Unified View Imputation and Feature Selection Learning for Incomplete Multi-view Data

Yanyong Huang[1] , **Zongxin Shen**[1*] , **Tianrui Li**[2] and **Fengmao Lv**[2*]

[1]Joint Laboratory of Data Science and Business Intelligence, School of Statistics, Southwestern University of Finance and Economics, Chengdu, China

[2]School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu, China

huangyy@swufe.edu.cn, shenzx@smail.swufe.edu.cn, trli@swjtu.edu.cn, fengmaolv@126.com

Abstract

Although multi-view unsupervised feature selection (MUFS) is an effective technology for reducing dimensionality in machine learning, existing methods cannot directly deal with incomplete multi-view data where some samples are missing in certain views. These methods should first apply predetermined values to impute missing data, then perform feature selection on the complete dataset. Separating imputation and feature selection processes fails to capitalize on the potential synergy where local structural information gleaned from feature selection could guide the imputation, thereby improving the feature selection performance in turn. Additionally, previous methods only focus on leveraging samples' local structure information, while ignoring the intrinsic locality of the feature space. To tackle these problems, a novel MUFS method, called UNified view Imputation and Feature selectIon lEaRning (UNIFIER), is proposed. UNIFIER explores the local structure of multi-view data by adaptively learning similarity-induced graphs from both the sample and feature spaces. Then, UNIFIER dynamically recovers the missing views, guided by the sample and feature similarity graphs during the feature selection procedure. Furthermore, the half-quadratic minimization technique is used to automatically weight different instances, alleviating the impact of outliers and unreliable restored data. Comprehensive experimental results demonstrate that UNIFIER outperforms other state-of-the-art methods.

1 Introduction

Multi-view data describes the same sample from different perspectives or forms, often represented by high-dimensional features [Liu *et al.*, 2018; Zhang *et al.*, 2021]. In practical applications, acquiring labels for multi-view data is typically difficult due to the considerable amount of time and effort required. High-dimensional, unlabeled multi-view data presents challenges for machine learning applications, such as the curse of dimensionality, performance degradation in downstream tasks, and high computational costs. The recently developed dimensionality reduction technique, multi-view unsupervised feature selection (MUFS), aims to tackle these difficulties by selecting informative features from multi-view data in an unsupervised manner [Zhang *et al.*, 2019; Shi *et al.*, 2023].

The existing MUFS approaches can generally be divided into two categories. The first category of methods sequentially connects features from multiple views into a vector, and subsequently employs single-view feature selection methods such as LPscore [He *et al.*, 2005], EGCFS [Zhang *et al.*, 2020], and FSDK [Nie *et al.*, 2023]. The second category of methods directly selects features from multi-view data by considering the correlations among different views. Typically, CvLP-DCL [Tang *et al.*, 2021] learns view-specific and view-common label spaces and constructs a cross-view similarity graph to capture consensus and diversity information from multiple views. TLR-MFS [Yuan *et al.*, 2022] enforces a tensor low-rank constraint on the similarity graph matrix to leverage high-order consensus information among different views in the feature selection procedure. [Cao and Xie, 2024] fully utilizes neighbor information of differing views to construct a similarity graph for feature selection. However, previous studies assumed that each sample is present across all views, which may not hold in real-world scenarios. For example, in Alzheimer's disease detection, many patients only have features from magnetic resonance imaging (MRI) and lack positron emission tomography (PET) scans due to the high cost of PET imaging [Cai *et al.*, 2018]. There are few studies on unsupervised feature selection in incomplete multi-view data. [Xu *et al.*, 2021] proposed a cross-view feature selection method for incomplete multi-view data. This method first fills missing samples with mean values and then uses weighted non-negative matrix factorization on the imputed data to select features. The previously mentioned MUFS methods, when applied to incomplete multi-view data, consist of two sequential stages: first, the imputation of missing data, and then the subsequent selection of features from the imputed multi-view data.

However, the aforementioned methods miss out on the potential synergy by treating imputation and feature selection as two separate processes. Local structural information gleaned from feature selection could guide the imputation

process and, conversely, the refined imputation could lead to improved feature selection performance. Furthermore, although these methods have leveraged samples' local structure information-indicating that samples close to each other in high-dimensional space also remain close in low-dimensional space, they have overlooked the exploration of the intrinsic locality of the feature space. Specifically, the capacity for features to be sparsely represented by similar features is a crucial aspect that could enhance feature selection performance.

To address the aforementioned issues, we propose a novel MUFS method named UNified view Imputation and Feature selectIon lEaRning (UNIFIER), which consists of three key components: 1) An adaptive dual-graph learning module that selects discriminative features by leveraging the local structures of both the feature and sample spaces. 2) A bi-level cooperative missing view completion module that effectively utilizes local structure graphs at both the feature and sample levels to impute missing views in samples. 3) A dynamic sample quality assessment module that automatically weighs samples to mitigate the effects of outliers and unreliable restored data. Fig. 1 illustrates the framework of the proposed method UNIFIER. The main contributions of this paper are as follows:

- To the best of our knowledge, this is the first work to integrate multi-view unsupervised feature selection and missing view imputation into a unified learning framework, enabling a seamless synergy between these two processes that culminates in enhanced feature selection performance.

- Simultaneous investigation of the local structures in both the feature and sample spaces is conducted to facilitate missing view imputation and discriminative feature selection. Additionally, a dynamic sample quality assessment based on half-quadratic minimization is proposed, which can alleviate the impact of outliers and unreliable restored data.

- An efficient alternative iterative algorithm is developed to solve the proposed UNIFIER method, and comprehensive experimental results demonstrate the superiority of UNIFIER over several state-of-the-art (SOTA) methods.

2 Proposed Method

We first summarize some notations throughout this paper. For any matrix $\mathbf{M} = (m_{ij}) \in \mathbb{R}^{p \times q}$, the i-th row and the j-th column of $\mathbf{M}$ are denoted as $\mathbf{m}_{i \cdot}$ and $\mathbf{m}_{\cdot j}$, respectively. Its Frobenius norm and $\ell_{2,1}$-norm are respectively defined as $\|\mathbf{M}\|_F = \sqrt{\sum_{i=1}^p \sum_{j=1}^q m_{ij}^2}$ and $\|\mathbf{M}\|_{2,1} = \sum_{i=1}^p \sqrt{\sum_{j=1}^q m_{ij}^2}$. $\mathrm{Tr}(\mathbf{M})$ and $\mathbf{M}^\top$ represent the trace and transpose of $\mathbf{M}$, respectively. In the following, we will introduce three modules of the proposed method UNIFIER in detail.

Adaptive Dual-graph Learning for MUFS. Given a multi-view dataset $\mathcal{X} = \{\mathbf{X}^{(v)} \in \mathbb{R}^{n \times d_v}\}_{v=1}^V$, where $\mathbf{X}^{(v)}$
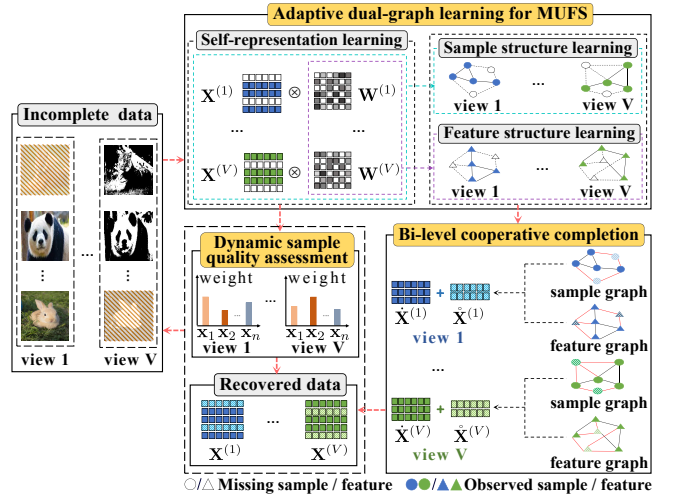

Figure 1. The framework of the proposed method UNIFIER.

represents the feature matrix of the v-th view, n and d_v denote the number of samples and dimensions of $\mathbf{X}^{(v)}$, and V is the number of views. In order to select discriminative features from the given multi-view dataset, a comprehensive exploration of the local structures within both the feature and sample spaces is conducted. By extending the concept of data self-representation [Liu *et al.*, 2021b] to the feature level, it is reasonable to assume that each feature can be reconstructed by a linear combination of a small set of representative features. This can be formalized as follows:

$$\min_{\mathbf{W}^{(v)}} \sum_{v=1}^V \|\mathbf{X}^{(v)} - \mathbf{X}^{(v)}\mathbf{W}^{(v)}\|_F^2 + \lambda\|\mathbf{W}^{(v)}\|_{2,1}, \quad (1)$$

where $\mathbf{W}^{(v)} \in \mathbb{R}^{d_v \times d_v}$ is the sparse representation graph or feature weight matrix in the v-th view. The i-th row of $\mathbf{W}^v$ indicates the contribution of the i-th feature $\mathbf{x}_{\cdot i}^{(v)}$ to the reconstruction of all other features, thereby reflecting the importance of the i-th feature in $\mathbf{X}^v$. Moreover, we employ the $\ell_{2,1}$-norm regularization $\|\mathbf{W}^{(v)}\|_{2,1}$ to induce sparsity within $\mathbf{W}^v$, which encourages the weights of less significant features to approach zero. Consequently, each feature is selectively associated with only the most representative features, effectively capturing the local structure of the feature space.

Furthermore, according to the spectral graph theory [Dong *et al.*, 2012], if two samples $\mathbf{x}_{i \cdot}^{(v)}$ and $\mathbf{x}_{j \cdot}^{(v)}$ are similar in high-dimensional space, they will also be similar in low-dimensional space. This allows us to preserve the local geometric structure of the sample space by adaptively learning the nearest neighborhood graph, which can be formulated as follows:

$$\min_{\mathbf{W}^{(v)}, \mathbf{S}^{(v)}} \sum_{v=1}^V \sum_{i,j=1}^n \|\mathbf{x}_{i \cdot}^{(v)}\mathbf{W}^{(v)} - \mathbf{x}_{j \cdot}^{(v)}\mathbf{W}^{(v)}\|_2^2 s_{ij}^{(v)} + \xi_v\|\mathbf{S}^{(v)}\|_F^2$$

$$\text{s.t. } s_{ii}^{(v)} = 0, s_{ij}^{(v)} \geq 0, \mathbf{1}^\top \mathbf{s}_{\cdot i}^{(v)} = 1, \|\mathbf{s}_{\cdot i}^{(v)}\|_0 = k, \quad (2)$$

where $\mathbf{S}^{(v)}$ is the similarity matrix of the v-th view and ξ_v is a regularization parameter. The constraint $\|\mathbf{s}_{\cdot i}^{(v)}\|_0 = k$ is

enforced to ensure that $\mathbf{x}_{i\cdot}^{(v)}$ is exclusively connected to its k nearest neighbors.

By combining Eqs. (1) and (2), we can obtain the following adaptive dual-graph learning-based MUFS module:

$$\min_{\mathbf{W}^{(v)},\mathbf{S}^{(v)}} \sum_{v=1}^{V} \Big[\alpha^{(v)} \|\mathbf{X}^{(v)} - \mathbf{X}^{(v)}\mathbf{W}^{(v)}\|_F^2 + \lambda\|\mathbf{W}^{(v)}\|_{2,1}$$

$$+ \frac{1}{2}\sum_{i,j=1}^{n} \|\mathbf{x}_{i\cdot}^{(v)}\mathbf{W}^{(v)} - \mathbf{x}_{j\cdot}^{(v)}\mathbf{W}^{(v)}\|_2^2 s_{ij}^{(v)} + \xi_v\|\mathbf{S}^{(v)}\|_F^2 \Big]$$

$$\text{s.t. } s_{ii}^{(v)} = 0, s_{ij}^{(v)} \geq 0, \mathbf{1}^\top\mathbf{s}_{\cdot i}^{(v)} = 1, \|\mathbf{s}_{\cdot i}^{(v)}\|_0 = k,$$

$$\tag{3}$$

where $\alpha^{(v)}$ is the view weight. Eq. (3) can simultaneously use the local structure information of both the feature and sample spaces to improve the feature selection performance.

Bi-level Cooperative Missing View Completion. In the context of an incomplete multi-view scenario, where some samples are present in certain views but absent in others, we use $\mathring{\mathbf{X}}^{(v)} \in \mathbb{R}^{m \times d_v}$ and $\dot{\mathbf{X}}^{(v)} \in \mathbb{R}^{(n-m) \times d_v}$ to represent the feature matrices of missing samples and available samples in $\mathbf{X}^{(v)}$, respectively. Unlike these two-step methods, which sequentially address missing value imputation followed by feature selection, we integrate feature selection and missing view imputation into a unified optimization process. To this end, we incorporate the missing data as a variable into the learning process described by Eq. (3), which will undergo alternate optimization with other variables until convergence. The corresponding objective function is presented below.

$$\min_{\mathbf{W}^{(v)},\mathring{\mathbf{X}}^{(v)},\mathbf{S}^{(v)}} \sum_{v=1}^{V} \Big[\alpha^{(v)} \|\widetilde{\mathbf{X}}^{(v)} - \widetilde{\mathbf{X}}^{(v)}\mathbf{W}^{(v)}\|_F^2 + \lambda\|\mathbf{W}^{(v)}\|_{2,1}$$

$$+ \frac{1}{2}\sum_{i,j=1}^{n} \|\widetilde{\mathbf{x}}_{i\cdot}^{(v)}\mathbf{W}^{(v)} - \widetilde{\mathbf{x}}_{j\cdot}^{(v)}\mathbf{W}^{(v)}\|_2^2 s_{ij}^{(v)} + \xi_v\|\mathbf{S}^{(v)}\|_F^2 \Big]$$

$$\text{s.t. } \widetilde{\mathbf{X}}^{(v)} = \dot{\mathbf{K}}^{(v)}\dot{\mathbf{X}}^{(v)} + \mathring{\mathbf{K}}^{(v)}\mathring{\mathbf{X}}^{(v)}, s_{ii}^{(v)} = 0, s_{ij}^{(v)} \geq 0,$$

$$\mathbf{1}^\top\mathbf{s}_{\cdot i}^{(v)} = 1, \|\mathbf{s}_{\cdot i}^{(v)}\|_0 = k,$$

$$\tag{4}$$

where $\mathring{\mathbf{K}}^{(v)} \in \mathbb{R}^{n \times m}$ and $\dot{\mathbf{K}}^{(v)} \in \mathbb{R}^{n \times (n-m)}$ serve as two indicator matrices, and m denotes the number of missing samples. These matrices are used to project the learned missing samples and available samples into a complete multi-view data $\widetilde{\mathbf{X}}^{(v)}$, and their specific definitions are as follows:

$$\dot{\mathbf{K}}_{ij}^{(v)} = \begin{cases} 1, & \text{if } \mathbf{x}_{i\cdot}^{(v)} \text{ corresponds to } \dot{\mathbf{x}}_{j\cdot}^{(v)} \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathring{\mathbf{K}}_{ij}^{(v)} = \begin{cases} 1, & \text{if } \mathbf{x}_{i\cdot}^{(v)} \text{ corresponds to } \mathring{\mathbf{x}}_{j\cdot}^{(v)} \\ 0, & \text{otherwise.} \end{cases}$$

$$\tag{5}$$

In Eq. (4), the missing data $\mathring{\mathbf{X}}^{(v)}$ is imputed with collaborative guidance from the sample-level and feature-level graphs obtained from Eq. (3). This imputation, in turn, enhances the learning of local both sample and feature structures. These two processes boost each other in an interplay manner to recover incomplete multi-view data and achieve better feature selection performance.

Dynamic Sample Quality Assessment. Due to the sensitivity of the Frobenius norm-based loss function to outliers [Nie *et al.*, 2010], we adopt the Geman-McClure estimator $\mathcal{L}(z) = \frac{\gamma z^2}{\gamma + z^2}$ (γ is the scale parameter) [Barron, 2019] as the loss function in Eq. (4), aiming to reduce the influence of outliers. However, the objective function based on the Geman-McClure loss function is difficult to solve. To cope with this, we replace $\mathcal{L}(z)$ with the following equivalent expression according to the half-quadratic minimization theory [Nikolova and Chan, 2007].

$$\min_z \mathcal{L}(z) = \frac{\gamma z^2}{\gamma + z^2} \iff \min_{z,e}\{ez^2 + \psi(e)\} \tag{6}$$

where e is an auxiliary variable, and $\psi(e) = \gamma(\sqrt{e} - 1)^2$ is the dual potential function of Geman-McClure loss function.

Then, the final objective function of the proposed UNIFIER is formulated as

$$\min_{\Theta}\sum_{v=1}^{V}\Big[\alpha^{(v)}\sum_{i=1}^{n}\Big(e_i^{(v)}\|\widetilde{\mathbf{x}}_{i\cdot}^{(v)} - \widetilde{\mathbf{x}}_{i\cdot}^{(v)}\mathbf{W}^{(v)}\|_2^2 + \gamma_v(\sqrt{e_i^{(v)}} - 1)^2\Big)$$

$$+ \frac{1}{2}\sum_{i,j=1}^{n}\|\widetilde{\mathbf{x}}_{i\cdot}^{(v)}\mathbf{W}^{(v)} - \widetilde{\mathbf{x}}_{j\cdot}^{(v)}\mathbf{W}^{(v)}\|_2^2 s_{ij}^{(v)} + \xi_v\sum_{i=1}^{n}\|\mathbf{s}_{\cdot i}^{(v)}\|_2^2$$

$$+ \lambda\|\mathbf{W}^{(v)}\|_{2,1}\Big]$$

$$\text{s.t. } \widetilde{\mathbf{X}}^{(v)} = \dot{\mathbf{K}}^{(v)}\dot{\mathbf{X}}^{(v)} + \mathring{\mathbf{K}}^{(v)}\mathring{\mathbf{X}}^{(v)}, s_{ii}^{(v)} = 0, s_{ij}^{(v)} \geq 0,$$

$$\mathbf{1}^\top\mathbf{s}_{\cdot i}^{(v)} = 1, \|\mathbf{s}_{\cdot i}^{(v)}\|_0 = k,$$

$$\tag{7}$$

where $\Theta = \{\mathbf{W}^{(v)}, \mathring{\mathbf{X}}^{(v)}, \mathbf{S}^{(v)}, \mathbf{e}^{(v)} | v = 1 \dots n\}$, and $\mathbf{e}^{(v)} = [e_1^{(v)}, e_2^{(v)}, \dots, e_n^{(v)}]^T \in \mathbb{R}^{n \times 1}$.

In Eq. (7), $e_i^{(v)}$ can assess the quality of the i-th sample $\widetilde{\mathbf{x}}_{i\cdot}^{(v)}$ based on the optimization result in Eq. (18). Specifically, if $\widetilde{\mathbf{x}}_{i\cdot}^{(v)}$ is an outlier or an unreliable restored sample, the reconstruction error $\|\widetilde{\mathbf{x}}_{i\cdot}^{(v)} - \widetilde{\mathbf{x}}_{i\cdot}^{(v)}\mathbf{W}^{(v)}\|_2^2$ will be large. As a result, the corresponding weight of $\widetilde{\mathbf{x}}_{i\cdot}^{(v)}$ will be small, and vice versa. Furthermore, $e_i^{(v)}$ is updated adaptively, enabling the automatic assignment of weights to samples of varying qualities.

3 Optimization

Since Eq. (7) is not jointly convex to all four groups of variables, including $\mathbf{W}^{(v)}, \mathring{\mathbf{X}}^{(v)}, \mathbf{S}^{(v)}$ and $\mathbf{e}^{(v)}$, we propose to optimize them alternatively, i.e., by fixing three groups of variables and optimizing the remaining one alternately.

Update $\mathbf{W}^{(v)}$ by Fixing Others. When other variables are fixed, the objective function w.r.t. $\mathbf{W}^{(v)}$ becomes:

$$\min_{\mathbf{W}^{(v)}} \alpha^{(v)}\sum_{i=1}^{n} e_i^{(v)}\|\widetilde{\mathbf{x}}_{i\cdot}^{(v)} - \widetilde{\mathbf{x}}_{i\cdot}^{(v)}\mathbf{W}^{(v)}\|_2^2 + \lambda\|\mathbf{W}^{(v)}\|_{2,1}$$

$$+ \frac{1}{2}\sum_{i,j=1}^{n}\|\widetilde{\mathbf{x}}_{i\cdot}^{(v)}\mathbf{W}^{(v)} - \widetilde{\mathbf{x}}_{j\cdot}^{(v)}\mathbf{W}^{(v)}\|_2^2 s_{ij}^{(v)}$$

$$\tag{8}$$

Algorithm 1 Iterative Algorithm of UNIFIER

Input:Incomplete multi-view data $\{\mathbf{X}^{(v)} \in \mathbb{R}^{n \times d_v}\}_{v=1}^{V}$; the parameters $\alpha^{(v)}$ and λ; the number of selected features h.

1: Initialize $\mathbf{D}^{(v)}, \mathbf{e}^{(v)}, \mathring{\mathbf{X}}^{(v)}$ and $\mathbf{S}^{(v)}$ ($v = 1, \ldots, V$).
2: **while** not convergent **do**
3:    Update $\{\mathbf{W}^{(v)}\}_{v=1}^{V}$ via Eq. (10);
4:    Update $\{\mathbf{D}^{(v)}(i,i) = 1/(2\sqrt{\|\mathbf{w}_{i\cdot}^{(v)}\|_2^2 + \epsilon})\}_{v=1}^{V}$;
5:    Update $\{\mathring{\mathbf{X}}^{(v)}\}_{v=1}^{V}$ by solving Eq. (13);
6:    Update $\{\mathbf{S}^{(v)}\}_{v=1}^{V}$ via Eq. (16);
7:    Update $\{\mathbf{e}^{(v)}\}_{v=1}^{V}$ via Eq. (18);
8: **end while**

Output:Sorting the ℓ_2-norm of the rows of $\{\mathbf{W}^{(v)}\}_{v=1}^{V}$ in descending order and selecting the top h features from $\mathbf{X}^{(v)}$.

By using the matrix trace property, we can optimize problem (8) by solving its equivalent form:

$$\min_{\mathbf{W}^{(v)}} \alpha^{(v)}\|\mathbf{E}^{(v)}(\widetilde{\mathbf{X}}^{(v)} - \widetilde{\mathbf{X}}^{(v)}\mathbf{W}^{(v)})\|_F^2 + \lambda\mathrm{Tr}(\mathbf{W}^{(v)^\top}\mathbf{D}^{(v)}\mathbf{W}^{(v)})$$
$$+ \mathrm{Tr}\left(\mathbf{W}^{(v)^\top}\widetilde{\mathbf{X}}^{(v)^\top}\mathbf{L}^{(v)}\widetilde{\mathbf{X}}^{(v)}\mathbf{W}^{(v)}\right)$$
(9)

where $\mathbf{D}^{(v)}$ is a diagonal matrix with its i-th diagonal entry given by $\mathbf{D}^{(v)}(i,i) = 1/2\sqrt{\|\mathbf{w}_{i\cdot}^{(v)}\|_2^2 + \epsilon}$ (ϵ is a small constant to prevent the denominator from vanishing, ϵ is set to 10^{-20} in the experiment), $\mathbf{L}^{(v)}$ is the Laplacian matrix of $\mathbf{S}^{(v)}$, and $\mathbf{E}^{(v)} = diag(\sqrt{\mathbf{e}^{(v)}})$.

By taking the derivative of Eq. (9) w.r.t. $\mathbf{W}^{(v)}$ and setting it to zero, we can obtain the solution of $\mathbf{W}^{(v)}$ as follows:

$$\mathbf{W}^{(v)} = \alpha^{(v)}\left(\mathbf{C}^{(v)} + \lambda\mathbf{D}^{(v)}\right)^{-1}\widetilde{\mathbf{X}}^{(v)^\top}\mathbf{E}^{(v)}\mathbf{E}^{(v)^\top}\widetilde{\mathbf{X}}^{(v)}$$
(10)

where $\mathbf{C}^{(v)} = \alpha^{(v)}\widetilde{\mathbf{X}}^{(v)^\top}\mathbf{E}^{(v)}\mathbf{E}^{(v)^\top}\widetilde{\mathbf{X}}^{(v)} + \widetilde{\mathbf{X}}^{(v)^\top}\mathbf{L}^{(v)}\widetilde{\mathbf{X}}^{(v)}$.

Update $\mathring{\mathbf{X}}^{(v)}$ by Fixing Others. After fixing the other variables, the objective function w.r.t. $\mathring{\mathbf{X}}^{(v)}$ is reduced to:

$$\min_{\mathring{\mathbf{X}}^{(v)}} \alpha^{(v)} \sum_{i=1}^{n} e_i^{(v)}\|\widetilde{\mathbf{x}}_{i\cdot}^{(v)} - \widetilde{\mathbf{x}}_{i\cdot}^{(v)}\mathbf{W}^{(v)}\|_2^2$$
$$+ \frac{1}{2}\sum_{i,j=1}^{n}\|\widetilde{\mathbf{x}}_{i\cdot}^{(v)}\mathbf{W}^{(v)} - \widetilde{\mathbf{x}}_{j\cdot}^{(v)}\mathbf{W}^{(v)}\|_2^2 s_{ij}^{(v)}$$
$$\text{s.t. } \widetilde{\mathbf{X}}^{(v)} = \dot{\mathbf{K}}^{(v)}\dot{\mathbf{X}}^{(v)} + \mathring{\mathbf{K}}^{(v)}\mathring{\mathbf{X}}^{(v)}$$
(11)

Incorporating the equality constraint $\widetilde{\mathbf{X}}^{(v)} = \dot{\mathbf{K}}^{(v)}\dot{\mathbf{X}}^{(v)} + \mathring{\mathbf{K}}^{(v)}\mathring{\mathbf{X}}^{(v)}$ into the objective function and removing the irrelevant terms w.r.t $\mathring{\mathbf{X}}^{(v)}$, we can reformulate problem (11) as follows:

$$\min_{\mathring{\mathbf{X}}^{(v)}}\mathrm{Tr}\{\alpha^{(v)}[2\mathbf{R}^{(v)^\top}\mathring{\mathbf{X}}^{(v)}(\mathbf{I}+\mathbf{W}^{(v)}\mathbf{W}^{(v)^\top}-\mathbf{W}^{(v)}-\mathbf{W}^{(v)^\top})$$
$$+\mathring{\mathbf{X}}^{(v)^\top}\mathring{\mathbf{K}}^{(v)^\top}\mathbf{H}^{(v)}\mathring{\mathbf{K}}^{(v)}\mathring{\mathbf{X}}^{(v)}(\mathbf{I}+\mathbf{W}^{(v)}\mathbf{W}^{(v)^\top}-2\mathbf{W}^{(v)})]+$$
$$\mathbf{W}^{(v)^\top}(2\dot{\mathbf{X}}^{(v)^\top}\dot{\mathbf{K}}^{(v)^\top}+\mathring{\mathbf{X}}^{(v)^\top}\mathring{\mathbf{K}}^{(v)^\top})\mathbf{L}^{(v)}\dot{\mathbf{K}}^{(v)}\mathring{\mathbf{X}}^{(v)}\mathbf{W}^{(v)}\},$$
(12)

where $\mathbf{H}^{(v)} = \mathbf{E}^{(v)}\mathbf{E}^{(v)^\top}$, $\mathbf{R}^{(v)} = \mathring{\mathbf{K}}^{(v)^\top}\mathbf{H}^{(v)}\dot{\mathbf{K}}^{(v)}\dot{\mathbf{X}}^{(v)}$.

Taking the derivative of Eq. (12) w.r.t. $\mathring{\mathbf{X}}^{(v)}$ and setting it to zero, we can obtain:

$$\mathbf{A}^{(v)}\mathring{\mathbf{X}}^{(v)}\mathbf{W}^{(v)}\mathbf{W}^{(v)^\top} + \mathbf{P}^{(v)}\mathring{\mathbf{X}}^{(v)}\mathbf{Q}^{(v)} = \mathbf{F}^{(v)}, \quad (13)$$

where $\mathbf{A}^{(v)} = \mathring{\mathbf{K}}^{(v)^\top}\mathbf{L}^{(v)}\mathring{\mathbf{K}}^{(v)}$, $\mathbf{Q}^{(v)} = \mathbf{I} - \mathbf{W}^{(v)} - \mathbf{W}^{(v)^\top} + \mathbf{W}^{(v)}\mathbf{W}^{(v)^\top}$, $\mathbf{P}^{(v)} = \alpha^{(v)}\mathring{\mathbf{K}}^{(v)^\top}\mathbf{H}^{(v)}\mathring{\mathbf{K}}^{(v)}$, $\mathbf{F}^{(v)} = -\alpha^{(v)}\mathbf{R}^{(v)}\mathbf{Q}^{(v)} - \mathring{\mathbf{K}}^{(v)^\top}\mathbf{L}^{(v)}\dot{\mathbf{K}}^{(v)}\dot{\mathbf{X}}^{(v)}\mathbf{W}^{(v)}\mathbf{W}^{(v)^\top}$. Problem (13) is a generalized Sylvester matrix equation, which can be solved according to the method proposed in [Zhang and Yin, 2017].

Update $\mathbf{S}^{(v)}$ by Fixing Others. By fixing the other variables, the optimization problem w.r.t $\mathbf{S}^{(v)}$ in Eq. (7) becomes independent for different i, allowing us to optimize $\mathbf{S}^{(v)}$ by separately solving each $\mathbf{s}_{\cdot i}^{(v)}$ as follows:

$$\min_{\mathbf{s}_{\cdot i}^{(v)}} \frac{1}{2}\sum_{j=1}^{n}\|\widetilde{\mathbf{x}}_{i\cdot}^{(v)}\mathbf{W}^{(v)} - \widetilde{\mathbf{x}}_{j\cdot}^{(v)}\mathbf{W}^{(v)}\|_2^2 s_{ij}^{(v)} + \xi_v\left\|\mathbf{s}_{\cdot i}^{(v)}\right\|_2^2$$
$$\text{s.t. } s_{ii}^{(v)} = 0, s_{ij}^{(v)} \geq 0, \mathbf{1}^\top\mathbf{s}_{\cdot i}^{(v)} = 1, \|\mathbf{s}_{\cdot i}^{(v)}\|_0 = k,$$
(14)

By defining a vector $\mathbf{b}_i^{(v)}$ with the j-th entry as $b_{ij}^{(v)} = \frac{1}{2}\|\widetilde{\mathbf{x}}_{i\cdot}^{(v)}\mathbf{W}^{(v)} - \widetilde{\mathbf{x}}_{j\cdot}^{(v)}\mathbf{W}^{(v)}\|_2^2$, problem (14) can be transformed into:

$$\min_{\mathbf{s}_{\cdot i}^{(v)}} \frac{1}{2}\left\|\mathbf{s}_{\cdot i}^{(v)} + \mathbf{b}_i^{(v)}/2\xi_v\right\|_2^2$$
$$\text{s.t. } s_{ii}^{(v)} = 0, s_{ij}^{(v)} \geq 0, \mathbf{1}^\top\mathbf{s}_{\cdot i}^{(v)} = 1, \|\mathbf{s}_{\cdot i}^{(v)}\|_0 = k.$$
(15)

Then, following the same derivation process in [Nie *et al.*, 2019], we can get the optimal solution of $s_{ij}^{(v)}$ as follows:

$$s_{ij}^{(v)} = \begin{cases} \frac{b_{i,k+1}^{(v)} - b_{ij}^{(v)}}{kb_{i,k+1}^{(v)} - \sum_{t=1}^{k}b_{it}^{(v)}} & j \leq k \\ 0 & j > k \end{cases}$$
(16)

Besides, the regularization parameter ξ_v is determined as $(kb_{i,k+1}^{(v)} - \sum_{t=1}^{k}b_{it}^{(v)})/2$ to ensure that $\mathbf{s}_{\cdot i}^{(v)}$ only has k nonzero entries.

Update $\mathbf{e}^{(v)}$ by Fixing Others. When other variables are fixed, the objective function w.r.t. $\mathbf{e}^{(v)}$ becomes:

$$\min_{\mathbf{e}^{(v)}} \sum_{i=1}^{n}\left(e_i^{(v)}\|\widetilde{\mathbf{x}}_{i\cdot}^{(v)} - \widetilde{\mathbf{x}}_{i\cdot}^{(v)}\mathbf{W}^{(v)}\|_2^2 + \gamma_v(\sqrt{e_i^{(v)}} - 1)^2\right)$$
(17)

By taking the derivative of Eq. (17) w.r.t. $e_i^{(v)}$ and setting it to zero, we can get the solution of $e_i^{(v)}$ as follows:

$$e_i^{(v)} = \left(\frac{\gamma_v}{\gamma_v + \|\widetilde{\mathbf{x}}_{i\cdot}^{(v)} - \widetilde{\mathbf{x}}_{i\cdot}^{(v)}\mathbf{W}^{(v)}\|_2^2}\right)^2$$
(18)

Algorithm 1 summarizes the detailed steps to solve Eq. (7). In Algorithm 1, $\mathbf{D}^{(v)}$ is initialized as an identity matrix, $e_i^{(v)}$ is set as $1/n$ for all views, $\{\mathring{\mathbf{X}}^{(v)}\}_{v=1}^{l}$ is initialized by the mean values of $\dot{\mathbf{X}}^{(v)}$, and $\mathbf{S}^{(v)}$ is initialized by constructing the k-nearest neighbor graph according to [Li *et al.*, 2019].

4 Algorithm Analysis

Time complexity analysis. In Algorithm 1, the updates for $\mathbf{W}^{(v)}$, $\mathring{\mathbf{X}}^{(v)}$, $\mathbf{S}^{(v)}$ and $\mathbf{e}^v$ are performed alternately. In each iteration, updating $\mathbf{W}^{(v)}$ costs $\mathcal{O}(\max(n, d_v)nd_v + d_v^3)$ and updating $\mathring{\mathbf{X}}^{(v)}$ costs $\mathcal{O}((\max(m, d_v)md_v)t)$, where t denotes the iteration number of conjugate gradient descent. The updates of $\mathbf{S}^{(v)}$ and $\mathbf{e}^v$ only involve element-based operations, thus their computational cost can be ignored. In summary, the time complexity of each iteration in Algorithm 1 is $\mathcal{O}(\sum_{v=1}^V [(\max(n, d_v)nd_v + (\max(m, d_v)md_v t + d_v^3)])$.

Convergence analysis. Following the similar derivation in [Hou *et al.*, 2013], it is easy to demonstrate that updating $\mathbf{W}^{(v)}$ by solving problem (9) leads to a monotonic decrease in the objective value of Eq. (7). Besides, the convergence of optimizing $\mathring{\mathbf{X}}^{(v)}$ can be guaranteed according to [Zhang and Yin, 2017]. Moreover, the closed-form solutions of Eqs. (16) and (18) demonstrate the convergence of the updates for $\mathbf{S}^{(v)}$ and $\mathbf{e}^{(v)}$. Thus, the objective function value of Eq. (7) will monotonically decrease in each iteration until convergence.

5 Experiments

Datasets. We evaluate the performance of the proposed UNIFIER on six real-world multi-view datasets, including two text datasets: BBCSport [Wen *et al.*, 2019] and B-BC4views [Chen *et al.*, 2021]; a face image dataset: Yale [Liu *et al.*, 2021a]; two object image datasets: Aloi [Rocha and Goldenstein, 2013] and Caltech101-20 [Huang *et al.*, 2019]; and a handwritten digit image dataset: USPS2View [Liu *et al.*, 2016]. A detailed description of the datasets is summarized in Table 1. To simulate the incomplete multi-view setting, we follow the method in [Lin *et al.*, 2022] by randomly removing $\lfloor n \times r \rfloor$ instances from each view, with r representing the missing data ratio and $\lfloor \cdot \rfloor$ denoting the round-down operator. In the experiment, we vary the value of r over the range of $\{0.1, 0.15, 0.2, 0.25, 0.3\}$.

Comparison methods. We compare UNIFIER with the following SOTA methods:

1) **AllFea** utilizes all the original features.

2) **LPscore** [He *et al.*, 2005] measures the importance of each feature by considering the locality-preserving power.

3) **EGCFS** [Zhang *et al.*, 2020] embeds between-class scatter matrix maximization into adaptive graph learning.

4) **NSGL** [Bai *et al.*, 2020] simultaneously learns the similarity graph and pseudo labels to select features.

5) **HMUFS** [Shen *et al.*, 2020] utilizes an indicator matrix to exclude unobserved data from feature selection.

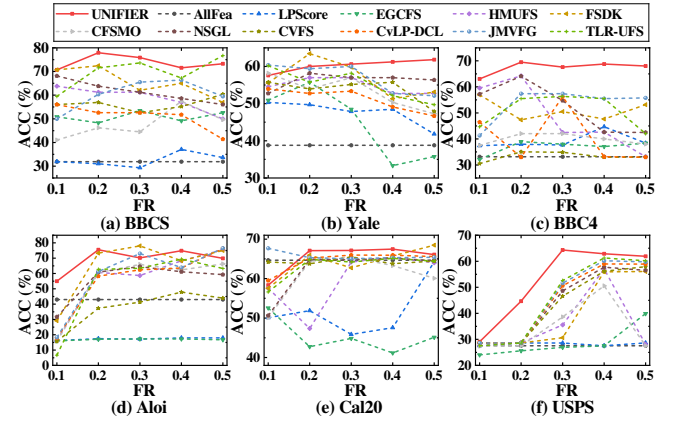| Datasets	Abbr.	Views	Instances	Features	Classes
BBCSport	BBCS	4	116	1991/2063/2113/2158	5
Yale	Yale	3	165	4096/3304/6750	15
BBC4views	BBC4	4	685	4659/4633/4665/4684	5
Aloi	Aloi	4	1869	77/13/64/64	17
Caltech101-20	Cal20	6	2386	48/40/254/1984/512/928	20
USPS2View	USPS	2	5427	256/32	5

Table 1. Dataset description

Figure 2. ACC of different methods on six datasets under different feature selection ratios.
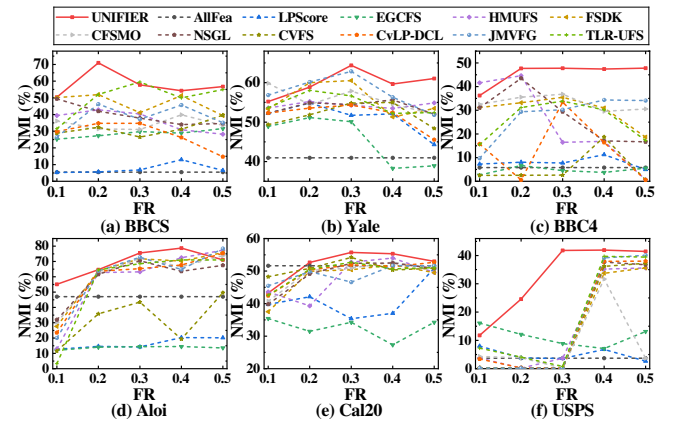

Figure 3. NMI of different methods on six datasets under different feature selection ratios.

6) **FSDK** [Nie *et al.*, 2023] proposes a fast sparse discriminative K-means method for feature selection.

7) **CVFS** [Xu *et al.*, 2021] combines Hilbert-Schmidt Independence Criterion to perform cross-view feature selection.

8) **CvLP-DCL** [Tang *et al.*, 2021] learns cross-view similarity graphs for feature selection.

9) **TLR-MFS** [Yuan *et al.*, 2022] uses tensor low-rank regularization to learn similarity graphs and select features.

10) **JMVFG** [Fang *et al.*, 2023] integrates feature selection and multi-view graph learning into a unified framework.

11) **CFSMO** [Cao and Xie, 2024] introduces multi-order similarity learning for the selection of relevant features.

Since most comparison methods cannot be directly utilized with incomplete data, we fill the missing data with the mean values of features before employing these methods. To ensure a fair comparison, we tune the parameters of all methods using a grid search strategy and report the best performance. The parameters $\alpha^{(v)}$ and λ of our method are set within the range of $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$. To simplify, the weight parameters for each view are set to be the same and the scale parameter γ_v of Geman-McClure function is set to 1. As determining the optimal number of selected

Methods	BBCS		Yale		BBC4		Aloi		Cal20		USPS	
	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
UNIFIER	**71.55**	**54.26**	**61.21**	**59.66**	**68.76**	**47.45**	**74.91**	**78.74**	**67.48**	**55.32**	**62.87**	**41.93**
AllFea	31.90	5.47	38.79	40.96	33.14	5.80	43.02	47.01	64.67	51.63	27.55	3.69
LPscore	37.07	12.96	48.48	52.09	44.67	11.24	17.98	20.31	47.57	37.00	27.31	6.79
EGCFS	49.14	28.86	33.33	38.27	37.08	3.73	17.28	14.43	41.16	27.33	27.40	7.12
HMUFS	56.90	30.89	52.73	53.48	42.63	16.88	68.33	72.50	65.47	53.99	57.01	35.20
FSDK	65.00	51.39	51.30	51.31	47.63	31.02	68.35	70.27	65.72	51.92	55.92	33.72
NSGL	58.62	34.00	56.97	55.64	42.63	17.06	61.16	63.59	65.21	52.45	57.71	37.68
CVFS	55.17	31.12	55.76	54.98	32.99	18.75	47.87	19.33	63.79	50.34	56.13	36.14
CvLP-DCL	51.72	26.24	49.09	52.39	33.14	16.39	63.62	67.66	65.93	51.36	58.95	37.96
TLR-MFS	67.24	49.97	52.12	52.21	55.47	30.16	69.40	71.00	65.13	50.67	61.27	39.69
JMVFG	66.38	45.67	52.73	56.31	55.47	34.36	63.90	64.84	65.30	51.65	60.20	39.17
CFSMO	57.41	39.82	50.30	54.57	39.99	29.55	62.38	66.57	63.29	52.73	50.55	31.69

Table 2. Means (%) of ACC and NMI of different methods on six datasets with missing ratio 0.2 while selecting 40% of all features.
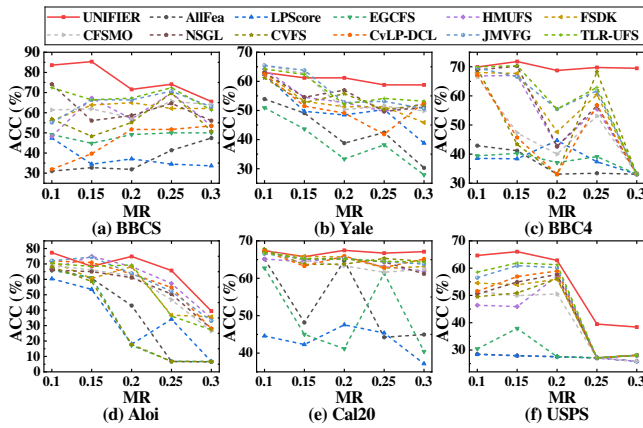

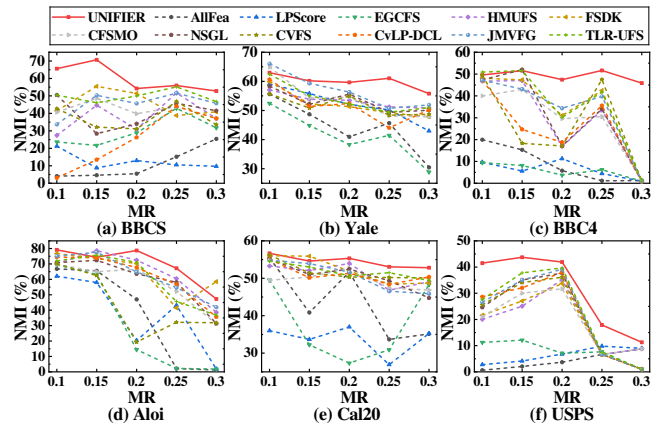Figure 4. ACC of different methods on six datasets under different missing rate.

Figure 5. NMI of different methods on six datasets under different missing rate.

features is still a challenging problem, we vary the proportion of selected features from $\{0.1, 0.2, 0.3, 0.4, 0.5\}$. We adopt a commonly used approach to assess MUFS [Tang *et al.*, 2021; Zhang *et al.*, 2019], employing clustering performance to evaluate the quality of selected features. In this paper, we run the graph-based multi-view clustering algorithm (GMC) [Wang *et al.*, 2019] 30 times on the selected features and use two widely recognized metrics, namely clustering accuracy (ACC) and normalized mutual information (NMI), to measure performance. As GMC is not sensitive to initialization, we only present the average results and omit the standard deviation.

Performance Comparison Table 2 summarizes the performance of UNIFIER and other compared methods on six benchmark datasets, where we achieve the best ACC and NMI in all cases. As can be seen, on two text datasets BBCS and BBC4, UNIFIER outperforms other competitors with average improvements of 21.91% and 24.89% in terms of ACC and NMI, respectively. For two object image datasets Aloi and Cal20, UNIFIER achieves average improvements of 14.08% and 16.64% in ACC and NMI, respectively. And in terms of ACC and NMI, the average improvements of U-

NIFIER on the handwritten image dataset USPS are 13.78% and 13.85%, respectively. Additionally, for the face image dataset Yale, UNIFIER continues to outperform other compared methods, respectively showing average improvements of 11.97% and 8.55% in ACC and NMI.

Furthermore, to comprehensively validate the effectiveness of UNIFIER, we also present the results of all methods across various feature selection ratios and missing ratios. In Figs. 2 and 3, the x-axis (FR) denotes the feature selection ratio, while the y-axis represents the ACC and NMI values of comparative methods, respectively, with the missing ratio fixed at 0.2. It can be observed that UNIFIER yields the best performance in most cases compared with other methods when feature selection ratio varies from 0.1 to 0.5. Additionally, in Figs. 4 and 5, the x-axis (MR) indicates the missing ratio, while the y-axis shows the ACC and NMI results of the comparison methods, respectively, with a fixed feature selection ratio of 0.4. As observed, UNIFIER consistently outperforms other comparison methods in most situations. The superior performance of the proposed method is attributed to the integration of feature selection, missing view imputation, and exploration of local structures within both sample and

feature spaces in a unified learning framework, where these three components mutually enhance each other. Additionally, the dynamic sample quality assessment module in UNIFIER helps mitigate the impact of unreliable restored data and outliers.

Parameters Sensitivity and Convergence   The objective function in Eq.(7) involves three parameters: $\alpha^{(v)}$, λ, and ξ_v. The regularization parameter ξ_v is automatically determined during the process of solving $\mathbf{S}^{(v)}$ in Eq.(16). Hence, we only investigate the performance variation of our method at different values of $\alpha^{(v)}$ and λ. Fig. 6 shows how the ACC and NMI results of our method vary with different parameter combinations. We can observe that the parameter λ is somewhat sensitive for our method, while the parameter $\alpha^{(v)}$ exhibits even more sensitive than λ. Moreover, in most cases, our method achieves prominent performances with the parameter combinations of $\alpha^{(v)}$=0.01 or 0.001 and λ=100. Therefore, we can empirically fine-tune $\alpha^{(v)}$ and λ based on the smaller range discussed previously. Fig. 7 illustrates the variation in the objective values of Eq.(7) across different iterations on the BBCS and USPS datasets. It is evident that the proposed optimization algorithm converges rapidly within 20 iterations.

Ablation Study   In this section, we conduct ablation experiments to show the significance of the component terms in the proposed UNIFIER. Two variants of UNIFIER are conducted as follows:

(1) UNIFIER-I (without the bi-level collaborative missing view completion module):

$$\min_{\Theta} \sum_{v=1}^{V}\Bigg[\alpha^{(v)}\sum_{i=1}^{n}\Big(e_i^{(v)}\|\mathbf{x}_{i.}^{(v)}-\mathbf{x}_{i.}^{(v)}\mathbf{W}^{(v)}\|_2^2+\gamma(\sqrt{\mathbf{e}^{(v)}}-1)^2\Big) \\ +\frac{1}{2}\sum_{i,j=1}^{n}\|\mathbf{x}_{i.}^{(v)}\mathbf{W}^{(v)}-\mathbf{x}_{j.}^{(v)}\mathbf{W}^{(v)}\|_2^2 s_{ij}^{(v)}+\xi_v\|\mathbf{S}^{(v)}\|_F^2\Bigg]$$

(19)

In UNIFIER-I, the missing samples in $\mathbf{X}^{(v)}$ are first filled with the means of the features, followed by the same feature selection process as in UNIFIER.

(2) UNIFIER-II (without the dynamic sample quality assessment module):

$$\min_{\Theta} \sum_{v=1}^{V}\Bigg[\alpha^{(v)}\sum_{i=1}^{n}\|\widetilde{\mathbf{x}}_{i.}^{(v)}-\widetilde{\mathbf{x}}_{i.}^{(v)}\mathbf{W}^{(v)}\|_2^2+\lambda\|\mathbf{W}^{(v)}\|_{2,1} \\ +\frac{1}{2}\sum_{i,j=1}^{n}\|\widetilde{\mathbf{x}}_{i.}^{(v)}\mathbf{W}^{(v)}-\widetilde{\mathbf{x}}_{j.}^{(v)}\mathbf{W}^{(v)}\|_2^2 s_{ij}^{(v)}+\xi_v\|\mathbf{S}^{(v)}\|_F^2\Bigg]$$

(20)

Table 3 shows the ablation experiment results on six datasets. We can observe a significant decrease in the performance of UNIFIER-I compared to UNIFIER in terms of ACC and NMI. This underscores the benefit of integrating feature selection with missing view completion to improve performance. Furthermore, UNIFIER outperforms UNIFIER-II, highlighting the effectiveness of the dynamic sample quality assessment module in mitigating the influence of unreliable restored data and outliers.
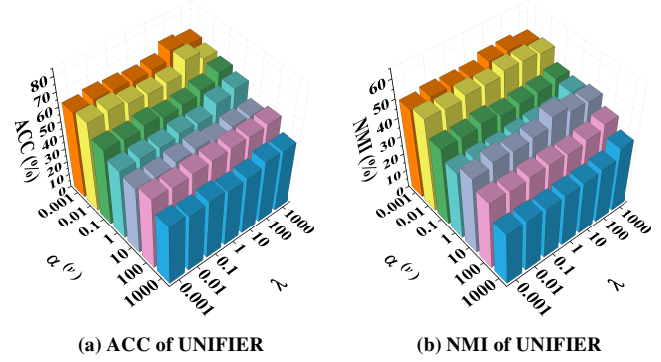

(a) ACC of UNIFIER   (b) NMI of UNIFIER

Figure 6. ACC of UNIFIER with varying parameters $\alpha^{(v)}$ and λ on BBCS dataset.
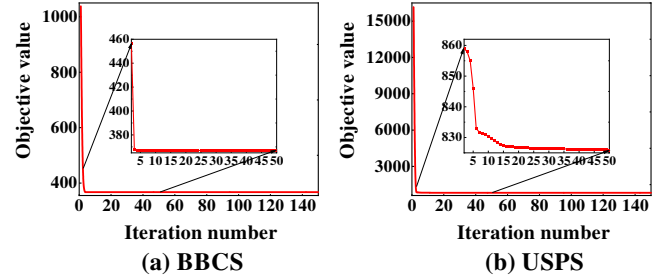

(a) BBCS   (b) USPS

Figure 7. Object values of UNIFIER with different numbers of iterations on BBCS and USPS datasets.

6   Conclusions

In this paper, we have proposed a novel MUFS method for incomplete multi-view data, addressing the limitation of current methods that treat missing data imputation and feature selection as two separate processes. The proposed method UNIFIER integrates feature selection, missing view completion, and the exploration of both local sample and feature structures into a unified learning framework. Furthermore, UNIFIER can automatically assign lower weights to low-quality samples, which is beneficial for alleviating the impact of unreliable restored data and outliers. Experiment results on benchmark datasets demonstrated the superior performance of UNIFIER compared with SOTA methods. In our future work, we will extend our framework to conduct parallel and incremental unsupervised feature selection on large-scale incomplete multi-view datasets.

Datasets	UNIFIER		UNIFIER-I		UNIFIER-II	
	ACC	NMI	ACC	NMI	ACC	NMI
BBCS	**71.55**	**54.26**	63.79	42.90	58.62	42.17
Yale	**61.21**	**59.66**	53.94	55.37	59.01	57.53
BBC4	**68.76**	**47.45**	42.77	17.48	65.22	44.03
Aloi	**74.91**	**78.74**	42.37	47.53	72.98	73.51
Cal20	**67.48**	**55.32**	63.29	48.67	65.56	53.21
USPS	**62.87**	**41.93**	58.98	38.52	58.24	36.28

Table 3. Means (%) of ACC and NMI for different variants of UNIFIER on six multi-view datasets.

Acknowledgments

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.

[Barron, 2019] Jonathan T. Barron. A general and adaptive robust loss function. In *Proc. IEEE/CVF Int. Conf. Comput. Vis. Pattern Recognit.*, pages 4331–4339, 2019.

[Cai *et al.*, 2018] Lei Cai, Zhengyang Wang, Hongyang Gao, Dinggang Shen, and Shuiwang Ji. Deep adversarial learning for multi-modality missing data completion. In *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, pages 1158–1166, 2018.

[Cao and Xie, 2024] Zhiwen Cao and Xijiong Xie. Multi-view unsupervised complementary feature selection with multi-order similarity learning. *Knowl. Based Syst.*, 283:111172, 2024.

[Chen *et al.*, 2021] Yongyong Chen, Shuqin Wang, Chong Peng, Zhongyun Hua, and Yicong Zhou. Generalized non-convex low-rank tensor approximation for multi-view subspace clustering. *IEEE Trans. Image Process.*, 30:4022–4035, 2021.

[Dong *et al.*, 2012] Xiaowen Dong, Pascal Frossard, Pierre Vandergheynst, and Nikolai Nefedov. Clustering with multi-layer graphs: A spectral perspective. *IEEE Trans. Signal Process.*, 60(11):5820–5831, 2012.

[Fang *et al.*, 2023] Siguo Fang, Dong Huang, Changdong Wang, and Yong Tang. Joint multi-view unsupervised feature selection and graph learning. *IEEE Trans. Emerg. Top. Comput. Intell.*, 1:1–16, 2023.

[He *et al.*, 2005] Xiaofei He, Deng Cai, and Partha Niyogi. Laplacian score for feature selection. In *Adv. Neural Inf. Process. Syst.*, pages 507–514, 2005.

[Hou *et al.*, 2013] Chenping Hou, Feiping Nie, Xuelong Li, Dongyun Yi, and Yi Wu. Joint embedding learning and sparse regression: A framework for unsupervised feature selection. *IEEE Trans. Cybern.*, 44(6):793–804, 2013.

[Huang *et al.*, 2019] Zhenyu Huang, Joey Tianyi Zhou, Xi Peng, Changqing Zhang, Hongyuan Zhu, and Jiancheng Lv. Multi-view spectral clustering network. In *Proc. Int. Joint Conf. Artif. Intell.*, pages 2563–2569, 2019.

[Li *et al.*, 2019] Xuelong Li, Han Zhang, Rui Zhang, and Feiping Nie. Discriminative and uncorrelated feature selection with constrained spectral analysis in unsupervised learning. *IEEE Trans. Image Process.*, 29:2139–2149, 2019.

[Lin *et al.*, 2022] Yijie Lin, Yuanbiao Gou, Xiaotian Liu, Jinfeng Bai, Jiancheng Lv, and Xi Pen. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(4):4447–4461, 2022.

[Liu *et al.*, 2016] Hongfu Liu, Haiyi Mao, and Yun Fu. Robust multi-view feature selection. In *Proc. IEEE Int. Conf. Data Mining*, pages 281–290, 2016.

[Liu *et al.*, 2018] Weifeng Liu, Xinghao Yang, Dapeng Tao, Jun Cheng, and Yuanyan Tang. Multiview dimension reduction via hessian multiset canonical correlations. *Inf. Fusion*, 100(41):119–128, 2018.

[Liu *et al.*, 2021a] Haifeng Liu, Zheng Yang, and Zhaohui Wu. Locality-constrained concept factorization. In *Proc. Int. Joint Conf. Artif. Intell.*, pages 1378–1383, 2021.

[Liu *et al.*, 2021b] Jiyuan Liu, Xinwang Liu, Yi Zhang, Pei Zhang, Wenxuan Tu, Siwei Wang, Sihang Zhou, Weixuan Liang, Siqi Wang, and Yuexiang Yang. Self-representation subspace clustering for incomplete multi-view data. In *Proc. ACM Int. Conf. Multimedia*, pages 2726–2734, 2021.

[Nie *et al.*, 2010] Feiping Nie, Heng Huang, Xiao Cai, and Chris Ding. Efficient and robust feature selection via joint $\ell_{2,1}$-norms minimization. In *Adv. Neural Inf. Process. Syst.*, pages 1–9, 2010.

[Nie *et al.*, 2019] Feiping Nie, Wei Zhu, and Xuelong Li. Structured graph optimization for unsupervised feature selection. *IEEE Trans. Knowl. Data Eng.*, 33(3):1210–1222, 2019.

[Nie *et al.*, 2023] Feiping Nie, Zhenyu Ma, Jingyu Wang, and Xuelong Li. Fast sparse discriminative k-means for unsupervised feature selection. *IEEE Trans. Neural Netw. Learn. Syst.*, 1:1–15, 2023.

[Nikolova and Chan, 2007] Mila Nikolova and Raymond H. Chan. The equivalence of half-quadratic minimization and the gradient linearization iteration. *IEEE Trans. Image Process.*, 16(6):1623–1627, 2007.

[Rocha and Goldenstein, 2013] Anderson Rocha and Siome Klein Goldenstein. Multiclass from binary: Expanding one-versus-all, one-versus-one and ecoc-based approaches. *IEEE Trans. Neural Netw. Learn. Syst.*, 25(2):289–302, 2013.

[Shen *et al.*, 2020] Hengtao Shen, Yonghua Zhu, Wei Zheng, and Xiaofeng Zhu. Half-quadratic minimization for unsupervised feature selection on incomplete data. *IEEE Trans. Neural Netw. Learn. Syst.*, 32(7):3122–3135, 2020.

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

[Tang *et al.*, 2021] Chang Tang, Xiao Zheng, Xinwang Liu, Wei Zhang, Jing Zhang, Jian Xiong, and Lizhe Wang. Cross-view locality preserved diversity and consensus learning for multi-view unsupervised feature selection. *IEEE Trans. Knowl. Data Eng.*, 34(10):4705–4716, 2021.

[Wang *et al.*, 2019] Hao Wang, Yan Yang, and Bing Liu. Gmc: Graph-based multi-view clustering. *IEEE Trans. Knowl. Data Eng.*, 32(6):1116–1129, 2019.

[Wen *et al.*, 2019] Jie Wen, Zheng Zhang, Yong Xu, Bob Zhang, Lunke Fei, and Hong Liu. Unified embedding alignment with missing views inferring for incomplete multi-view clustering. In *Proc. AAAI Conf. Artif. Intell.*, pages 5393–5400, 2019.

[Xu *et al.*, 2021] Yuanyuan Xu, Yu Yin, Jun Wang, Jinmao Wei, Jian Liu, Lina Yao, and Wenjie Zhang. Unsupervised cross-view feature selection on incomplete data. *Knowl. Based Syst.*, 234:107595, 2021.

[Yuan *et al.*, 2022] Haoliang Yuan, Junyu Li, Yong Liang, and Yuanyan Tang. Multi-view unsupervised feature selection with tensor low-rank minimization. *Neurocomputing*, 487:75–85, 2022.

[Zhang and Yin, 2017] Huamin Zhang and Hongcai Yin. Conjugate gradient least squares algorithm for solving the generalized coupled sylvester matrix equations. *Comput. Math. with Appl.*, 73(12):2529–2547, 2017.

[Zhang *et al.*, 2019] Rui Zhang, Feiping Nie, Xuelong Li, and Xian Wei. Feature selection with multi-view data: A survey. *Inf. Fusion*, 50:158–167, 2019.

[Zhang *et al.*, 2020] Rui Zhang, Yunxing Zhang, and Xuelong Li. Unsupervised feature selection via adaptive graph learning and constraint. *IEEE Trans. Neural Netw. Learn. Syst.*, 33(3):1355–1362, 2020.

[Zhang *et al.*, 2021] Bin Zhang, Qianyao Qiang, Fei Wang, and Feiping Nie. Flexible multi-view unsupervised graph embedding. *IEEE Trans. Image Process.*, 30:4143–4156, 2021.