^T\mathbf{W}-\mathbf{F}\|_F^2+\gamma\sum_{i=1}^{n}\sum_{j=1}^{m}\|\mathbf{W}\mathbf{x}_i-\mathbf{W}\mathbf{z}_j\|_2^2 s_{ij}\\&+\frac{\mu}{2}\|\mathbf{E}-\mathbf{W}+\frac{\mathbf{\Pi}}{\mu}\|_F^2.\end{aligned} \tag{14}$$

Setting the derivative of Eq. (14) w.r.t. $\mathbf{W}$ to zero, we have:

$$\mathbf{W} = (\mathbf{X}\mathbf{X}^T + \gamma\mathbf{M} + \frac{\mu}{2}\mathbf{I})^{-1}(\mathbf{X}\mathbf{F} + \frac{\mu}{2}\mathbf{E} + \frac{\mathbf{\Pi}}{2}), \quad (15)$$

where $\mathbf{M} = \mathbf{X}\mathbf{X}^T + \mathbf{Z}\mathbf{\Lambda}\mathbf{Z}^T - \mathbf{X}\mathbf{S}\mathbf{Z}^T - \mathbf{Z}\mathbf{S}^T\mathbf{X}^T$.

• **E subproblem:** By fixing the other variables except for $\mathbf{E}$, we have the following subproblem:

$$\min_{\|\mathbf{E}\|_{2,0}=k} \|\mathbf{E} - \mathbf{W} + \frac{\mathbf{\Pi}}{\mu}\|_F^2. \quad (16)$$

$\mathbf{E}$ can be directly solved by setting the $d-k$ rows of $\mathbf{W} - \frac{\mathbf{\Pi}}{\mu}$ with the smallest l_2-norm to zeros.

• **S subproblem:** By fixing the other variables except for $\mathbf{S}$, we have the following subproblem:

$$\min_{\mathbf{S}} \lambda\|\mathbf{F} - \mathbf{S}\mathbf{U}\|_F^2 + \gamma \sum_{i=1}^{n}\sum_{j=1}^{m} \|\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{z}_j\|_2^2 s_{ij}$$
$$+ \beta\|\mathbf{S} - \sum_{v=1}^{V}\alpha_v\mathbf{S}_v\|_F^2 \quad \text{s.t.} \ \mathbf{S}\mathbf{1} = \mathbf{1}, \mathbf{S} \geq 0. \quad (17)$$

Noting that the optimization problem in Eq. (17) is independent for each row (i.e., $\mathbf{s}_i$), we can calculate $\mathbf{S}$ by rows:

$$\min_{\mathbf{s}_i\mathbf{1}=1, \mathbf{s}_i \geq 0} \lambda\|\mathbf{f}_i - \mathbf{s}_i\mathbf{U}\|_2^2 + \beta\|\mathbf{s}_i - \mathbf{a}_i\|_2^2 + \gamma\mathbf{s}_i\mathbf{d}_i^T \quad (18)$$

where $\mathbf{f}_i$ and $\mathbf{a}_i$ denote the i-th row of $\mathbf{F}$ and $\mathbf{A} = \sum_{v=1}^{V}\alpha_v\mathbf{S}_v$, and $\mathbf{d}_i$ is a row vector with $d_{ij} = \|\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{z}_j\|_2^2$. Setting the derivative of Eq. (18) w.r.t. $\mathbf{s}_i$ to zero, the latent solution of $\mathbf{s}_i$ is obtained as: $\mathbf{s}_i^* = (\lambda\mathbf{f}_i\mathbf{U}^T + \beta\mathbf{a}_i - \frac{\gamma}{2}\mathbf{d}_i)(\lambda\mathbf{U}\mathbf{U}^T + \beta\mathbf{I})^{-1}$. Thus, the optimal solution of $\mathbf{s}_i$ can be obtained by solving:

$$\min_{\mathbf{s}_i\mathbf{1}=1, \mathbf{s}_i \geq 0} \|\mathbf{s}_i - \mathbf{s}_i^*\|_2^2, \quad (19)$$

which can be effectively solved by [Huang *et al.*, 2015].

• **F subproblem:** Fixing other variables, we can optimize $\mathbf{F}$ by rows:

$$\min_{\mathbf{f}_i\mathbf{1}=1, \mathbf{f}_i \geq 0} \|\mathbf{x}_i^T\mathbf{W} - \mathbf{f}_i\|_2^2 + \lambda\|\mathbf{f}_i - \mathbf{s}_i\mathbf{U}\|_F^2. \quad (20)$$

Eq. (20) can be solved with a closed-form solution.

• **Update ALM parameters:** In each iteration, the penalty parameter μ and the Lagrange multipliers $\mathbf{\Pi}$ are updated as:

$$\mathbf{\Pi} = \mathbf{\Pi} + \mu(\mathbf{E} - \mathbf{W})$$
$$\mu = \rho\mu. \quad (21)$$

where ρ is a constant updated rate. During the optimization, EMUFS separately optimizes each subproblem, decreasing the objective value monotonically until convergence.

Computational Complexity

The steps for solving the problem (9) are summarized in Algorithm 1. Specifically, selecting m anchors takes $\mathcal{O}(nm)$, calculating $\{\mathbf{U}_v\}_{v=1}^V$ by FCM requires $\mathcal{O}(mdcV)$, and learning the permutation matrices needs $\mathcal{O}(mc^2V)$. Afterward, updating $\mathbf{U}$, $\mathbf{W}$ and $\mathbf{S}$ involve the inverses of matrices, requiring $\mathcal{O}(nmd+m^2c+m^3)$, $\mathcal{O}(nd^2+nmd+d^3)$ and $\mathcal{O}(nm^2+ndc+m^3)$. Besides, calculating $\boldsymbol{\alpha}$ takes $\mathcal{O}(nmV^2)$, and the optimization of $\mathbf{E}$ and $\mathbf{F}$ respectively take $\mathcal{O}(dk+dc)$ and $\mathcal{O}(nc)$. Since c and V are small constants, the computational complexity of EMUFS in each iteration is $\mathcal{O}(nm^2+nd^2+m^3+d^3)$, which is linearly related to the sample scale.

Dataset	Classes	Data size	Feature size
flower-17	17	1360	3011(512/630/1239/630)
Leaves	100	1600	192(64/64/64)
NUS	12	2400	1134(64/144/73/128/225/ 500)
Scene	8	2688	1248(512/432/256/48)
ALOI	100	10800	279(77/13/64/125)
Youtube	31	24800	832(113/322/110/287)

Table 2: Detailed information on multi-view datasets.

Methods	Total computational complexity
SOGFS	$\mathcal{O}(n^2d + d^2c)$
MVCSS	$\mathcal{O}(nd + d^3)$
ASVW	$\mathcal{O}(n^2d + d^3)$
NSGL	$\mathcal{O}(n^2d + d^3)$
MAMFS	$\mathcal{O}(nd^2 + n^2d + d^3)$
MGAGR	$\mathcal{O}(n^2d + d^3)$
EMUFS	$\mathcal{O}(nm^2 + nd^2 + m^3 + d^3)$

Table 3: The computational complexity of multi-view methods.

4 Experiments

4.1 Experimental Settings

In this section, six real-word datasets are employed, including flower-17[1], Leaves[2], NUS[3], Scene[4], ALOI[5] and Youtube[6]. The details of each dataset are listed in Table 2. To comprehensively verify the superiority and effectiveness of EMUFS, we conduct experiments with six state-of-the-art competitors, including (1) Unsupervised Feature Selection with Structured Graph Optimization (**SOGFS**) [Nie *et al.*, 2016]; (2) Multi-View Clustering and Feature Learning via Structured Sparsity (**MVCSS**) [Wang *et al.*, 2013]; (3) Multi-view Unsupervised Feature Selection with Adaptive Similarity and View Weight (**ASVW**) [Hou *et al.*, 2017]; (4) Multi-view Feature Selection via Nonnegative Structured Graph Learning (**NSGL**) [Bai *et al.*, 2020]; (5) Multilevel Projections with Adaptive Neighbor Graph for Unsupervised Multi-View Feature Selection (**MAMFS**) [Zhang *et al.*, 2021]; (6) Robust Unsupervised Feature Selection via Multi-Group Adaptive Graph Representation (**MGAGR**) [You *et al.*, 2023]. To ensure comparison fairness, the parameters of all competitors are tuned following their respective works. The regularization parameters for EMUFS are searched in a grid of $\{10^{-3}, 10^{-2}, \cdots, 10^3\}$, with the number of anchors set as $m = 10\% \times n$. The K-means clustering is independently executed 20 times on the selected feature subsets, and the average results, including the clustering accuracy (ACC) and the normalized mutual information (NMI), are reported to evaluate the performance.

4.2 Comparison Results

The means and standard deviations of ACC and NMI are presented in Tables 4 and 5, where the optimal and second-best results are prominently indicated in bold and underlined formatting, respectively. The following conclusions can

[1]https://www.robots.ox.ac.uk/ vgg/data/flowers/

[2]https://archive.ics.uci.edu/dataset/

[3]https://lms.comp.nus.edu.sg/wp-content/uploads/2019/research/nuswide/NUS-WIDE.html

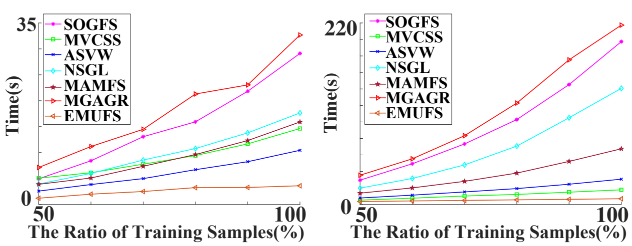
[4]http://people.csail.mit.edu/torralba/code/spatialenvelope/

[5]https://aloi.science.uva.nl/

[6]https://archive.ics.uci.edu/dataset/269/

Datasets	Feature ratio	10%	15%	20%	25%	30%	35%
flower-17	SOGFS	18.12±0.59	18.35±0.78	18.08±0.35	18.49±0.59	19.76±0.70	19.22±0.56
	MVCSS	23.20±0.98	23.34±1.21	23.51±1.30	24.50±1.17	25.33±1.55	24.95±0.83
	ASVW	22.21±1.04	22.49±0.91	22.28±0.71	22.37±0.81	21.40±0.90	21.68±0.69
	NSGL	26.62±0.79	25.81±2.04	26.40±1.56	26.54±1.04	26.76±1.53	27.65±1.21
	MAMFS	22.18±1.07	24.70±0.95	25.32±1.09	26.36±1.25	26.59±1.20	27.98±1.22
	MGAGR	25.82±1.33	25.29±1.15	25.09±0.89	25.36±1.35	26.51±0.79	26.47±1.09
	EMUFS	**30.49**±1.30	**31.51**±1.87	**30.82**±1.59	**30.05**±1.15	**31.20**±1.14	**29.15**±0.99
Leaves	SOGFS	43.91±1.57	50.64±1.29	52.84±1.88	55.29±1.33	60.39±1.58	60.51±1.98
	MVCSS	36.51±0.84	42.84±0.75	49.13±1.54	53.09±0.85	54.12±1.34	57.98±1.72
	ASVW	46.88±0.94	53.11±1.25	60.11±1.37	62.70±1.59	65.98±1.45	68.32±1.51
	NSGL	35.41±0.86	44.43±1.26	48.07±1.45	52.22±1.60	54.41±1.54	58.22±1.78
	MAMFS	54.53±1.25	60.05±1.63	63.57±1.58	64.59±1.27	66.45±1.14	67.83±2.19
	MGAGR	53.16±1.61	59.51±1.49	61.79±1.72	63.58±2.17	66.96±1.65	68.71±1.45
	EMUFS	**56.31**±1.08	**62.92**±1.34	**67.06**±1.76	**69.95**±1.47	**71.45**±1.64	**72.34**±1.50
NUS	SOGFS	19.08±0.24	19.22±0.56	20.53±0.27	20.63±0.24	19.35±0.25	20.32±0.43
	MVCSS	23.93±0.56	24.62±0.80	25.44±0.72	25.43±0.98	25.98±0.84	26.21±0.58
	ASVW	23.08±0.63	24.21±0.79	23.75±0.85	23.62±0.41	23.33±0.46	23.94±0.59
	NSGL	23.18±1.09	24.58±1.47	25.51±0.76	25.58±0.66	26.00±0.87	25.76±0.90
	MAMFS	23.65±0.56	22.60±0.74	23.18±0.86	23.69±0.22	23.62±0.42	23.36±0.67
	MGAGR	22.35±0.35	24.63±1.05	24.81±0.88	25.04±0.73	24.81±0.61	25.79±0.51
	EMUFS	**25.70**±0.78	**26.50**±0.92	**27.31**±0.94	**27.29**±0.69	**27.68**±1.18	**27.26**±1.02
Scene	SOGFS	36.46±0.12	37.72±0.22	37.61±0.31	38.41±0.10	39.45±0.16	39.75±0.15
	MVCSS	42.66±0.35	41.78±0.19	42.85±0.42	44.36±0.30	44.58±0.60	44.82±0.43
	ASVW	42.63±0.24	41.51±0.22	42.24±0.50	42.56±0.21	43.38±1.52	44.26±0.10
	NSGL	40.33±0.47	41.78±0.46	40.76±0.16	40.36±0.18	40.31±0.36	40.57±0.37
	MAMFS	41.78±0.74	41.74±0.84	44.25±0.33	43.77±0.27	44.19±0.30	45.63±0.22
	MGAGR	42.93±0.60	44.59±0.12	44.13±0.85	44.27±0.18	44.06±0.09	45.51±0.95
	EMUFS	**44.08**±0.15	**45.18**±0.21	**45.53**±0.17	**46.09**±0.23	**46.30**±0.24	**46.53**±0.35
ALOI	SOGFS	51.34±1.05	53.07±1.17	55.30±1.37	57.53±0.95	57.78±1.07	60.95±1.17
	MVCSS	56.52±0.64	59.75±1.39	**61.85**±0.90	64.13±1.36	64.02±1.34	64.11±1.67
	ASVW	58.94±1.19	59.71±1.13	61.21±1.90	63.29±1.18	63.56±1.43	**64.42**±1.75
	NSGL	29.35±0.73	45.87±1.23	48.26±1.36	48.74±1.46	50.31±1.26	55.44±1.67
	MAMFS	47.59±0.59	50.28±0.88	54.36±2.16	57.44±1.43	57.50±1.51	61.60±1.78
	MGAGR	59.45±1.09	61.59±0.82	61.08±1.07	61.48±0.69	62.97±1.36	62.27±1.63
	EMUFS	**61.53**±1.40	**62.29**±1.32	61.56±1.27	**64.66**±1.15	**64.06**±1.24	64.27±1.26
Youtube	SOGFS	16.79±0.65	19.57±1.26	21.73±1.78	23.63±1.74	24.75±2.38	27.50±3.20
	MVCSS	20.21±0.80	24.86±1.08	25.20±1.52	27.97±2.20	28.24±2.21	32.54±2.71
	ASVW	16.68±1.66	17.03±1.47	17.73±1.48	25.69±2.83	26.47±2.34	26.94±2.44
	NSGL	21.91±0.74	25.97±1.80	26.73±1.60	**28.41**±1.33	**28.58**±1.49	31.56±2.28
	MAMFS	20.59±0.86	**27.01**±1.34	26.47±1.11	27.66±1.30	28.43±2.16	31.08±1.81
	MGAGR	21.12±0.92	21.01±1.06	22.62±1.41	22.16±1.30	23.88±2.23	25.22±1.67
	EMUFS	**22.78**±1.04	26.59±1.80	**26.92**±0.91	28.04±1.84	28.21±1.48	**33.22**±2.75

Table 4: ACC of different methods with different numbers of features. The best and second results are in bold and underlined.



(a) ALOI (b) Youtube

Figure 3: Running time versus the number of samples.

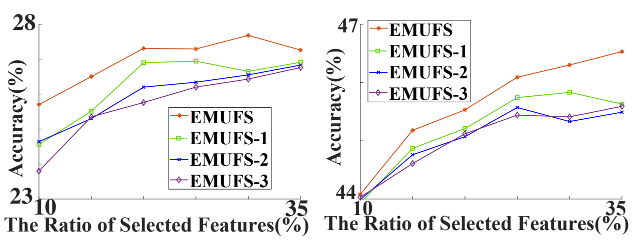
be drawn: (1) Across varying numbers of selected features, EMUFS consistently exhibits competitive or superior results, fully demonstrating its effectiveness in multi-view feature selection. (2) The ACC and NMI scores achieved through the EMUFS outperform those of the single-view methodology. This underscores the assertion that harnessing distinct information from multiple views enhances the selection of informative features. (3) Compared with other multi-view feature selection methods, EMUFS consistently attains superior performance, highlighting the effectiveness of collaborative utilization of the cluster structure and the similarity structure. Meanwhile, to assess the efficiency of EMUFS, Table 3 summarizes the computation complexity of each method and Fig. 3 shows the running times versus the training sample scale on the ALOI and Youtube datasets. We find that the running times of EMUFS exhibit a linear increase, while other methods emerge an exponentially increasing trend with the increased sample scale, fully validating the efficiency of propagating cluster information on the bipartite graph.

4.3 Ablation Study

To investigate the significance of adaptive structure learning and inference, an ablation study is conducted to design three variants of EMUFS: EMUFS_1, wherein the similarity structure is only derived from original data, overlooking the information from projection space; EMUFS_2, a variation of EMUFS that excludes the collaborative fusion model, guiding feature selection by fixed graphs and membership matrices; and EMUFS_3 employs the unaligned membership matrices to learn the cluster structure. The results of EMUFS and its distilled versions are depicted in Fig. 4. We can conclude that: (1) The ACC of EMUFS_1 is inferior to EMUFS, indicating that the dynamic graph learning on selected feature space has an effective influence on the overall performance; (2) EMUFS

Datasets	Feature ratio	10%	15%	20%	25%	30%	35%
flower-17	SOGFS	14.92±0.63	15.41±0.81	15.69±0.53	16.11±0.62	17.51±0.97	17.48±0.51
	MVCSS	21.48±0.96	21.98±1.11	22.95±0.59	23.28±0.70	24.56±0.96	24.33±1.01
	ASVW	20.83±0.78	21.51±0.72	21.29±0.84	21.16±0.93	20.43±0.67	21.45±0.45
	NSGL	24.50±0.80	25.57±1.57	25.21±1.14	27.12±0.77	25.42±1.51	25.66±1.19
	MAMFS	19.88±0.50	21.88±0.97	23.56±0.74	24.48±0.82	25.63±1.01	26.81±0.70
	MGAGR	24.07±0.84	23.82±0.90	24.29±0.73	25.02±1.15	25.97±0.61	26.26±0.89
	EMUFS	**29.96**±0.86	**30.94**±1.09	**30.51**±1.36	**30.00**±0.96	**30.81**±0.74	**28.49**±0.85
Leaves	SOGFS	70.99±0.49	75.51±0.57	77.70±0.61	78.50±0.51	81.89±0.50	82.12±0.64
	MVCSS	65.57±0.52	69.22±0.35	74.21±0.63	76.71±0.34	77.57±0.73	80.08±0.51
	ASVW	72.20±0.45	77.60±0.42	80.92±0.49	82.31±0.62	84.24±0.47	86.14±0.49
	NSGL	65.72±0.39	71.96±0.55	74.64±0.48	77.36±0.55	79.13±0.52	81.26±0.52
	MAMFS	76.81±0.41	80.84±0.49	82.56±0.50	83.55±0.48	84.16±0.38	85.85±0.89
	MGAGR	74.79±0.35	80.23±0.67	81.83±0.59	83.14±0.80	85.10±0.54	86.62±0.56
	EMUFS	**78.18**±1.41	**82.33**±0.39	**84.96**±0.37	**86.74**±0.53	**88.06**±0.50	**88.96**±0.47
NUS	SOGFS	8.59±0.31	8.95±0.40	11.74±0.25	11.91±0.14	12.28±0.19	13.04±0.24
	MVCSS	12.03±0.11	14.15±0.31	13.58±0.23	**15.56**±0.54	15.55±0.56	15.27±0.33
	ASVW	12.27±0.25	12.86±0.29	12.80±0.31	13.02±0.19	12.99±0.24	13.32±0.26
	NSGL	12.74±0.44	13.63±0.56	14.33±0.76	14.48±0.45	14.59±0.53	14.64±0.40
	MAMFS	11.97±0.20	12.19±0.44	11.83±0.76	12.16±0.37	12.45±0.36	12.27±0.74
	MGAGR	11.31±0.16	12.97±0.46	12.71±0.66	13.27±0.16	13.21±0.36	15.13±0.24
	EMUFS	**13.82**±0.32	**14.77**±0.54	**14.94**±0.51	15.21±0.30	**16.13**±0.57	**15.41**±0.33
SCENE	SOGFS	26.63±0.14	27.92±0.13	28.18±0.21	28.24±0.26	28.33±0.10	30.49±0.21
	MVCSS	32.20±0.15	33.53±0.26	33.49±0.23	**33.96**±0.30	35.65±0.35	36.24±0.27
	ASVW	30.99±0.18	29.77±0.28	31.66±0.26	32.26±0.16	33.09±0.59	34.37±1.11
	NSGL	27.56±0.12	30.54±0.18	31.00±0.11	31.59±0.14	33.09±0.11	34.18±0.21
	MAMFS	27.75±0.42	28.08±0.10	30.36±0.13	31.13±0.11	31.40±0.68	33.28±0.43
	MGAGR	31.49±0.65	32.56±0.02	32.68±0.12	33.46±0.34	32.67±0.38	35.84±0.57
	EMUFS	**32.78**±0.18	**33.80**±0.17	**34.16**±0.11	33.53±0.14	**36.13**±0.17	**36.46**±0.18
ALOI	SOGFS	71.37±0.28	72.27±0.27	73.76±0.45	75.02±0.19	75.56±0.26	77.39±0.30
	MVCSS	73.75±0.48	75.81±0.41	77.72±0.33	79.53±0.63	79.57±0.90	80.70±0.75
	ASVW	77.56±0.23	77.42±0.38	77.92±0.52	79.38±0.40	80.13±0.35	**81.00**±0.54
	NSGL	53.33±0.40	67.72±0.49	69.33±0.37	70.19±0.73	72.12±0.45	75.15±0.70
	MAMFS	66.97±0.25	69.25±0.16	76.43±0.53	77.97±0.60	78.01±0.56	77.95±0.41
	MGAGR	76.54±0.22	77.62±0.27	77.74±0.44	78.50±0.23	78.64±0.27	78.90±0.54
	EMUFS	**79.71**±0.35	**79.76**±0.52	**78.25**±0.62	**80.57**±0.56	**82.23**±0.46	80.76±0.44
Youtube	SOGFS	17.87±1.25	21.26±1.37	25.91±2.23	27.72±2.15	30.82±2.76	34.46±3.20
	MVCSS	21.87±0.71	28.67±1.15	29.34±0.84	33.71±2.10	**34.29**±2.62	40.31±1.79
	ASVW	16.68±1.44	17.47±1.27	20.04±1.32	29.74±1.71	33.69±2.56	33.70±2.20
	NSGL	23.12±0.59	27.69±1.29	28.38±1.17	30.74±1.20	32.40±1.30	34.37±1.97
	MAMFS	21.58±1.04	**32.71**±1.26	30.64±1.48	33.25±1.55	33.66±2.17	38.15±2.57
	MGAGR	22.43±0.82	23.22±0.71	25.14±1.34	25.42±1.34	27.59±2.32	29.79±2.05
	EMUFS	**23.40**±1.10	29.29±1.02	**31.02**±1.12	**33.85**±1.22	32.11±1.46	**42.17**±2.65

Table 5: NMI of different methods with different numbers of features. The best and second results are in bold and underlined.



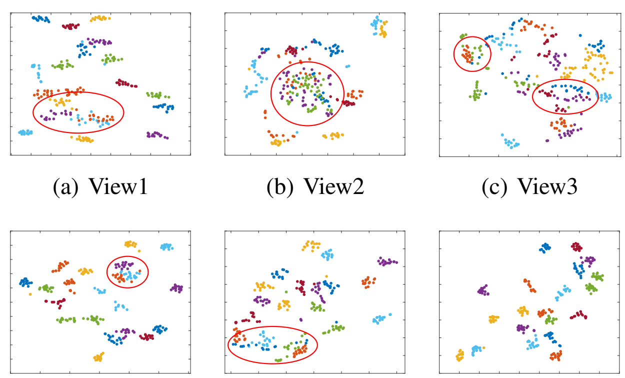
(a) NUS (b) Scene

Figure 4: ACC of EMUFS and its simplified versions.

achieves superior results than $EMUFS_2$, validating that the collaborative fusion of similarity structure and cluster information can facilitate feature selection; (3) Comparative analyses between EMUFS and $EMUFS_3$ demonstrate that simply concatenating different membership matrices impairs the consistency of clustering centers among views.

4.4 Visualization

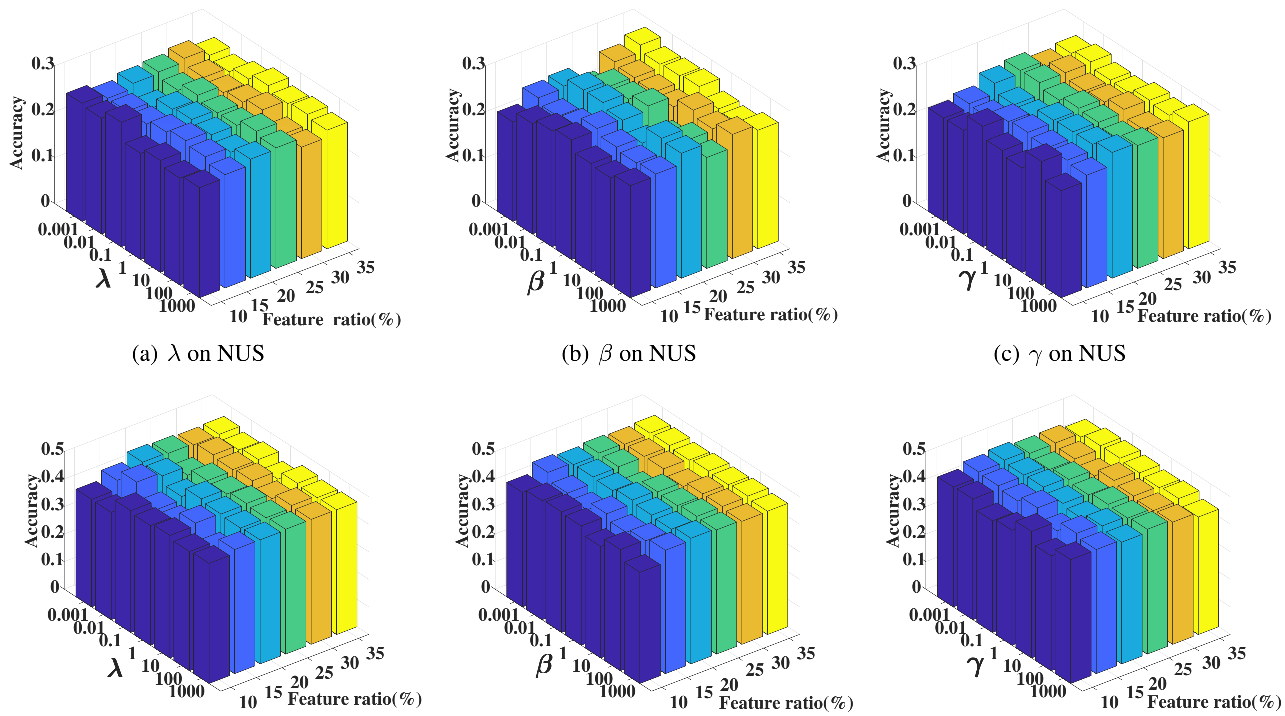
To visually evaluate the quality of selected features, the T-SNE is employed to project the selected features into a two-dimensional space [Van der Maaten and Hinton, 2008]. In this context, we select 320 samples from 20 clusters of the Leaves dataset for visualization, in which each sample has three views (i.e. View #1, #2 and #3), and each view contains 64 features. Figs. 5 (a)-(c) show the results of the original three views, while Figs. 5 (d)-(f) depict the visualizations



(a) View1 (b) View2 (c) View3

(d) ASVW (e) MAMFS (f) EMUFS

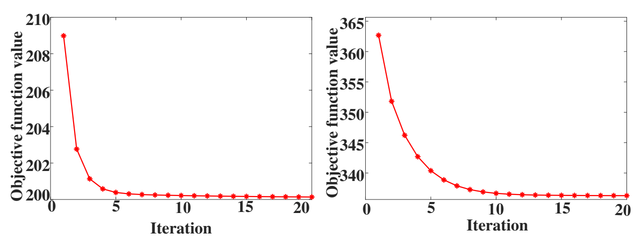
Figure 5: The T-SNE visualizations on the Leaves dataset.



(a) λ on NUS (b) β on NUS (c) γ on NUS

(d) λ on SCENE (e) β on SCENE (f) γ on SCENE

Figure 6: ACC with different parameters on the NUS and SCENE datasets.



(a) NUS (b) Scene

Figure 7: Variation curves of objective function values.

of feature subsets selected by ASVW, MAMFS, and EMUFS, respectively. To ensure a fair evaluation, the number of selected features is set to 64 for each method. As depicted in Fig. 5, the visualizations of the original views exhibit significant overlaps between different clusters. In contrast, ASVW, MAMFS, and EMUFS achieve separable results, highlighting the significance of multi-view feature selection. Moreover, the inter-cluster distance obtained by EMUFS is larger than those of ASVW and MAMFS, demonstrating that EMUFS can select a compact subset of discriminative features.

4.5 Parameter Sensitivity and Convergence

EMUFS encompasses three intrinsic parameters, namely λ, β and γ. To assess the impacts of these parameters, we present the clustering accuracies under distinct parameter settings, alongside varying numbers of selected features, in Fig. 6. This depiction shows that EMUFS can perform better, particularly when λ assumes a value smaller than 1. This observation effectively indicates that the adaptive structure learning and fusion indeed contribute to identifying informative features. Meanwhile, to illustrate the convergence behavior of EMUFS, Fig. 7 provides the variations of the objective function values, which shows that the objective function exhibits a rapid decrease and converges to a stable value quickly.

5 Conclusion

In this paper, we propose an efficient multi-view unsupervised feature selection with adaptive structure learning and inference (EMUFS). Unlike existing methods that exploit the unreliable similarity structures as well as overlook the cluster structures of data, EMUFS exploits the similarity structure from the selected feature space to alleviate the impacts of low-quality features and explores the cluster structure by membership matrices, so that the feature selection process can be effectively guided by the collaborative cooperation of structures. Moreover, EMUFS learns the cluster structure of samples from anchors via the information propagation on the bipartite graph, reducing the computational cost and strengthening the scalability for large-scale tasks. Extensive experiments demonstrate the effectiveness of EMUFS and its superiority against the state-of-the-art competitors.

Acknowledgments

This work was supported by the "Pioneer" and "Leading Goose" R&D Program of Zhejiang Province, China (Grant No. 2023C01022), the National Natural Science Foundation of China (Nos. 62306171 and 62176001), the Science and Technology Major Project of Shanxi (No. 202201020101006), the Natural Science Foundation of Chongqing (No. CSTB2022NSCQ-MSX1202).

References

[Bai *et al.*, 2020] Xiangpin Bai, Lei Zhu, Cheng Liang, Jingjing Li, Xiushan Nie, and Xiaojun Chang. Multi-view feature selection via nonnegative structured graph learning. *Neurocomputing*, 387:110–122, 2020.

[Bai *et al.*, 2021] Zechen Bai, Zhigang Wang, Jian Wang, Di Hu, and Errui Ding. Unsupervised multi-source domain adaptation for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12914–12923, 2021.

[Bezdek *et al.*, 1984] James C Bezdek, Robert Ehrlich, and William Full. Fcm: The fuzzy c-means clustering algorithm. *Computers & geosciences*, 10(2-3):191–203, 1984.

[Chen *et al.*, 2023] Hong Chen, Feiping Nie, Rong Wang, and Xuelong Li. Fast unsupervised feature selection with bipartite graph and $l_{2,0}$-norm constraint. *IEEE Transactions on Knowledge and Data Engineering*, 35(5):4781–4793, 2023.

[Dong *et al.*, 2018] Xiao Dong, Lei Zhu, Xuemeng Song, Jingjing Li, and Zhiyong Cheng. Adaptive collaborative similarity learning for unsupervised multi-view feature selection. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 2064–2070, 2018.

[Feng *et al.*, 2013] Yinfu Feng, Jun Xiao, Yueting Zhuang, and Xiaoming Liu. Adaptive unsupervised multi-view feature selection for visual concept recognition. In *Proceedings of the Asian Conference on Computer Vision*, pages 343–357. Springer, 2013.

[Hou *et al.*, 2017] Chenping Hou, Feiping Nie, Hong Tao, and Dongyun Yi. Multi-view unsupervised feature selection with adaptive similarity and view weight. *IEEE Transactions on Knowledge and Data Engineering*, 29(9):1998–2011, 2017.

[Hu *et al.*, 2019] Di Hu, Feiping Nie, and Xuelong Li. Deep multimodal clustering for unsupervised audiovisual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9248–9257, 2019.

[Hu *et al.*, 2021] Rongyao Hu, Ziwen Peng, Xiaofeng Zhu, Jiangzhang Gan, Yonghua Zhu, Junbo Ma, and Guorong Wu. Multi-band brain network analysis for functional neuroimaging biomarker identification. *IEEE Transactions on Medical Imaging*, 40(12):3843–3855, 2021.

[Hu *et al.*, 2022] Rongyao Hu, Liang Peng, Jiangzhang Gan, Xiaoshuang Shi, and Xiaofeng Zhu. Complementary graph representation learning for functional neuroimaging identification. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 3385–3393, 2022.

[Huang *et al.*, 2015] Jin Huang, Feiping Nie, and Heng Huang. A new simplex sparse learning model to measure data similarity for clustering. In *Proceedings of the International Joint Jonference on Artificial Intelligence*, 2015.

[Jiang *et al.*, 2022] Bingbing Jiang, Junhao Xiang, Xingyu Wu, Yadi Wang, Huanhuan Chen, Weiwei Cao, and Weiguo Sheng. Robust multi-view learning via adaptive regression. *Information Sciences*, 610:916–937, 2022.

[Jiang *et al.*, 2023] Bingbing Jiang, Chenglong Zhang, Yan Zhong, Yi Liu, Yingwei Zhang, Xingyu Wu, and Weiguo Sheng. Adaptive collaborative fusion for multi-view semi-supervised classification. *Information Fusion*, 96:37–50, 2023.

[Li and Tang, 2015] Zechao Li and Jinhui Tang. Unsupervised feature selection via nonnegative spectral analysis and redundancy control. *IEEE Transactions on Image Processing*, 24(12):5343–5355, 2015.

[Li *et al.*, 2022] Xuelong Li, Han Zhang, Rong Wang, and Feiping Nie. Multi-view clustering: A scalable and parameter-free bipartite graph fusion method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):330–344, 2022.

[Li *et al.*, 2024a] Jing Li, Quanxue Gao, Qianqian Wang, and Wei Xia. Tensorized label learning on anchor graph. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 13537–13544, 2024.

[Li *et al.*, 2024b] Tao Li, Yuhua Qian, Feijiang Li, Xinyan Liang, and Zhi-hui Zhan. Feature subspace learning-based binary differential evolution algorithm for unsupervised feature selection. *IEEE Transactions on Big Data*, 2024.

[Liang *et al.*, 2022a] Weixuan Liang, Xinwang Liu, Sihang Zhou, Jiyuan Liu, Siwei Wang, and En Zhu. Robust graph-based multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 7462–7469, 2022.

[Liang *et al.*, 2022b] Xinyan Liang, Yuhua Qian, Qian Guo, Honghong Cheng, and Jiye Liang. Af: An association-based fusion method for multi-modal classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):9236–9254, 2022.

[Liang *et al.*, 2024] Xinyan Liang, Pinhan Fu, Qian Guo, Keyin Zheng, and Yuhua Qian. Dc-nas: Divide-and-conquer neural architecture search for multi-modal classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 13754–13762, 2024.

[Liu *et al.*, 2022] Wei Liu, Xiaodong Yue, Yufei Chen, and Thierry Denoeux. Trusted multi-view deep learning with opinion aggregation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 7585–7593, 2022.

[Lu *et al.*, 2023] Han Lu, Quanxue Gao, Qianqian Wang, Ming Yang, and Wei Xia. Centerless multi-view k-means based on the adjacency matrix. In *Proceedings of the AAAI*

Conference on Artificial Intelligence, pages 8949–8956, 2023.

[Nie *et al.*, 2014] Feiping Nie, Xiaoqian Wang, and Heng Huang. Clustering and projected clustering with adaptive neighbors. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 977–986, 2014.

[Nie *et al.*, 2016] Feiping Nie, Wei Zhu, and Xuelong Li. Unsupervised feature selection with structured graph optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1302–1308, 2016.

[Peng *et al.*, 2023] Liang Peng, Xin Wang, and Xiaofeng Zhu. Unsupervised multiplex graph learning with complementary and consistent information. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 454–462, 2023.

[Shi *et al.*, 2023] Dan Shi, Lei Zhu, Jingjing Li, Zheng Zhang, and Xiaojun Chang. Unsupervised adaptive feature selection with binary hashing. *IEEE Transactions on Image Processing*, 32:838–853, 2023.

[Tang *et al.*, 2019] Chang Tang, Xinzhong Zhu, Xinwang Liu, and Lizhe Wang. Cross-view local structure preserved diversity and consensus learning for multi-view unsupervised feature selection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 5101–5108, 2019.

[Tang *et al.*, 2023] Chang Tang, Xiao Zheng, Wei Zhang, Xinwang Liu, Xinzhong Zhu, and En Zhu. Unsupervised feature selection via multiple graph fusion and feature weight learning. *Science China Information Sciences*, 66(5):1–17, 2023.

[Van der Maaten and Hinton, 2008] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(11):2579–2605, 2008.

[Wang *et al.*, 2013] Hua Wang, Feiping Nie, and Heng Huang. Multi-view clustering and feature learning via structured sparsity. In *Proceedings of the International Conference on Machine Learning*, pages 352–360, 2013.

[Wang *et al.*, 2022] Siwei Wang, Xinwang Liu, Suyuan Liu, Jiaqi Jin, Wenxuan Tu, Xinzhong Zhu, and En Zhu. Align then fusion: Generalized large-scale multi-view clustering with anchor matching correspondences. *Advances in Neural Information Processing Systems*, 35:5882–5895, 2022.

[Xu *et al.*, 2024] Cai Xu, Jiajun Si, Ziyu Guan, Wei Zhao, Yue Wu, and Xiyue Gao. Reliable conflictive multi-view learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 16129–16137, 2024.

[Yang *et al.*, 2024] Zequn Yang, Han Zhang, Yake Wei, Zheng Wang, Feiping Nie, and Di Hu. Geometric-inspired graph-based incomplete multi-view clustering. *Pattern Recognition*, 147:110082, 2024.

[You *et al.*, 2023] Mengbo You, Aihong Yuan, Min Zou, Dong jian He, and Xuelong Li. Robust unsupervised feature selection via multi-group adaptive graph representation. *IEEE Transactions on Knowledge and Data Engineering*, 35(3):3030–3044, 2023.

[Zhang *et al.*, 2021] Han Zhang, Danyang Wu, Feiping Nie, Rong Wang, and Xuelong Li. Multilevel projections with adaptive neighbor graph for unsupervised multi-view feature selection. *Information Fusion*, 70:129–140, 2021.

[Zhang *et al.*, 2023a] Chao Zhang, Huaxiong Li, Wei Lv, Zizheng Huang, Yang Gao, and Chunlin Chen. Enhanced tensor low-rank and sparse representation recovery for incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 11174–11182, 2023.

[Zhang *et al.*, 2023b] Chenglong Zhang, Bingbing Jiang, Zidong Wang, Jie Yang, Yangfeng Lu, Xingyu Wu, and Weiguo Sheng. Efficient multi-view semi-supervised feature selection. *Information Sciences*, 649:119675, 2023.

[Zhao and Liu, 2007] Zheng Zhao and Huan Liu. Spectral feature selection for supervised and unsupervised learning. In *Proceedings of the International Conference on Machine Learning*, pages 1151–1157, 2007.

[Zhao *et al.*, 2023] Shuping Zhao, Jie Wen, Lunke Fei, and Bob Zhang. Tensorized incomplete multi-view clustering with intrinsic graph completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 11327–11335, 2023.

[Zhao *et al.*, 2024] Wenhui Zhao, Guangfei Li, Haizhou Yang, Quanxue Gao, and Qianqian Wang. Embedded feature selection on graph-based multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 17016–17023, 2024.

[Zhong and Pun, 2020] Guo Zhong and Chi-Man Pun. Data representation by joint hypergraph embedding and sparse coding. *IEEE Transactions on Knowledge and Data Engineering*, 34(5):2106–2119, 2020.

[Zhong and Pun, 2021] Guo Zhong and Chi-Man Pun. Improved normalized cut for multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):10244–10251, 2021.

[Zhou *et al.*, 2023] Peng Zhou, Jiangyong Chen, Liang Du, and Xuejun Li. Balanced spectral feature selection. *IEEE Transactions on Cybernetics*, 53(7):4232–4244, 2023.