Unsupervised Kernel-based Multi-view Feature Selection with Robust Self-representation and Binary Hashing

Rongyao Hu[1, 2], Jiangzhang Gan[3], Mengmeng Zhan[1], Li Li[4*], Mengling Wei[1]

[1] School of Computer Science and Technology,
University of Electronic Science and Technology of China, Chengdu 611731, China
[2] Knowledge and Data Engineering Laboratory of Chinese Medicine, School of Information and Software Engineering,
University of Electronic Science and Technology of China, Chengdu 610054, China
[3] School of Computer Science and Technology, Hainan University, Haikou 570228, China
[4] Computer School, Beijing Information Science and Technology University, Beijing 100101, China
hurongyao123@gmail.com, ganjzgxnu@163.com, memos.zhan@gmail.com, 20212599@bistu.edu.cn, ml_wei@126.com

Abstract

Unsupervised multi-view feature selection involves selecting a subset of crucial features across diverse views to diminish feature dimensionality without leveraging label information. While numerous studies may risk losing semantic information when applied to real-world multi-view datasets. In this study, we introduce a novel model, Unsupervised Kernel-based Multi-view Feature selection with Robust self-representation and Binary hashing (UKMFS), which aims to identify robust consistent graph representation across views and leverage binary hashing codes to guide feature selection. Specifically, we first explore the underlying geometry by unifying the dimension of multi-view data with non-linear kernel mapping. Then, we search for consistent graphs across views by fusing unique graph representations of each view in a self-representation manner. Additionally, we impose low-rank constraints on the graph of each view to mitigate noise and unimportant parts to preserve the main structures and patterns. Furthermore, we design an unsupervised hashing feature selection model to exploit reliable binary labels across views and weighted matrices from each view. Finally, an effective optimization method is customized to solve the formulated problem. Comprehensive experiments on public multi-view datasets indicate that our proposed method achieves state-of-the-art performance compared with the representative comparison methods regarding the clustering and the feature selection task.

Introduction

With the advancement of information theory and data storage technology, high-dimensional multi-view data has become increasingly prevalent, evident in diverse applications, such as multi-pose pedestrian retrieval, multi-view image registration, and multi-functional drug molecule mining (Yu et al. 2024; Bolkart, Li, and Black 2023; Hua et al. 2023). However, handling high-dimensional multi-view data presents significant challenges since it demands greater computational cost and incurs more storage expense. This reality poses substantial obstacles for researchers in the development of practical applications. Furthermore, directly dealing with high-dimensional multi-view data often leads to the "curse of dimensionality" to occur. Therefore, multi-view feature selection has been developed to select a small subset of features as a substitute for the original data, offering an effective way to improve the discriminative ability of these features (Tang et al. 2023; Fan et al. 2024).

According to the usage of labels, existing strategies for multi-view feature selection sort into three groups, *i.e.,* supervised-based methods (Han, Hu, and Gao 2024; Liu et al. 2023), semi-supervised-based methods (Zhang et al. 2023; Jiang et al. 2023), and unsupervised-based methods (Cao and Xie 2024a; Gu et al. 2023). Supervised multi-view feature selection methods strongly rely on label information to supervise the modeling and feature selection, yet they fall short in excavating the latent information hidden in unlabeled data which results in an incomplete or inaccurate feature selection. Furthermore, as the labeling information plummets, semi-supervised multi-view feature selection methods fuse information from both labeled and unlabeled data, but they grapple with balancing these two categories of data. Additionally, constructing complex algorithms to integrate labeled and unlabeled data leads to steep computational costs. Consequently, unsupervised multi-view feature selection has garnered the focus of most researchers given its independence from labeled data and its extensive applicability.

For unsupervised multi-view feature selection, the important objective is effectively grasping relations within the data across multiple views and then capitalizing on these correlations to steer the crucial task of feature selection. To fulfill the goal of searching for these data connections and relationships between views, existing methods turn to the spectral graph representation way (Yang et al. 2024b; Fang et al. 2023) and the cluster structure way (Xiang et al. 2024; Zhao, Yang, and Nie 2023). They are always involved in a two-step learning procedure. (1) Spectral graph representation is conducted by graph Laplacian matrix, and then both the number of nearest neighbors and the graph representation are predefined from the original multi-view data manually. (2) Projection matrix (*i.e.,* feature selection) is learned from each view and is guided by the spectral graph representation across views. Although remarkable performances

*Corresponding Author

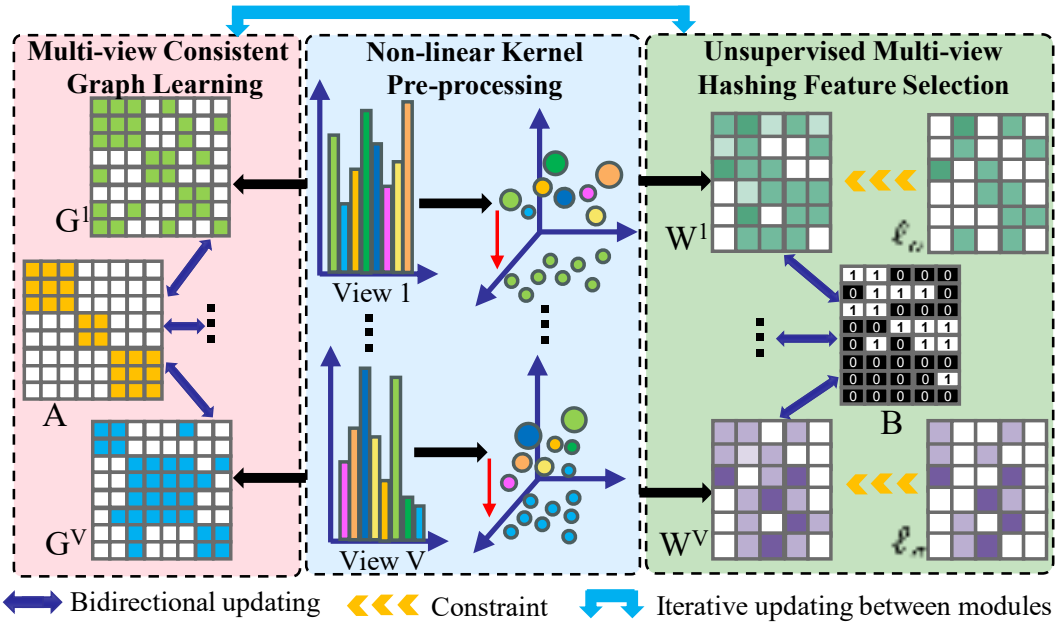


Figure 1: The proposed framework comprises three parts: (1) **Non-linear Kernel Pre-processing Module** unifies the dimensions across views. (2) **Multi-view Consistent Graph Learning Module** considers the weight of each view and obtains consistent graph from multi-view data. (3) **Unsupervised Multi-view Hashing Feature Selection Module** obtains the projection matrix of each view.

have been obtained, there are four limitations that inevitably need to be confronted: (1) With graph learning and feature selection separated, the interaction of features might not be adequately captured across views. The inability to account for these complex correlations can limit the performance of multi-view learning and also constrain the search for significant features of each view. (2) The process of feature selection may not always converge to an optimal result, since the pre-constructed graph is unchangeable in the model training. (3) The influence of noise and outliers can be amplified due to the increased dimensionality of multiple views. It becomes easier for irrelevant features or outliers to dominate one view in graph learning, which skews the learned model and misguides the learning process, leading to the inaccuracy of model construction. (4) Both the given labels and the learnt labels would be beneficial to improve the performance of unsupervised learning since labels provide the semantic information to guide common similarity learning and feature selection among multiple views.

Several works have been proposed to address parts of the above issues, via using the joint learning framework consists of dynamic graph learning and feature selection for multi-view data, such as Joint Multi-view unsupervised Feature selection and Graph learning (JMVFG) (Fang et al. 2023), unsupervised Complementary Feature Selection with Multi-Order similarity learning (CFSMO) (Cao and Xie 2024a). However, they fall short in dealing with the lack of semantic depth in the feature selection, as well as failing to obtain a reliable similarity representation from unified multi-view data. Therefore, Shi *et al.* propose a multi-View unsupervised Feature selection with the binary hashing method (Shi et al. 2023) to learn weakly-supervised labels to guide the feature selection procedure by giving multiple graphs from different views. Wang *et al.* propose a Graph-Collaborated Auto-Encoder hashing method (Wang et al. 2023) to conduct the reliable graph representation from unified multi-view data to guide weakly-supervised labels. Unfortunately, These methods use either Frobenious-norm or $\ell_{2,1}$-norm to remove the unimportant features in feature selection, ignoring the negative effect of outliers and non-differentiable.

To solve the challenges mentioned above, we propose a novel Unsupervised Kernel-based Multi-view Feature selection with Robust self-representation and Binary hashing (UKMFS) model to adaptively explore consistent graph representation among multi-view and to search weakly-supervised multi-labels with binary hashing for guiding feature selection and similarity learning of each view. Specifically, we first adopt the non-linear RBF kernel mapping method to unify the dimension of multi-view for preserving the underlying geometric structure. Then, we search multi-view consistent graph representation by adaptively learning the weight of each view, while affinity graph representation of each view with low-rank constraint is learned to preserve significant information and the underlying structure of multi-view data. Moreover, we design unsupervised multi-view feature selection with binary hashing to exploit the representative feature in every view by considering project matrix of each view and by employing weakly-supervised labels with binary hashing codes across multiple views. Furthermore, we employ adaptive loss function to search the

essential features in project matrix, as well as to find the robustness control between Frobenious-norm or $\ell_{2,1}$-norm. Finally, an effective optimization scheme is proposed to search the optimal solution of each variables iteratively. Experiments are tested on public multi-view datasets, which have proved the superior result of the proposed method for unsupervised multi-view feature selection task. Figure 1 illustrates the framework of the proposed method. It is worthwhile to highlight the main contributions of this paper as follows:

- We present a novel adaptive unsupervised feature selection structure referring to two types of modules, including the multi-view consistent graph learning adaptively searches the affinity graph to each view and to all views, and the unsupervised multi-view hashing feature selection that adaptively investigates the weakly-supervised labels to all views and the projection matrix to each view.
- The proposed method unifies robust self-representation learning, weakly-supervised multi-label learning, and adaptive graph learning into one framework. This joint learning scheme guarantees that the three structures can connect with each other to better serve unsupervised multi-view feature selection.
- We propose a novel optimization algorithm and experimentally verify the effectiveness of the proposed unsupervised multi-view feature selection method.

Approach

Notations and Definitions

In this paper, matrices, vectors, and scalars are defined in boldface uppercase, boldface lowercase, and normal italic, respectively, *e.g.,* $\mathbf{X}$, $\mathbf{x}$, and x.

Non-linear Kernel Pre-processing

Given a multi-view dataset composed of V views for n samples, which are represented by a set of matrices $\{\mathbf{X}^1, ..., \mathbf{X}^V\}$. Each view $\mathbf{X}^v \in \mathbb{R}^{n\times d^v}$ consists of n samples and each sample contains d^v features. For simplicity, we assume that the data samples in each view are zero-centered, *i.e.,* $\sum_i \mathbf{x}^v_{i,\cdot} = 0$.

Generally, the features of each view should undergo normalization to guarantee balance, especially considering the potential varying dimensions. However, the original multi-view data consists of redundant and noisy features with a large value, leading to normalised amplification of unimportant features. To address this, a non-linear RBF kernel scheme is employed to map the original multi-view data into a lower-dimensional feature space. We encode i-th sample $\mathbf{x}^v_{i,\cdot}$ from the v-th view into a s-dimensional non-linear embedding $\phi(\mathbf{x}^v_{s,\cdot}) \in \mathbb{R}^s$ by the non-linear RBF kernel mapping, and each view can be calculated by

$$[exp(-\frac{||\mathbf{x}^v_{i,\cdot}-\mathbf{a}^v_{1,\cdot}||_2^2}{\eta}); ...; exp(-\frac{||\mathbf{x}^v_{i,\cdot}-\mathbf{a}^v_{s,\cdot}||_2^2}{\eta})], \tag{1}$$

where η denotes the kernel width, $\{\mathbf{a}^v_{1,\cdot}, ..., \mathbf{a}^v_{s,\cdot}\}$ denotes the randomly selected s anchor samples from the v-th view. Its purpose is to guarantee the original data structure can be maintained after non-linear projection and to maintain the balance of the data across views. To clarify, we let $\hat{\mathbf{X}}^v = \mathbf{X}^v\mathbf{Z}^v$, where $\hat{\mathbf{X}}^v \in \mathbb{R}^{n\times s}$ denotes the reconstruction representation, and $\mathbf{Z}^v \in \mathbb{R}^{s\times d^v}$ denotes the fixed projection by the operation in Eq. (1).

Multi-view Consistent Graph Learning

Motivated by the low-rank representation way (Wang et al. 2023) and sample-level self-representation (Zhang et al. 2022), the affinity graph $\mathbf{G}^v$ is learned from $\hat{\mathbf{X}}^v$ in each view. Then, the model is conducted by

$$\begin{aligned} &\min_{\mathbf{G}^v} \textstyle\sum_{v=1}^V (||\hat{\mathbf{X}}^v - \mathbf{G}^v\hat{\mathbf{X}}^v||_F^2) \\ &\quad s.t.,\ \ rank(\mathbf{G}^v) \le min(n, s), \end{aligned} \tag{2}$$

where $\mathbf{G}^v \in \mathbb{R}^{n\times n}$ denotes the learned graph for v-th view, $rank(.)$ denotes the low-rank constraint of $\mathbf{G}^v$. By minimizing Eq. (2), the obtained graph preserves the signification information of each view data in the low-dimensional subspace, as well as explores the low-rank representation from each view data.

To explore the collaboration across views, a common graph is learned from multiple affinity graphs in multi-view data since it describes the underlying common structure for diverse views, as well as it provides improvement to guide individual affinity graph in each view. We have

$$\begin{aligned} &\min_{\mathbf{G}^v,\mathbf{A}} \textstyle\sum_{v=1}^V (||\hat{\mathbf{X}}^v - \mathbf{G}^v\hat{\mathbf{X}}^v||_F^2 + \lambda^v||\mathbf{A} - \mathbf{G}^v||_F^2) \\ &\quad s.t.,\ \ rank(\mathbf{G}^v) \le min(n, s), \\ &\qquad \textstyle\sum_j a_{i,j} = 1, a_{i,j} \ge 0, \end{aligned} \tag{3}$$

where λ^v denotes the control parameter for each view, $\mathbf{A} \in \mathbb{R}^{n\times n}$ denotes the common graph representation across views.

However, the control parameter λ^v needs to be adjusted manually for each view, leading to a significant expenditure of manpower and is impractical on a larger scale of multi-view data. To alleviate this issue, an automatic weight learning method is proposed as follows:

$$\begin{aligned} &\min_{\mathbf{G}^v,\mathbf{A}} \textstyle\sum_{v=1}^V (||\hat{\mathbf{X}}^v - \mathbf{G}^v\hat{\mathbf{X}}^v||_F^2 + ||\mathbf{A} - \mathbf{G}^v||_F) \\ &\quad s.t.,\ \ rank(\mathbf{G}^v) \le min(n, s), \\ &\qquad \textstyle\sum_j a_{i,j} = 1, a_{i,j} \ge 0. \end{aligned} \tag{4}$$

Eq. (4) provides parameter-free way between common graph $\mathbf{A}$ and multiple graphs $\{\mathbf{G}^1, ..., \mathbf{G}^V\}$.

Unsupervised Multi-view Hashing Feature Selection

To search the informative features under the unsupervised learning framework, an effective approach involves acquiring weakly-supervised labels through spectral analysis (Jiang et al. 2021). These weakly-supervised labels serve as guidance for the feature selection procedure. Existing methods (Yuan et al. 2019; Shi et al. 2023) usually conduct a unified model to exploit the single weakly-supervised label matrix $\mathbf{Y} \in \mathbb{R}^{n\times c}$ and feature selection matrix $\mathbf{W} \in \mathbb{R}^{d\times c}$ together. The general model is designed by

$$\begin{aligned} &\min_{\mathbf{Y},\mathbf{W}} ||\mathbf{XW} - \mathbf{Y}||_F^2 + \alpha\vartheta(\mathbf{Y}) + \gamma\vartheta(\mathbf{W}) \\ &\quad s.t.,\ \ \mathbf{Y} \in C, \end{aligned} \tag{5}$$

Dataset	Feature rates	UAFSBH	ACSL	NSGL	AMFS	JMVFG	SCMvFS	UKMFS
Caltech101	10%	43.99±3.77	53.98±4.44	44.79±5.69	54.39±2.12	**68.19**±2.13	67.73±2.88	67.89±2.23
	15%	47.83±3.56	54.81±4.17	49.31±6.03	57.88±2.05	67.31±2.47	67.95±2.95	**68.82**±2.19
	20%	52.18±3.20	55.43±3.99	48.22±6.31	60.17±2.44	69.33±3.02	67.06±2.34	**71.10**±1.78
	25%	53.41±3.19	55.11±2.80	49.46±5.37	63.41±2.56	70.42±2.44	66.02±2.17	**70.62**±2.01
	30%	51.18±3.44	54.71±3.01	48.83±6.77	59.83±2.17	68.21±4.02	62.35±2.09	**70.17**±2.05
HW	10%	77.37±3.51	79.74±5.90	54.62±5.62	65.07±1.04	82.50±8.33	81.82±4.22	**83.43**±3.37
	15%	74.04±4.20	80.47±5.82	65.04±5.81	63.02±2.49	**84.19**±8.27	83.53±4.39	83.98±3.42
	20%	70.69±4.26	81.66±6.20	69.29±5.63	72.37±1.85	83.42±8.55	82.44±4.76	**85.52**±3.26
	25%	75.10±4.13	84.31±6.76	72.92±5.33	74.99±2.72	**85.83**±8.41	82.19±4.19	85.39±3.17
	30%	74.81±4.76	81.17±6.33	74.53±5.11	77.55±1.81	82.51±7.94	80.30±3.77	**84.77**±3.10
MSRCV1	10%	69.34±6.63	77.60±5.81	73.41±4.52	58.94±1.72	**80.59**±6.61	78.31±6.33	80.32±2.99
	15%	71.18±5.83	**80.69**±5.77	76.21±4.38	59.52±1.61	79.55±6.34	79.92±5.72	80.03±3.41
	20%	75.85±3.73	77.48±5.63	75.09±4.60	59.23±1.52	78.20±6.11	77.44±5.50	**78.88**±3.22
	25%	73.38±4.11	75.71±5.29	73.74±4.19	59.90±1.36	76.60±5.84	78.09±5.39	**78.40**±3.17
	30%	72.49±4.66	74.82±5.03	72.44±4.52	58.74±1.99	72.83±5.47	76.53±5.44	**78.21**±3.32
Scene	10%	60.32±3.49	64.42±5.51	48.37±3.07	63.13±1.64	64.79±4.50	61.58±3.43	**69.31**±2.10
	15%	60.05±3.39	65.34±5.79	53.16±3.25	68.16±1.72	66.38±4.22	62.77±3.27	**69.43**±2.03
	20%	60.19±4.16	66.53±5.43	55.18±2.85	**68.53**±1.69	67.37±4.17	63.96±3.18	67.12±2.77
	25%	60.59±3.69	63.41±6.21	58.41±3.41	**69.26**±1.71	64.11±5.63	63.17±3.77	62.35±2.81
	30%	59.71±3.68	60.66±6.39	54.26±2.92	60.45±1.52	61.73±5.98	60.42±2.88	**64.40**±2.45
ORL	10%	63.32±1.87	61.77±3.28	43.79±1.83	52.73±3.33	60.55±4.22	62.09±2.85	**65.14**±1.11
	15%	60.97±1.83	62.59±3.43	42.43±2.32	58.17±3.54	61.76±4.44	60.16±2.55	**65.32**±1.70
	20%	61.84±1.35	60.82±3.88	41.78±2.17	56.32±3.71	59.78±4.92	61.27±2.47	**64.40**±1.41
	25%	61.33±1.23	60.39±4.02	41.33±2.43	55.83±3.10	57.92±5.31	61.51±2.23	**64.08**±1.52
	30%	61.87±1.59	59.54±4.41	40.75±2.71	54.55±2.87	57.38±5.19	60.33±1.87	**64.22**±1.83
WebKB	10%	57.18±5.02	71.37±3.79	72.18±4.13	64.45±3.89	66.43±3.12	71.48±2.37	**74.41**±1.97
	15%	64.32±5.77	72.33±3.88	73.41±3.97	66.21±4.15	69.76±3.41	**79.57**±2.68	78.83±2.20
	20%	65.28±5.80	70.47±4.55	75.58±4.57	67.03±3.97	67.17±4.03	**79.03**±2.75	78.50±2.21
	25%	64.31±5.66	66.58±4.47	74.63±3.95	65.72±3.88	66.25±3.58	76.23±3.05	**78.01**±2.03
	30%	64.02±5.52	64.20±4.23	**73.90**±4.07	61.14±4.55	64.31±3.27	73.41±3.27	73.52±2.22

Table 1: ACC of all methods with different percentages of features. The best results are denoted in bold.

where α and γ are adjusted parameters for controlling the balance of regularization terms, C is the employed discrete and binary constraint on the label $\mathbf{Y}$. With the constraints, each row of matrix $\mathbf{Y}$ has only one non-zero element, *i.e.,* 1, and the rest being 0.

However, Eq. (5) has limitations when it applies for multiple views data in the real-world. Therefore, an unsupervised multi-view hashing feature selection model is proposed. Specifically, we dynamically acquire a specific quantity of weakly-supervised labels using binary hashing techniques from the available unsupervised samples, thoroughly leveraging the inherent data structure. These labels are then utilized as semantic guidance for the final feature selection process across views. Therefore, we generate the binary labels by explicitly incorporating hash constraints during the spectral embedding phase. It is worth noting that the quantity of weakly-supervised labels is flexibly determined based on the particular characteristics of the data, thus preventing any loss of semantic information. We rewrite Eq. (5) to

$$\begin{aligned} &\min_{\mathbf{B},\mathbf{W}^v} ||\mathbf{X}^v\mathbf{W}^v - \mathbf{B}||_F^2 + \alpha\vartheta(\mathbf{B}) + \gamma\vartheta(\mathbf{W}^v) \\ &s.t.,\ \ \mathbf{B} \in \{0,1\}^{n\times l}, \end{aligned} \tag{6}$$

where $\mathbf{W}^v \in \mathbb{R}^{d^v\times l}$ denotes the feature selection matrix of each view, $\mathbf{B}$ denotes the learnable binary label, l defines the length of binary hash codes.

The spectral embedding has indicated the successful application (Shi et al. 2021, 2023), so we enforce the binary hashing constraint during the spectral embedding process to generate binary labels while leveraging the data structure. In this way, we introduce a constraint term in the label learning terms, *i.e.,*

$$\vartheta(\mathbf{B}) = Tr(\mathbf{B}^T\mathbf{L}\mathbf{B}) \tag{7}$$

where $\mathbf{L} = \mathbf{D}_a - (\mathbf{A}^T + \mathbf{A})/2$ denotes the Laplacian matrix of an affinity matrix $\mathbf{A}$, and $\mathbf{D}_a$ denotes a diagonal degree matrix and each diagonal element is defined as $\sum_j (a_{i,j} + a_{j,i})/2$.

Although the Frobenious-norm constraint is always used in the project matrix $\mathbf{W}^v$ providing a clear geometric interpretation and a simple computation, it is sensitive to outliers and lacks sparsity. Comparatively, the $\ell_{2,1}$-norm provides a sparsity solution and robustness to outliers, but it is non-differentiable and is non-smooth increasing the difficulty of optimization problems.

In order to consider their advantages, the adaptive loss function is used (Zhang et al. 2019) and is defined as

$$||\mathbf{W}^v||_\sigma = \textstyle\sum_i \frac{(1+\sigma)||\mathbf{w}^v_{i,\cdot}||_2^2}{||\mathbf{w}^v_{i,\cdot}||_2+\sigma}, \tag{8}$$

where σ denotes a trade-off parameter to control the robustness for outliers.

Eq. (8) is twice differential, convex and non-negative, and the relationship between it and the Frobenious-norm and the $\ell_{2,1}$-norm is listed as follows:

$$||\mathbf{W}^v||_\sigma \rightarrow \begin{cases} ||\mathbf{W}^v||_F^2, & \sigma \rightarrow \infty, \\ ||\mathbf{W}^v||_{2,1}, & \sigma \rightarrow 0, \\ \frac{1+\sigma}{\sigma}||\mathbf{W}^v||_F^2, & \forall i, ||\mathbf{w}_{i,\cdot}^v|| \ll \sigma, \\ (1+\sigma)||\mathbf{W}^v||_{2,1}, & \forall i, ||\mathbf{w}_{i,\cdot}^v|| \gg \sigma. \end{cases} \tag{9}$$

By combing with Eq. (6), Eq. (7) and Eq. (8), we obtain an unsupervised hashing feature selection model as follows,

$$\begin{aligned} &\min_{\mathbf{B},\mathbf{W}^v} ||\mathbf{X}^v\mathbf{W}^v - \mathbf{B}||_F^2 + \alpha Tr(\mathbf{B}^T\mathbf{L}\mathbf{B}) + \gamma||\mathbf{W}^v||_\sigma \\ &s.t., \ \ \mathbf{B} \in \{0,1\}^{n\times l}. \end{aligned} \tag{10}$$

In this way, $\mathbf{B}$ is derived from the "new" data (*i.e.,* $\hat{\mathbf{X}}^v\mathbf{W}^v$) of each view while $\mathbf{W}^v$ is obtained by the guidance of the common binary label across views.

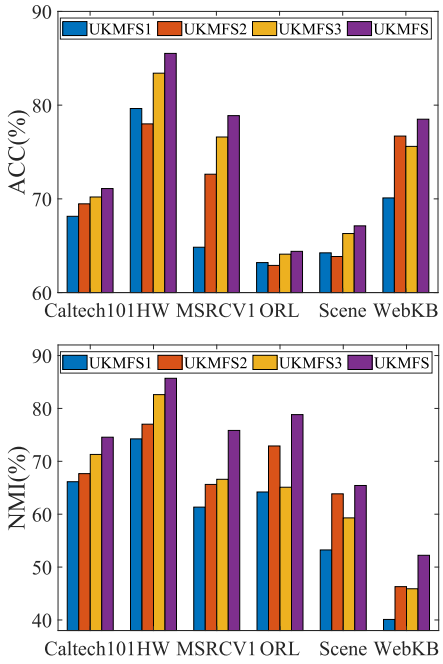


Figure 2: ACC and NMI of Proposed method and its variants.

Overall Objective Function

In Eq. (10), each original view data usually contains redundant information and outliers, and the initialized graph matrix needs to be provided first, respectively, leading to unreliable data support and the time-consuming graph process. Therefore, By combing with Eq. (4), we derive our final objective function as follows,

$$\begin{aligned} \min_{\mathbf{G}^v,\mathbf{A},\mathbf{B},\mathbf{W}^v,\mu^v} & \sum_{v=1}^V ||\hat{\mathbf{X}}^v - \mathbf{G}^v\hat{\mathbf{X}}^v||_F^2 \\ & + \tfrac{1}{\mu^v}||\mathbf{A} - \mathbf{G}^v||_F^2 + \alpha Tr(\mathbf{B}^T\mathbf{L}_A\mathbf{B}) \\ & + \beta||\hat{\mathbf{X}}^v\mathbf{W}^v - \mathbf{B}||_F^2 + \gamma||\mathbf{W}^v||_\sigma \\ s.t., \ \ & rank(\mathbf{G}^v) \leq min(n,s), \textstyle\sum_j a_{i,j} = 1, \\ & a_{i,j} \geq 0, \mu^v \geq 0, \mu^v\mathbf{1} = 1, \mathbf{B} \in \{0,1\}^{n\times l}. \end{aligned} \tag{11}$$

where α, β and γ are tuning parameters. The first term learns affinity graph of each view by sample self-representation and kernel way, the second term learns a common affinity common across views by fusing all affinity graph of each view, the third term performs binary spectral embedding on the common graph, the fourth term conducts binary learning and feature selection with unsupervised learning together, and the fifth term designs adaptive sparse learning to project matrix.

Experiments

Experimental Settings

In this section, six multi-view datasets are tested, including Caltech101 [1], HW [2], MSRCV1 (Yang et al. 2024a), Scene (Shi et al. 2023), ORL [3], and WebKB [4]. To fully evaluate the efficiency and effectiveness of UKMFS, we design the experiments with six state-of-the-art feature selection approaches, including (1) Unsupervised Adaptive Feature Selection with Binary Hashing (UAFSBH) (Shi et al. 2023); (2) Adaptive Collaborative Similarity Learning (ACSL) (Dong et al. 2018); (3) Non-negative Structure Graph Learning (NSGL) (Bai et al. 2020); (4) Adaptive Multi-view Feature Selection (AMFS) (Wang et al. 2016); (5) Joint Multi-view Unsupervised Feature Selection and Graph Learning (JMVFG) (Fang et al. 2023); (6) Structure Learning and Consensus Label Information for Multi-view Feature Selection (SCMvFS) (Cao and Xie 2024b).

We used all methods to obtain significant features to guarantee fair comparisons and employed the K-means on the selected features to conduct clustering tasks. The parameters for all comparison approaches are tuned on the basis of their studies. For the proposed method, the parameters α, β and γ are adjusted from $\{10^{-2}, ..., 10^2\}$ by the grid-search way, the length of hash codes l is selected from $\{4, 8, 16, 32, 64\}$, the low-rank r is defined as the range of $min\{n, s\}$. For the operation ϕ, η is set to 0.5, and anchor number s is set to one value from the range of $\{100, 1000\}$ based on the order to magnitude of the samples for each dataset. While obtaining an important feature subset, K-means is used to obtain the clustering performance 20 times independently, and the results are reported on average. The evaluation metrics contain average clustering accuracy (ACC) and normalized mutual information (NMI).

Experimental Results and Analysis

Tables 1-2 demonstrate the average and standard deviation of ACC and NMI, and the best results are denoted in bold. Then, we obtain the following conclusions: (1) As the percentage of features grows, UKMFS usually obtains superior performance to the selected comparison methods, demonstrating its effectiveness while coping with multi-view unsupervised feature selection tasks. (2) UKMFS

[1]http://www.vision.caltech.edu/ImageDatasets/Caltech101/..

[2]https://archive.ics.uci.edu/dataset/72/multiple+features..

[3]https://www.kaggle.com/datasets/tavarez/the-orl-database-for-training-and-testing.

[4]http://www.webkb.org/.

Dataset	Feature rates	UAFSBH	ACSL	NSGL	AMFS	JMVFG	SCMvFS	UKMFS
Caltech101	10%	67.71±4.77	66.31±3.76	65.87±5.49	72.13±2.89	63.11±3.25	**74.41**±2.17	73.70±2.78
	15%	67.16±4.56	66.53±3.55	65.32±5.73	72.01±2.77	65.72±3.67	73.82±2.09	**74.44**±3.05
	20%	66.83±4.43	65.39±3.44	64.78±5.50	70.64±2.55	68.31±3.61	73.44±1.89	**74.56**±2.87
	25%	66.19±4.16	61.19±4.02	62.05±4.87	68.54±2.37	69.33±3.03	72.11±2.10	**73.41**±3.01
	30%	65.18±3.82	62.81±4.19	62.48±4.66	66.19±2.15	68.01±4.77	72.67±2.23	**72.69**±2.96
HW	10%	76.58±1.17	68.41±3.09	43.52±3.79	72.61±2.45	**79.50**±3.79	78.21±3.18	79.46±2.22
	15%	76.21±1.09	75.74±3.58	55.70±3.63	77.44±2.13	82.19±3.55	80.32±2.94	**83.47**±2.37
	20%	73.71±1.40	82.41±3.61	64.57±3.44	77.18±2.53	85.43±3.13	80.55±2.77	**85.70**±2.41
	25%	75.01±1.79	81.37±3.44	67.89±3.17	80.13±2.43	**83.52**±2.86	80.19±2.85	81.08±2.20
	30%	74.62±1.55	78.77±3.02	72.09±3.25	78.23±2.61	82.77±2.75	78.43±2.61	**83.19**±2.14
MSRCV1	10%	65.45±5.81	71.82±4.52	60.92±4.63	62.68±3.90	72.19±3.47	69.52±6.17	**72.55**±3.18
	15%	68.63±5.76	72.41±4.79	61.33±4.38	65.92±2.98	74.81±4.03	71.23±5.89	**75.92**±3.55
	20%	72.15±5.73	75.50±4.66	63.05±4.19	66.13±2.67	75.83±3.12	70.18±5.71	**75.84**±3.66
	25%	69.05±5.46	73.32±4.83	61.79±3.98	65.75±2.72	73.63±3.32	67.91±5.54	**74.18**±3.30
	30%	67.72±5.21	**74.04**±5.19	59.41±4.27	66.70±2.05	74.17±3.20	64.79±5.20	73.37±3.51
Scene	10%	47.65±1.19	55.83±2.91	43.41±1.73	55.10±1.71	53.59±2.08	44.38±1.44	**58.80**±2.28
	15%	48.31±0.92	56.44±2.62	44.88±1.37	**62.67**±1.73	56.32±1.87	48.71±1.27	59.19±2.99
	20%	49.05±1.83	56.87±2.37	46.39±0.99	**67.72**±1.31	57.10±1.61	51.67±0.89	65.43±3.37
	25%	48.76±1.45	54.80±3.05	47.68±0.62	64.75±1.52	55.53±2.17	50.36±1.38	**65.01**±3.91
	30%	48.44±1.81	51.95±3.76	45.26±0.88	**66.23**±1.48	52.11±2.49	49.03±0.98	64.80±4.33
ORL	10%	73.66±1.27	78.44±1.91	64.76±1.08	71.23±3.38	78.90±2.44	79.35±2.27	**80.04**±3.31
	15%	73.16±1.49	77.11±2.03	64.37±1.33	75.52±3.99	77.43±2.67	77.29±2.45	**79.92**±4.50
	20%	73.25±1.23	75.83±2.41	63.69±1.52	75.02±4.21	76.72±2.89	78.51±2.69	**78.83**±3.54
	25%	73.48±1.54	71.71±2.88	62.90±1.44	74.72±3.75	73.48±3.01	76.19±3.11	**76.60**±4.00
	30%	73.07±1.83	73.43±2.65	61.76±1.77	74.41±3.46	74.94±3.31	75.39±2.77	**75.59**±3.31
WebKB	10%	52.18±4.55	43.53±3.51	40.66±6.32	48.83±4.21	41.85±4.98	46.59±3.39	**54.40**±5.33
	15%	53.39±4.67	41.74±3.58	41.33±6.45	52.20±4.12	40.55±4.76	46.22±3.52	**54.79**±5.18
	20%	**52.80**±4.88	41.96±3.90	40.54±6.12	53.49±3.39	41.63±4.55	45.31±3.76	52.23±5.02
	25%	52.22±5.19	40.22±3.61	39.88±5.97	51.08±4.55	39.79±4.11	43.74±4.07	**53.07**±4.77
	30%	52.31±5.30	38.87±4.76	40.36±5.70	47.62±4.72	38.43±4.33	41.98±4.32	**52.44**±4.81

Table 2: NMI of all methods with different percentages of features. The best results are denoted in bold.

outperforms UAFSBH which also considers the weakly-supervised multi-label learning with graph learning, indicating the learning procedure in subspace benefits the feature selection. (3) UKMFS achieves superior results than the methods that focus on selecting features with cluster-based similarity preservation (*i.e.,* JMVFG and SCMvFS), showing the importance of leveraging discriminative features guided by the multi-label. (4) Compared with the methods that utilize multiple graphs learning and feature selection (*i.e.,* ACSL, NSGL, and AMFS), UKMFS obtains better results in most cases, emphasizing the importance of both multi-label information and reliable similarity structure.

Ablation Study

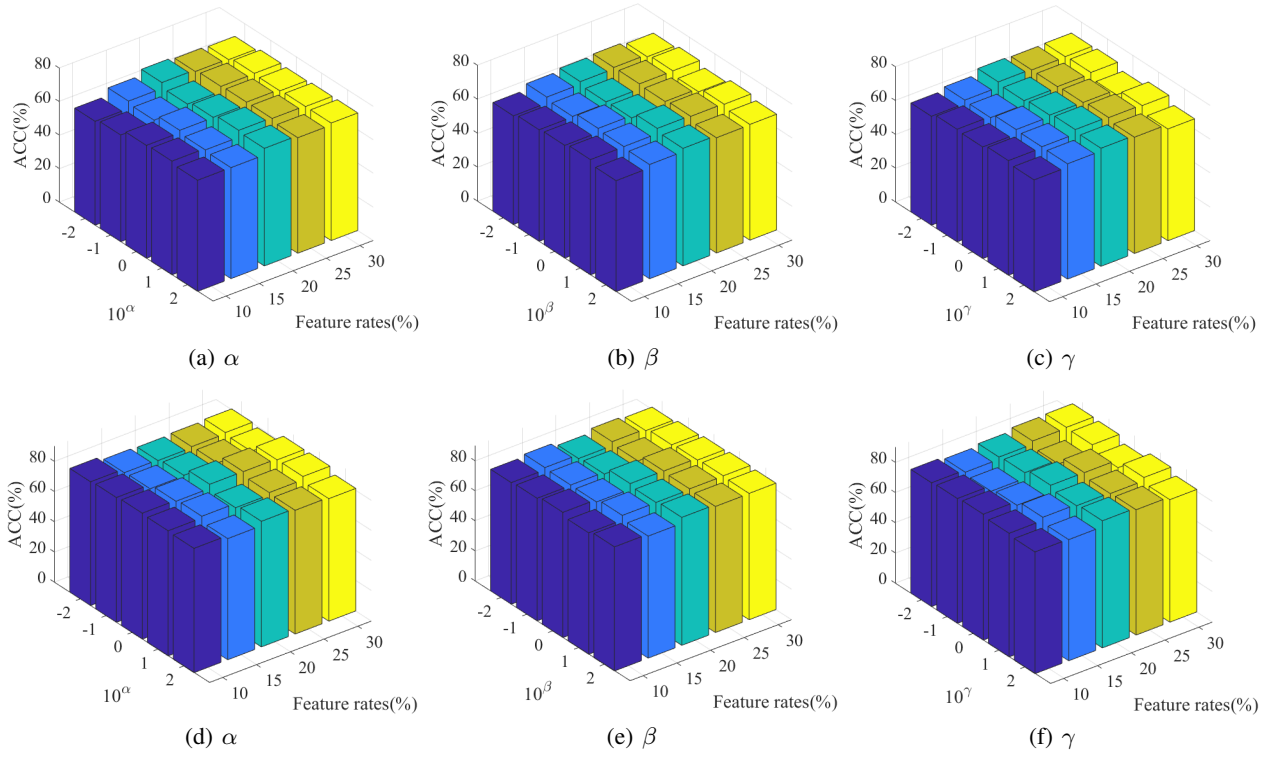
To verify the effectiveness of the signification components in the proposed method, an ablation experiment is designed. Specifically, three variants of UKMFS have been exploited, that is, UKMFS1 ignores the non-linear kernel process for multi-view data, and only conducts similarity structure from the original data; UKMFS2 removes the adaptive weight learning while fusing all view-specific graphs and the low-rank constraint for each view-specific graph; UKMFS3 neglects the learnable binary label and uses the single-label way while executing feature selection. The experimental results are shown in Figure 2, where the percentage of selected features is fixed as 20%. We have the following conclusion, (1) the non-linear kernel process indicates the important step in our proposed method while UKMFS outperforms UKMFS1. (2) Comparative results between UKMFS2 and UKMFS3 validate that the adaptive weight learning and low-rank constraint play an important role in the proposed method than the learnable binary label way.

Parameters Sensitivity and Convergence Analysis

UKMFS considers three kinds of parameters, where α controls the smoothness of the consistency graph and multi-label information across views, β controls the loss function regarding the multi-label learning across views and γ balances the influence of feature selection matrix of each view. Figure 3 shows the clustering accuracy with different parameter settings. We observe that when the values of three parameters are set to 10, UKMFS obtains superior performance. Moreover, Figure 4 illustrates the values of the objective function with the various number of iterations, showing that the value of the objective function decreases rapidly and converges within a fixed iteration, which validates the effectiveness of the proposed optimization scheme.

Conclusion

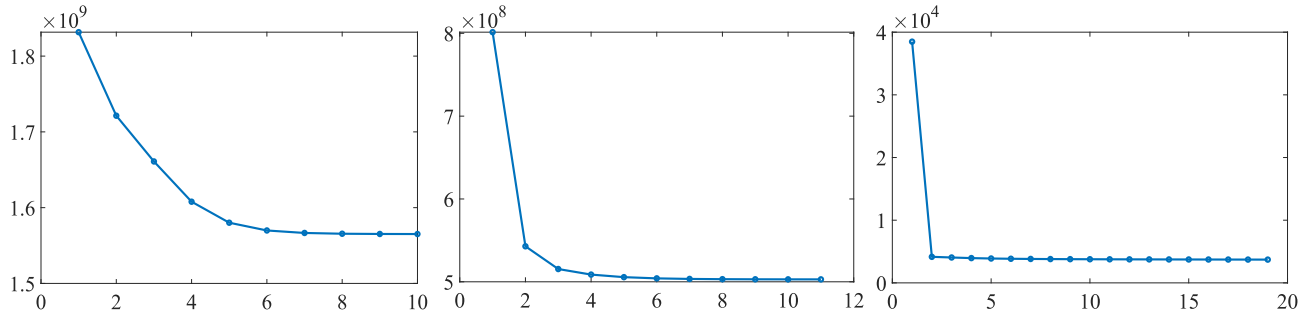
This paper introduces a novel unsupervised feature selection model to incorporate robust self-representation learning, weakly-supervised multi-label learning, and adaptive



(a) α (b) β (c) γ

(d) α (e) β (f) γ

Figure 3: ACC with different parameters on Caltech101 (top) and HW (bottom).



(a) Caltech101 (b) HW (c) MSRCV1

Figure 4: The visualization of objective functional values.

graph learning for multi-view data. We employ a non-linear kernel to pre-process the original data to obtain the robust multi-view data with cluster guidance in the subspace. Then, multi-view consistent graph learning and unsupervised multi-view hashing feature selection are designed, respectively, to fuse the view-specific graphs adaptively and to obtain the feature selection matrix guided by the weakly-supervised multi-label information. Moreover, a novel optimization algorithm is conducted to solve the proposed method and the corresponding experiments on six real-world multi-view datasets demonstrated that our proposed method outperforms the comparison methods, concerning diverse evaluation metrics and tasks.

The future work will focus on reducing time complexity and improving efficiency in processing.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under (Grant No. 62306066), the Sichuan Province Innovative Talent Funding Project for Postdoctoral Fellows under (Grant No. BX202313), and the R&D Program of Beijing Municipal Education Commission (Grant No. KM202411232004).

References

Bai, X.; Zhu, L.; Liang, C.; Li, J.; Nie, X.; and Chang, X. 2020. Multi-view feature selection via nonnegative structured graph learning. *Neurocomputing*, 387: 110–122.

Bolkart, T.; Li, T.; and Black, M. J. 2023. Instant multi-view head capture through learnable registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 768–779.

Cao, Z.; and Xie, X. 2024a. Multi-view unsupervised complementary feature selection with multi-order similarity learning. *Knowledge-Based Systems*, 283: 111172.

Cao, Z.; and Xie, X. 2024b. Structure learning with consensus label information for multi-view unsupervised feature selection. *Expert Systems with Applications*, 238: 121893.

Dong, X.; Zhu, L.; Song, X.; Li, J.; and Cheng, Z. 2018. Adaptive collaborative similarity learning for unsupervised multi-view feature selection. In *IJCAI*, 2064–2070.

Fan, Y.; Liu, J.; Tang, J.; Liu, P.; Lin, Y.; and Du, Y. 2024. Learning correlation information for multi-label feature selection. *Pattern Recognition*, 145: 109899.

Fang, S.-G.; Huang, D.; Wang, C.-D.; and Tang, Y. 2023. Joint multi-view unsupervised feature selection and graph learning. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 8(1): 16–31.

Gu, Z.; Feng, S.; Hu, R.; and Lyu, G. 2023. ONION: Joint unsupervised feature selection and robust subspace extraction for graph-based multi-view clustering. *ACM Transactions on Knowledge Discovery from Data*, 17(5): 1–23.

Han, Q.; Hu, L.; and Gao, W. 2024. Feature relevance and redundancy coefficients for multi-view multi-label feature selection. *Information Sciences*, 652: 119747.

Hua, Y.; Song, X.; Feng, Z.; and Wu, X. 2023. MFR-DTA: a multi-functional and robust model for predicting drug–target binding affinity and region. *Bioinformatics*, 39(2): 1–9.

Jiang, B.; Zhang, C.; Zhong, Y.; Liu, Y.; Zhang, Y.; Wu, X.; and Sheng, W. 2023. Adaptive collaborative fusion for multi-view semi-supervised classification. *Information Fusion*, 96: 37–50.

Jiang, G.; Wang, H.; Peng, J.; Chen, D.; and Fu, X. 2021. Graph-based multi-view binary learning for image clustering. *Neurocomputing*, 427: 225–237.

Liu, B.; Li, W.; Xiao, Y.; Chen, X.; Liu, L.; Liu, C.; Wang, K.; and Sun, P. 2023. Multi-view multi-label learning with high-order label correlation. *Information Sciences*, 624: 165–184.

Shi, D.; Zhu, L.; Li, J.; Cheng, Z.; and Liu, Z. 2021. Binary label learning for semi-supervised feature selection. *IEEE Transactions on Knowledge and Data Engineering*, 35(3): 2299–2312.

Shi, D.; Zhu, L.; Li, J.; Zhang, Z.; and Chang, X. 2023. Unsupervised adaptive feature selection with binary hashing. *IEEE Transactions on Image Processing*, 32: 838–853.

Tang, C.; Zheng, X.; Zhang, W.; Liu, X.; Zhu, X.; and Zhu, E. 2023. Unsupervised feature selection via multiple graph fusion and feature weight learning. *Science China Information Sciences*, 66(5): 152101.

Wang, H.; Yao, M.; Jiang, G.; Mi, Z.; and Fu, X. 2023. Graph-collaborated auto-encoder hashing for multiview binary clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 1–12.

Wang, Z.; Feng, Y.; Qi, T.; Yang, X.; and Zhang, J. J. 2016. Adaptive multi-view feature selection for human motion retrieval. *Signal Processing*, 120: 691–701.

Xiang, S.-J.; Li, H.-C.; Yang, J.-H.; and Feng, X.-R. 2024. Dual auto-weighted multi-view clustering via autoencoder-like nonnegative matrix factorization. *Information Sciences*, 667: 120458.

Yang, B.; Wu, J.; Zhang, X.; Zheng, X.; Nie, F.; and Chen, B. 2024a. Discrete correntropy-based multi-view anchor-graph clustering. *Information Fusion*, 103: 102097.

Yang, X.; Che, H.; Leung, M.-F.; and Wen, S. 2024b. Self-paced regularized adaptive multi-view unsupervised feature selection. *Neural Networks*, 106295.

Yu, Z.; Tiwari, P.; Hou, L.; Li, L.; Li, W.; Jiang, L.; and Ning, X. 2024. MV-ReID: 3D Multi-view Transformation Network for Occluded Person Re-Identification. *Knowledge-Based Systems*, 283: 111200.

Yuan, H.; Li, J.; Lai, L. L.; and Tang, Y. Y. 2019. Joint sparse matrix regression and nonnegative spectral analysis for two-dimensional unsupervised feature selection. *Pattern Recognition*, 89: 119–133.

Zhang, C.; Jiang, B.; Wang, Z.; Yang, J.; Lu, Y.; Wu, X.; and Sheng, W. 2023. Efficient multi-view semi-supervised feature selection. *Information Sciences*, 649: 119675.

Zhang, R.; Li, X.; Zhang, H.; and Nie, F. 2019. Deep fuzzy k-means with adaptive loss and entropy regularization. *IEEE Transactions on Fuzzy Systems*, 28(11): 2814–2824.

Zhang, Y.; Wang, X.; Jiang, X.; and Zhou, Y. 2022. Robust dual graph self-representation for unsupervised hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–13.

Zhao, M.; Yang, W.; and Nie, F. 2023. Auto-weighted orthogonal and nonnegative graph reconstruction for multi-view clustering. *Information Sciences*, 632: 324–339.