Unified Embedding Alignment with Missing Views Inferring for Incomplete Multi-View Clustering

Jie Wen,[†,1] **Zheng Zhang,**[†,2] **Yong Xu,**[∗,1] **Bob Zhang,**[3] **Lunke Fei,**[4] **Hong Liu**[5]

[1]Bio-Computing Research Center, Harbin Institute of Technology, Shenzhen, Shenzhen, China
[2]The University of Queensland, Australia
[3]Department of Computer and Information Science, University of Macau, Taipa, Macau, PR China
[4]School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, China
[5]Engineering Lab on Intelligent Perception for Internet of Things, Shenzhen Graduate School, Peking University, China
jiewen_pr@126.com, darrenzz219@gmail.com, yongxu@ymail.com, bobzhang@umac.mo,
flksxm@126.com, hongliu@pku.edu.cn

Abstract

Multi-view clustering aims to partition data collected from diverse sources based on the assumption that all views are complete. However, such prior assumption is hardly satisfied in many real-world applications, resulting in the incomplete multi-view learning problem. The existing attempts on this problem still have the following limitations: 1) the underlying semantic information of the missing views is commonly ignored; 2) The local structure of data is not well explored; 3) The importance of different views is not effectively evaluated. To address these issues, this paper proposes a Unified Embedding Alignment Framework (UEAF) for robust incomplete multi-view clustering. In particular, a locality-preserved reconstruction term is introduced to infer the missing views such that all views can be naturally aligned. A consensus graph is adaptively learned and embedded via the reverse graph regularization to guarantee the common local structure of multiple views and in turn can further align the incomplete views and inferred views. Moreover, an adaptive weighting strategy is designed to capture the importance of different views. Extensive experimental results show that the proposed method can significantly improve the clustering performance in comparison with some state-of-the-art methods.

Introduction

Multi-view clustering has attracted much attention in the communities of computer vision and data mining in recent years. Compared with the single-view clustering, multi-view clustering generally can achieve a better performance due to the complementary properties of the heterogeneous features collected from diverse sources or described by multiple features (Xu, Tao, and Xu 2013; Qian and Zhai 2014).

Notably, the conventional single-view clustering methods are not suitable to group multi-view data, because they ignore the diversity and complementary characteristics of multiple features from different views (Chao, Sun, and Bi 2017).

∗'†' indicates co-first authors with equal contributions; '∗' indicates the corresponding author.

Recently, many researchers focused on learning a common indicator matrix or shared similarity graph of different views for multi-view clustering. For instance, based on the spectral clustering, the diversity-induced multi-view subspace clustering accumulates multiple graphs of all views into a common graph for clustering (Cao et al. 2015). Multi-view subspace clustering directly learns a common and consistent representation from the representation graphs of all views (Gao et al. 2015). Xia et al. proposed to learn a common low-rank graph from multiple similarity graphs constructed from all views independently (Xia et al. 2014). Moreover, the kernel based multi-view clustering (Tzortzis and Likas 2012) and deep multi-view clustering (Zhao, Ding, and Fu 2017) have been developed to handle nonlinear features and achieve a very promising performance. *It is worth noting that these multi-view clustering methods commonly require that all views of data should be complete.* However, it is very common that the available multi-view data tends to be incomplete, which means features of some views are missing. For example, in document clustering, different translated languages of documents can be regarded as the features of multiple views. However, some documents may do not have all translated versions (Hu and Chen 2018). Another typical example is in the audio-visual appearance, where the audio and visual represent the two views of a speaker, while we may only have the samples of the audio or the visual view for some speakers in some cases (Li, Jiang, and Zhou 2014).

It is clear that the existing multi-view clustering methods fail to cluster the multi-view data with incomplete views, because it is impossible to learn the common similarity graph or low-dimensional representation for all views. Moreover, the available complementary information among multiple views is very limited due to the unpaired missing views. These factors make the incomplete views a challenging clustering problem. Some attempts have been made to adapt such incomplete learning problem. For example, a kernel canonical correlation analysis based method is proposed to complete the incomplete kernel matrices of the incomplete

views (Trivedi et al. 2010). However, it requires at least one complete view as reference for view completion. Gao et al. proposed to learn the consensus representation based on the kernel alignment (Gao, Peng, and Jian 2016). However, this method can not handle the case with large incomplete rate of views. The prevalent incomplete multi-view clustering methods are mainly based on the matrix factorization strategy (Hu and Chen 2018; Wen et al. 2018c; Li, Jiang, and Zhou 2014; Rai et al. 2016; Zhao, Liu, and Fu 2016; Shao, He, and Philip 2015). Compared with the above spectral clustering based methods, these methods can directly obtain the consensus representation from the incomplete views. However, there are still some limitations for these methods: 1) The local geometric structure of data is not well explored, and they cannot guarantee the compactness and discriminability of the obtained common features; 2) The missing views are not precisely recovered, and they cannot fully capture the hidden information of the missing views for consensus representation learning; 3) The importance of different views is ignored, and it is obviously unreasonable for incomplete multi-view learning when all views are treated equally during the learning phases.

In this paper, we propose a novel Unified Embedding Alignment Framework, dubbed as UEAF, to address the above issues. Firstly, we introduce an error matrix with Laplacian regularizer to recover the missing views. This approach not only allows all incomplete views to be aligned for the consensus representation learning, but also provides a novel way to exploit the hidden information of the missing views, which is beneficial to the final clustering tasks. Secondly, we develop a novel reverse graph regularization term on all recovered samples view by view to guarantee the consensus of local structure of multiple views. Thirdly, we impose an adaptively learned weight vector on the learning model of each view to identify their importance and in turn can enhance the discriminability of the learned common representation. Experimental results demonstrate that UEAF can greatly improve the clustering performance by integrating these factors into a unified framework. In brief, the proposed UEAF has the following contributions.

(1) To the best of our knowledge, it is a new way of thinking to simultaneously reconstruct the missing views and learn the common representation of multiple views for incomplete multi-view clustering.

(2) Different from the existing methods, the missing views completion and common representation learning are seamlessly incorporated into one unified framework to interactively facilitate its counterpart, rather than simply discarding the missing views or using suboptimal two-phase learning.

(3) A novel reverse graph regularization term is developed to guarantee the consensus of local structure of multiple views, which jointly enables the incomplete views to be further aligned and ensures the compactness and consistence of the learned common representation of multiple views.

Related Work

Multiple incomplete views clustering (MIC) provides a flexible framework to address the incomplete problem of multi-view clustering. The objective function of MIC is designed as follows (Shao, He, and Philip 2015):

$$\min_{P^{(v)}, U^{(v)}, P^*} \sum_v^l \left[\begin{array}{l} \left\| W^{(v)} \left(X^{(v)} - P^{(v)} U^{(v)} \right) \right\|_F^2 + \\ \beta_v \| P^{(v)} \|_{2,1} + \alpha_v \left\| W^{(v)} \left(P^{(v)} - P^* \right) \right\|_F^2 \end{array} \right]$$
$$s.t. \ U^{(v)} \ge 0, P^{(v)} \ge 0, P^* \ge 0, \tag{1}$$

where β_v and α_v are the penalty parameters of the vth view, $U^{(v)}$ is the basis matrix of the vth view, and l is the number of views. In MIC, all missing views are filled in the average vector of samples in the corresponding view, and then $X^{(v)} \in R^{n \times m_v}$ includes the missing and non-missing instances of the vth view, where n and m_v are the number of samples and feature dimensionality of the vth view, respectively. $P^* \in R^{n \times c}$ is the common representation of all samples shared by all views. In model (1), the weighting matrix $W^{(v)} \in R^{n \times n}$ is defined as:

$$W_{i,i}^{(v)} = \left\{ \begin{array}{ll} 1, & if\,the\,v th\,view\,contains\,the\,ith\,sample \\ \delta_v, & otherwise, \end{array} \right. \tag{2}$$

where $\delta_v = n_v/n$, n_v denotes the number of non-missing instances of the vth view. By introducing the weighting matrix $W^{(v)}$, MIC can exploit the non-missing instances of all views to learn the common representation.

Proposed Approach

In this section, we present the proposed UEAF, which infers the missing views and exploits the graph embedding technique to align all views for effective consensus representation learning. The proposed UEAF is mainly composed of three sub-models, *i.e.*, consensus representation learning, reverse graph regularization, and adaptively weighted incomplete multi-view learning.

Consensus Representation Learning

Learning a consensus representation for all views is one of the most favorable approaches for incomplete multi-view clustering. However, all the existing works only exploit the available non-missing views to learn the common latent representation while ignoring the underlying information of the missing views. In this paper, we propose the following model to simultaneously exploit the available information of non-missing views and hidden information of the missing views for the common representation learning:

$$\min_{E^{(v)}, U^{(v)}, P} \sum_{v=1}^l \left(\begin{array}{l} \left\| X^{(v)} + E^{(v)} W^{(v)} - U^{(v)} P \right\|_F^2 \\ + \frac{\lambda_1}{2} \sum_{j=1}^{m_v} \sum_{i=1}^{m_v} \left\| E_{i,:}^{(v)} - E_{j,:}^{(v)} \right\|_2^2 G_{i,j}^{(v)} \end{array} \right)$$
$$s.t. \ U^{(v)T} U^{(v)} = I, \tag{3}$$

where λ_1 is a positive penalty parameter, $X^{(v)} \in R^{m_v \times n}$ denotes the instances from the vth view. For each set $X^{(v)}$, all elements of the missing instances are filled in to be 0. $E^{(v)} \in R^{m_v \times n_v^m}$ is the error matrix used to model the missing instances of the vth view, n_v^m denotes the number of missing instances of the vth view, $E_{i,:}^{(v)}$ and $E_{j,:}^{(v)}$ denote the ith row and jth row vector of matrix $E^{(v)}$. $U^{(v)} \in R^{m_v \times c}$ is the basis matrix of the vth view, $P \in R^{c \times n}$ is the latent common representation of all views, where c is the cluster number or dimensionality of the latent representation. I is

the identity matrix. $W^{(v)} \in R^{n_v^m \times n}$ is an index matrix of the vth view and is defined as follows:

$$W_{i,j}^{(v)} = \begin{cases} 1, & \begin{array}{c} \textit{if the } j-\textit{th instance is the } i-\textit{th missing} \\ \textit{instance in the } v-\textit{th view} \end{array} \\ 0, & \textit{otherwise.} \end{cases} \tag{4}$$

In model (3), $G^{(v)} \in R^{m_v \times m_v}$ is the neighbor graph of features from the vth view, which is pre-constructed as:

$$G_{i,j}^{(v)} = \begin{cases} 1, & \textit{if } \bar{X}_{i,:}^{(v)} \in \psi\left(\bar{X}_{j,:}^{(v)}\right) \textit{ or } \bar{X}_{j,:}^{(v)} \in \psi\left(\bar{X}_{i,:}^{(v)}\right) \\ 0, & \textit{otherwise,} \end{cases} \tag{5}$$

where $\psi\left(\bar{X}_{j,:}^{(v)}\right)$ denotes the set of k nearest neighbors of the jth feature, *i.e.*, $\bar{X}_{j,:}^{(v)}$, $\bar{X}^{(v)}$ is the set of the available instances of the vth view.

We can easily deduce that the problem (3) can be transformed into the following equivalent formula:

$$\min_{E^{(v)}, U^{(v)}, P} \sum_{v=1}^{l} \left(\begin{array}{c} \left\|X^{(v)} + E^{(v)}W^{(v)} - U^{(v)}P\right\|_F^2 \\ +\lambda_1 Tr\left(E^{(v)T}L^{(v)}E^{(v)}\right) \end{array} \right)$$
$$s.t. \; U^{(v)T}U^{(v)} = I, \tag{6}$$

where $L^{(v)}$ is the Laplacian matrix of graph $G^{(v)}$ and is calculated as $L^{(v)} = D^{(v)} - G^{(v)}$, $D^{(v)}$ is a diagonal matrix and its ith diagonal element is calculated by $D_{i,j}^{(v)} = \sum_{j=1}^{m_v} G_{i,j}^{(v)}$. In (6), the orthogonal constraint $U^{(v)T}U^{(v)} = I$ is introduced to make the basis matrix independent.

From (6), we can find that the complete instances of each view is $\left(X^{(v)} + E^{(v)}W^{(v)}\right)$. By recovering the missing views, all incomplete views are naturally aligned to learn the consensus representation. Importantly, we can obtain more reasonable missing instances by introducing the graph regularization term, which further efficiently guides the model to learn the consensus representations of all views.

Reverse Graph Regularization

For multi-view data, it naturally shares some common information across different views, such as the common representation and common local structure (Nie, Cai, and Li 2017). The local structure preserving is significant and indispensable in subspace learning, which is beneficial to improve the compactness of the learned low-dimensional representation. However, for the incomplete multi-view data, it is obviously impossible to capture the local manifold structure of the data via the existing methods as far as we know since the numbers of available instances of all views are less than the total sample number. To overcome this difficulty, we propose the following reverse graph learning model:

$$\min_{S} \sum_{v=1}^{l} \sum_{j=1}^{n} \sum_{i=1}^{n} \left\|U^{(v)}P_{:,i} - U^{(v)}P_{:,j}\right\|_2^2 S_{i,j}^2$$
$$s.t. \;\; \forall i, S_{i,:}\mathcal{1} = 1, 0 \leq S_{i,j} \leq 1, S_{i,i} = 0, \tag{7}$$

where $S \in R^{n \times n}$ denotes the nearest neighbor graph with each element representing the similarity degree between the corresponding two instances. $\mathcal{1}$ is a vector with all 1s. Introducing the constraint $S_{i,:}\mathcal{1} = 1$ can avoid the trivial solution that any sample is not connected with its neighbors.

For the non-negative similarity graph S and its Laplacian matrix L_S, we have the following Theorem (Nie, Cai, and Li 2017; Nie et al. 2016).

Theorem 1 (Mohar et al. 1991; Chung 1996): The multiplicity c of eigenvalue 0 of the Laplacian matrix L_S is equal to the number of the connected components in graph S.

It is notable that for the multi-view data with c clusters, the constructed ideal similarity graph should also have exactly c connected components, in which samples connected by each component are naturally regarded as one cluster (Nie, Cai, and Li 2017). Many works have also proved that learning such an ideal graph is beneficial to obtain a better clustering performance (Ren et al. 2018; Wen et al. 2018a; Zhan et al. 2018). Based on this observation, we add a rank constraint to model (7) based on Theorem 1 as follows:

$$\min_{S} \sum_{v=1}^{l} \sum_{j=1}^{n} \sum_{i=1}^{n} \left\|U^{(v)}P_{:,i} - U^{(v)}P_{:,j}\right\|_2^2 S_{i,j}^2 \tag{8}$$
$$s.t. \; \forall i, S_{i,:}\mathcal{1} = 1, 0 \leq S_{i,j} \leq 1, S_{i,i} = 0, rank\left(L_S\right) = n - c,$$

where $rank(L_S)$ denotes the rank of matrix L_S.

By exploiting the reverse graph regularization, the consensus manifold structure across multiple views can be guaranteed, which further aligns all recovered incomplete views.

Adaptively Weighted Incomplete Multi-view Learning

For multi-view data, different views generally have different physical meanings and discriminant powers. Especially for the incomplete multi-view cases, the available discriminative information of different views will have huge differences due to the different feature dimensions and numbers of the available instances of multiple views. To this end, we exploit the following adaptive weighting strategy to balance the importance of different views:

$$\min_{\alpha^{(v)}} \sum_{v=1}^{l} \left(\alpha^{(v)}\right)^r \Gamma^{(v)}, \; s.t. \; \sum_{v=1}^{l} \alpha^{(v)} = 1, \alpha^{(v)} \geq 0, \tag{9}$$

where $\Gamma^{(v)}$ denotes the objective learning model of the vth view. $\alpha^{(v)}$ is a positive weight to balance the significance of the vth view. Parameter $r > 1$ is used to control the smoothness of the weights distribution. In this way, the model can adaptively figure out the complementary effects of different views with varying discriminant powers for model training.

Overall Objective Function

Finally, by incorporating the consensus representation learning and reverse graph regularization into the adaptive weighting model (*i.e.*, combing (6), (8) and (9)), we obtain the final incomplete multi-view clustering model as follows:

$$\min_{\Upsilon} \sum_{v=1}^{l} \left(\alpha^{(v)}\right)^r \left(\begin{array}{c} \left\|X^{(v)} + E^{(v)}W^{(v)} - U^{(v)}P\right\|_F^2 \\ +\lambda_1 Tr\left(E^{(v)T}L^{(v)}E^{(v)}\right) \\ +\frac{\lambda_2}{2} \sum_{i,j=1}^{n} \left\|U^{(v)}P_{:,i} - U^{(v)}P_{:,j}\right\|_2^2 S_{i,j}^2 \end{array} \right)$$
$$s.t. \; U^{(v)T}U^{(v)} = I, \sum_{v=1}^{l} \alpha^{(v)} = 1, \alpha^{(v)} \geq 0, \forall i, S_{i,:}\mathcal{1} = 1,$$
$$0 \leq S_{i,j} \leq 1, S_{i,i} = 0, rank\left(L_S\right) = n - c, \tag{10}$$
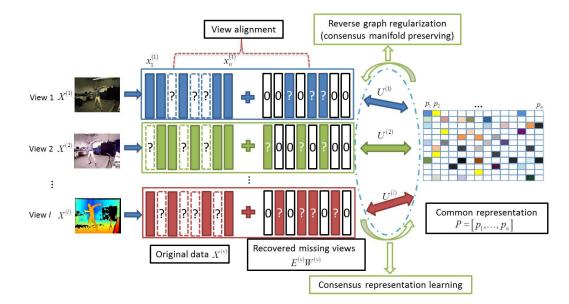
Figure 1: The framework of the proposed method.

where λ_2 is a penalty parameter, $\Upsilon = \{E^{(v)}, U^{(v)}, P, S, \alpha^{(v)}\}$ is the set of variables to compute. The overall learning framework of the proposed method is outlined in Fig.1.

Optimization

We introduce an alternating optimization algorithm to solve problem (10) (Shen, Wen, and Zhang 2014; Xu et al. 2012). According to the Ky Fan's Theorem (Nie, Cai, and Li 2017; Fan 1949) and constraint $U^{(v)T}U^{(v)} = I$, problem (10) is equivalent to the following optimization problem:

$$\min_{\Upsilon} \sum_{v=1}^{l} \left(\alpha^{(v)}\right)^r \left(\begin{array}{c} \|X^{(v)} + E^{(v)}W^{(v)} - U^{(v)}P\|_F^2 \\ +\lambda_1 Tr\left(E^{(v)T}L^{(v)}E^{(v)}\right) \\ +\lambda_2 Tr\left(PL_{S^2}P^T\right) \end{array} \right)$$

$$+ \lambda_3 Tr\left(F^T L_S F\right) \ s.t. \ U^{(v)T}U^{(v)} = I, \sum_{v=1}^{l}\alpha^{(v)} = 1,$$

$$\alpha^{(v)} \geq 0, \forall i, S_{i,:}I = 1, S_{i,i} = 0, 0 \leq S_{i,j} \leq 1, F^T F = I, \quad (11)$$

where λ_3 is a penalty parameter, $\Upsilon = \{E^{(v)}, U^{(v)}, P, S, \alpha^{(v)}, F\}$, L_{S^2} denotes the Laplacian matrix of graph $(S \odot S)$, $\odot$ is the hadamard product.

Step 1: Calculate P. Fixing all the other variables, we can calculate P by minimizing the following function:

$$\Psi(P) = \sum_{v=1}^{l} \left(\alpha^{(v)}\right)^r \left(\begin{array}{c} \|X^{(v)} + E^{(v)}W^{(v)} - U^{(v)}P\|_F^2 \\ +\lambda_2 Tr\left(PL_{S^2}P^T\right) \end{array} \right).$$
(12)

By setting the partial derivation of $\Psi(P)$ *w.r.t.* P to zero, we can obtain the closed-form solution to P as follows:

$$P = \left(\sum_{v=1}^{l}\left(\alpha^{(v)}\right)^r U^{(v)T}Y^{(v)}\right)(I + \lambda_2 L_{S^2})^{-1} \bigg/ \sum_{v=1}^{l}\left(\alpha^{(v)}\right)^r,$$
(13)

where $Y^{(v)} = X^{(v)} + E^{(v)}W^{(v)}$.

Step 2: Calculate S. Fixing the other variables, the sub-problem to variable S is simplified as follows:

$$\min_{S} \sum_{v=1}^{l} \left(\alpha^{(v)}\right)^r \left(\lambda_2 Tr\left(PL_{S^2}P^T\right)\right) + \lambda_3 Tr\left(F^T L_S F\right)$$

$$s.t. \ \forall i, S_{i,:}I = 1, 0 \leq S_{i,j} \leq 1, S_{i,i} = 0, \quad (14)$$

which can be further simplified as follows:

$$\sum_{i=1}^{n}\left(\min_{S_{i,:}I=1, 0 \leq S_{i,j} \leq 1, S_{i,i}=0} \sum_{j=1, j \neq i}^{n} \left(S_{i,j} + \frac{\lambda_3 H_{i,j}^F}{2\lambda_2 H_{i,j}^P}\right)^2 \right),$$
(15)

where $H_{i,j}^F = \|F_{i,:} - F_{j,:}\|_2^2$, $H_{i,j}^P = \sum_{v=1}^{l}\left(\alpha^{(v)}\right)^r \|P_{:,i} - P_{:,j}\|_2^2$. The closed-form solution of problem (15) can be achieved by an efficient algorithm presented in (Nie et al. 2016).

Step 3: Calculate F. Fixing the other variables, we can obtain the solution to F by solving the following problem:

$$\min_{F^T F = I} Tr\left(F^T L_S F\right), \quad (16)$$

which can be optimized by a set of Eigenvectors corresponding to the first c minimum Eigenvalues of L_S.

Step 4: Calculate $U^{(v)}$. Fixing the other variables, the sub-problem to variable $U^{(v)}$ is reformulated as:

$$\min_{U^{(v)T}U^{(v)}=I} \left\|X^{(v)} + E^{(v)}W^{(v)} - U^{(v)}P\right\|_F^2. \quad (17)$$

If we denote $Y^{(v)} = X^{(v)} + E^{(v)}W^{(v)}$, we can obtain the optimal solution of $U^{(v)} = BR^T$, where B and R are the left and right singular matrices of matrix $\left(Y^{(v)}P^T\right)$.

Step 5: Calculate $E^{(v)}$. When fixing all the other variables that are irrelevant to $E^{(v)}$, we have:

$$\min_{E^{(v)}} \left\|X^{(v)} + E^{(v)}W^{(v)} - U^{(v)}P\right\|_F^2 + \lambda_1 Tr\left(E^{(v)T}L^{(v)}E^{(v)}\right). \quad (18)$$

Algorithm 1 : UEAF (solving (11))

Input: Incomplete multi-view data $X = \left\{X^{(v)}\right\}_{v=1}^{l}$ whose missing views are filled in 0s, index matrix of missing views $W = \left\{W^{(v)}\right\}_{v=1}^{l}$, parameters λ_1, λ_2, λ_3, r, nearest neighbor graph $G = \left\{G^{(v)}\right\}_{v=1}^{l}$ of features.

Initialization: $\alpha^{(v)} = 1/l$, $U^{(v)}$ is an orthogonal matrix with random values, random matrix $E^{(v)}$, random graph S, initialize F by solving (16) based on S.

while not converged **do**

    1. Update P using (13);
    2. Update S by solving (15);
    3. Update F by solving (16);
    for v from 1 to l
        4. Update $U^{(v)}$ by solving (17);
        5. Update $E^{(v)}$ using (20);
    end
    6. Update $\alpha^{(v)}$ using (22).

end while

Output: $P, U^{(v)}, E^{(v)}, S$

Since all instances of $X^{(v)}$ corresponding to $E^{(v)}$ are zeros, problem (18) can be rewritten as:

$$\Psi\left(E^{(v)}\right) = \left\|E^{(v)} - U^{(v)}PW^{(v)T}\right\|_F^2 + \lambda_1 Tr\left(E^{(v)T}L^{(v)}E^{(v)}\right). \tag{19}$$

By setting the partial derivation $\partial\Psi\left(E^{(v)}\right)/\partial E^{(v)} = 0$, we can obtain the optimal solution of $E^{(v)}$ as follows:

$$E^{(v)} = \left(I + \lambda_1 L^{(v)}\right)^{-1} U^{(v)}PW^{(v)T}. \tag{20}$$

Step 6: Calculate $\alpha^{(v)}$. By fixing the other variables, $\alpha^{(v)}$ can be obtained by solving the following problem:

$$\min_{\alpha^{(v)}>0,\sum_{v=1}^{l}\alpha^{(v)}=1}\sum_{v=1}^{l}\left(\alpha^{(v)}\right)^r d^{(v)}, \tag{21}$$

where $d^{(v)} = \left\|X^{(v)} + E^{(v)}W^{(v)} - U^{(v)}P\right\|_F^2 + \lambda_1 Tr(E^{(v)T}L^{(v)}E^{(v)}) + \lambda_2 Tr(PL_{S^2}P^T)$. The optimal solution to (21) is given by (Zhang et al. 2018):

$$\alpha^{(v)} = \left(d^{(v)}\right)^{1/(1-r)} \Big/ \sum_{v=1}^{l}\left(d^{(v)}\right)^{1/(1-r)} \tag{22}$$

Algorithm 1 briefly summarizes the optimization procedures of the proposed method.

Convergence and Computational Complexity

We have provided an iterative algorithm to optimize the resulting optimization problem, which is divided into six sub-problems. Obviously, the six sub-problems are all convex and have the closed-form solution *w.r.t.* each parameter. Therefore, we conclude that the objective function is monotonically decreasing towards a stationary point. Moreover, the objective function (11) is lower bounded. The above two factors ensure the proposed optimization approach to finally

Table 1: Description of the used multi-view datasets.

Database	# Class	# View	# Samples	# Features
BUAA	10	2	90	100/100
Handwritten	10	2	2000	240/76
3 Sources	6	3	169	3560/3631/3068
BBCSport	5	4	116	1991/2063/2113/2158

find the local optimal point of the objective problem, which guarantees the convergence property.

Herein, we discuss the computational complexity of algorithm 1. We can find that the highest computational costs are the inverse operation in Steps 1 and 5, Eigenvalue decomposition in Step 3, and singular value decomposition (SVD) in Step 4. For an $n \times n$ matrix, the computational complexities of the inverse operation and Eigenvalue decomposition are $O(n^3)$. For an $m \times n$ matrix, the computational complexity of SVD is $O(mn^2)$. For Step 5, since the inverse operation of $(I + \lambda_1 L^{(v)})^{-1}$ can be calculated outside the loop, its computational complexity can be ignored. Thus the total computational complexity of algorithm 1 is about $O(\tau(2n^3 + \sum_{v=1}^{l} m_v c^2))$, where τ is the iteration number.

Experiments and Analysis

Experimental Settings

Dataset: (1) **BUAA-visnir face dataset (BUAA)** (Huang, Sun, and Wang 2012): Following the experimental settings in (Zhao, Liu, and Fu 2016), a subset of BUAA which is composed of 90 visual images and 90 near infrared images of the first 10 volunteers is chosen for comparison. (2) **Handwritten digit dataset** (Cai, Nie, and Huang 2013): The used handwritten digit dataset contains 2000 samples of 10 digits. The average pixels features with 240 dimensions and Fourier coefficient features with 76 dimensions are extracted as the two views for evaluation. (3) **3 Sources dataset**: In our experiments, we evaluate different methods on the subset of 3 Sources dataset[1], which is composed of 169 stories of six topical labels collected from the three well-known online news sources, *i.e.*, BBC, Reuters, and the Guardian. Each source can be regarded as a view. (4) **BBC-Sport**: The exploited BBCSport dataset contains 116 samples from 5 classes. Each sample is represented by 4 views. The above used datasets are briefly summarized in Table 1.

Compared methods: Eight incomplete multi-view clustering methods, including best single view (BSV), Concat, partial multi-view clustering (PVC) (Li, Jiang, and Zhou 2014), graph regularized partial multi-view clustering (GP-MVC) (Rai et al. 2016), incomplete multi-modality grouping (IMG) (Zhao, Liu, and Fu 2016), MIC (Shao, He, and Philip 2015), doubly aligned incomplete multi-view clustering (DAIMC) (Hu and Chen 2018), and online multi-view clustering (OMVC) (Shao et al. 2016) are selected to compare with the proposed UEAF. BSV fills in the missing views with the average instance of the corresponding view, then performs *k*-means on each view independently and reports the highest results among these views. Concat exploits the

---

[1]https://github.com/GPMVCDummy/GPMVC/tree/master/partialMV/PVC/recreateResults/data

Table 2: Mean NMIs (%), ACCs (%), and purities (%) of different methods on BUAA and Handwritten datasets.

Database	Method\Paired rate	NMI				ACC				Purity			
		0.1	0.3	0.5	0.7	0.1	0.3	0.5	0.7	0.1	0.3	0.5	0.7
BUAA	BSV	43.10	53.03	61.78	69.91	48.33	56.96	64.26	70.81	50.19	58.66	65.96	72.35
	Concat	51.22	51.95	52.43	56.51	45.62	46.61	47.46	52.34	47.99	49.41	49.68	54.81
	PVC	61.35	67.07	71.97	78.70	57.41	66.46	70.01	75.92	60.58	68.79	71.98	77.95
	GPMVC	62.12	70.25	74.33	81.63	58.98	68.75	74.28	78.28	61.26	71.66	75.36	81.71
	IMG	54.72	67.53	76.74	82.83	53.95	67.39	76.14	79.36	57.32	68.65	78.23	82.42
	MIC	59.92	72.59	69.56	74.11	57.77	70.00	68.22	74.89	61.11	73.33	70.44	75.33
	DAIMC	60.16	74.63	79.13	83.32	57.56	71.55	79.78	80.67	60.44	74.00	79.33	82.00
	OMVC	60.69	63.99	64.32	66.91	60.59	62.89	65.74	68.37	62.15	65.57	67.48	69.78
	Our UEAF	**72.24**	**80.76**	**82.98**	**88.54**	**72.88**	**82.22**	**83.77**	**88.00**	**74.00**	**82.68**	**84.67**	**88.89**
Handwritten	BSV	37.04	44.48	51.50	58.61	43.08	50.45	57.39	64.44	43.81	51.08	58.31	65.92
	Concat	47.71	54.43	61.12	70.30	46.01	57.46	66.45	78.64	48.20	58.57	66.73	78.72
	PVC	55.13	60.85	64.88	68.54	63.81	70.90	73.44	75.20	65.41	72.14	75.01	76.96
	GPMVC	60.99	63.99	72.23	73.68	65.60	74.04	76.94	79.06	66.58	74.85	76.21	79.70
	IMG	58.05	62.38	64.91	68.21	69.22	75.41	76.34	77.54	70.76	75.44	76.45	77.55
	MIC	44.59	51.09	57.41	66.32	49.13	59.99	64.98	74.70	50.25	60.56	66.73	75.70
	DAIMC	40.99	55.81	62.68	66.63	46.75	67.32	75.09	77.63	49.48	68.12	75.12	78.00
	OMVC	46.36	54.12	62.11	66.65	54.57	61.80	73.45	77.89	55.83	63.18	73.90	77.89
	Our UEAF	**62.34**	**67.94**	**72.52**	**74.69**	**70.15**	**79.00**	**82.72**	**85.11**	**72.15**	**79.00**	**82.61**	**85.11**

Table 3: Mean NMIs (%), ACCs (%), and purities (%) of different methods on the 3 Sources and BBCSport datasets.

Database	Method\Missing rate	NMI			ACC			Purity		
		0.1	0.3	0.5	0.1	0.3	0.5	0.1	0.3	0.5
3 Sources	BSV	50.07	34.46	22.34	56.90	47.38	39.24	68.14	57.63	48.99
	Concat	51.98	37.87	18.32	53.54	46.79	37.68	69.78	58.51	46.48
	GPMVC	34.82	30.44	28.15	48.24	44.50	42.01	60.47	58.58	57.40
	MIC	37.23	38.62	26.08	49.11	47.69	42.49	57.28	61.30	52.31
	DAIMC	52.98	49.07	41.64	56.33	52.43	50.73	68.99	67.21	63.56
	OMVC	36.48	28.42	24.34	43.95	41.11	39.53	59.37	48.76	45.44
	Our UEAF	**56.47**	**52.06**	**45.19**	**62.60**	**55.62**	**52.78**	**75.50**	**71.95**	**67.69**
BBCSport	BSV	43.73	31.03	21.40	58.62	51.31	44.03	65.79	55.07	47.59
	Concat	61.69	38.92	18.61	70.62	58.72	33.21	80.59	63.24	37.00
	GPMVC	28.23	20.04	15.48	51.44	46.89	43.91	58.39	52.76	45.29
	MIC	29.90	25.84	24.01	51.21	46.21	46.03	55.00	51.72	52.41
	DAIMC	56.62	50.17	37.89	68.62	63.45	56.89	76.90	71.72	61.03
	OMVC	30.64	41.57	40.63	53.33	51.38	48.79	56.49	59.20	57.47
	Our UEAF	**70.71**	**68.25**	**55.13**	**78.22**	**77.24**	**69.31**	**87.41**	**87.07**	**77.07**

same approach with BSV to complete the missing views, then concatenates all views into one single view, followed by k-means clustering. PVC and IMG can only deal with the special incomplete case that some samples have full views and the remaining samples have only one view.

Evaluation: The clustering accuracy (ACC), normalized mutual information (NMI), and purity are chosen as the criterion to compare these methods (Cai, Nie, and Huang 2013). It is noteworthy that: (1) For the above methods except BSV and Concat, we exploit the grid search approach to find the optimal penalty parameters of these methods and report their best clustering results. (2) For each dataset, all methods are performed on the same 5 randomly formed incomplete cases and their mean results are reported for fair comparison. (3) For the BUAA and Handwritten datasets, we randomly select 10%, 30%, 50%, and 70% samples as the paired samples. For the remaining samples, half of them miss the first view, while the other half of the samples remove the second view. For the BBCSport and 3 sources datasets, we randomly remove 10%, 30%, and 50% instances of each view to form the incomplete multi-view data.

Experimental Results and Analyses

Experimental results on the above datasets are shown in Tables 2-3. From these tables, we can observe the following interesting points.

(1) From Tables 2-3, we see that the ACC, NMI, and purity of all methods dramatically decrease as the missing rate increases. BSV and Concat achieve the worst performance in comparison with the other methods, especially for the case with large rate of missing views. This illustrates that simply filling in the missing views with the average vector is harmful to clustering. Compared with BSV and Concat, the other methods can capture more complementary information of the multiple views from the incomplete data. (2) From Table 2, we can find that IMG, GPMVC, and UEAF outperform PVC in most cases, which proves the effectiveness of preserving the local structure of data in multi-view clustering. (3) From Table 3, UEAF and DAIMC generally perform much better than MIC and OMVC, which demonstrates that UEAF and DAIMC can capture more complementary information from the incomplete multiple views. Furthermore, the superior performance of our UEAF compared with DAIMC indicates that reconstructing effective

missing views can greatly facilitate the incomplete multiview clustering. (4) From Table 2, UEAF achieves much higher ACCs, NMIs, and purities than IMG. The difference is that our method not only takes into account the importance of different views, but also exploits the hidden information of the reconstructed missing views to learn the latent representation. Therefore, the promising performance demonstrates the indispensability of two components.

Component Analysis

To analyze the effectiveness of the view completion and reverse graph regularization strategies, we compare our model with three degraded models. (1) **Degraded model 1**: We remove the reverse graph regularization terms (*i.e.,* $\lambda_2 Tr\left(PL_{S^2}P^T\right)$ and $\lambda_3 Tr\left(F^T L_S F\right)$) in (11) and only preserve the view completion term. (2) **Degraded model 2**: We remove both view completion term $\lambda_1 Tr\left(E^{(v)T}L^{(v)}E^{(v)}\right)$ and error matrix $E^{(v)}W^{(v)}$ from model (11), and the missing instances are filled in the average samples of the corresponding view. (3) **Degraded model 3**: We remove the same terms with the degraded model 2 from model (11), and exploit the similar approach as MIC to constrain the missing and non-missing views with index matrix $W^{(v)}$ defined in (2). Experimental results of the four models on the BBCSport and BUAA datasets are shown in Fig. 2. It is clear that UEAF achieves a better performance than the three models, which proves that both strategies are beneficial to incomplete multi-view clustering.

Parameter Sensitivity Analysis

We first fix parameters $r = 3$ and $k = 7$, and conduct some experiments on the BBCSport to analyze the sensitivity of ACC *w.r.t.* λ_1, λ_2 and λ_3. From Fig. 3, we can see that UEAF can obtain encouraging results when they are located in the ranges of $\left[10^1, 10^5\right]$, $\left[10^{-3}, 10^1\right]$, and $\left[10^{-4}, 10^1\right]$, respectively. In the experiments, we exploit the grid search strategy to find the three optimal parameters (Wen et al. 2018b). Moreover, we show the ACC (%) *w.r.t.* r on the BBCsport and BUAA datasets in Fig. 4. The proposed method achieves a satisfactory performance with a small parameter r (less than 5) and we simply set $r = 3$ in all experiments.

Experimental Convergence Study

In this subsection, we experimentally show the convergence of the proposed optimization algorithm (Xie et al. 2017). The objective function value and ACC *v.s.* iterations on the BBCSport and BUAA datasets are shown in Fig. 5. It is clear that the objective function value is no increasing and efficiently converges to the stationary point, which verifies the effectiveness of our method.

Conclusion

We developed a novel unified common embedding aligned with missing views inferring framework for incomplete multi-view clustering. The proposed method simultaneously considers the hidden information reconstruction of the missing views, local structure preservation of multi-view data, and the adaptive importance evaluation of different views
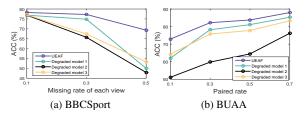

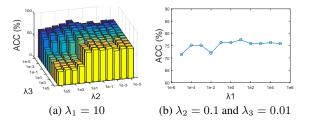Figure 2: ACCs (%) of the proposed method and different degraded models on (a) BBCsport and (b) BUAA.

Figure 3: ACC (%) v.s. (a) parameters λ_2 and λ_3 when $\lambda_1 = 10$, and (b) λ_1 when $\lambda_2 = 0.1$ and $\lambda_3 = 0.01$ on BBCSport, where 30% instances of each view are missing.
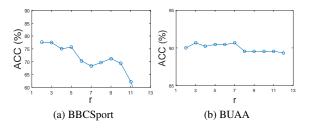

Figure 4: ACC (%) v.s. parameter r of the proposed method on (a) BBCSport with 30% missing instances of each view, and (b) BUAA with 70% paired samples.
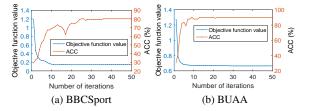

Figure 5: The objective function value and ACC (%) v.s. iterations on (a) BBCsport with 30% missing instances of each view, and (b) BUAA with 70% paired samples.

in one learning framework. Extensive experimental results demonstrate the superiority of our method.

References

Cai, X.; Nie, F.; and Huang, H. 2013. Multi-view k-means clustering on big data. In *IJCAI*, 2598–2604.

Cao, X.; Zhang, C.; Fu, H.; Liu, S.; and Zhang, H. 2015. Diversity-induced multi-view subspace clustering. In *CVPR*, 586–594.

Chao, G.; Sun, S.; and Bi, J. 2017. A survey on multi-view clustering. *arXiv preprint arXiv:1712.06246*.

Chung, F. R. 1996. Spectral graph theory (cbms regional conference series in mathematics, no. 92).

Fan, K. 1949. On a theorem of weyl concerning eigenvalues of linear transformations i. *Proceedings Nat. Academy Sci.* 35(11):652–655.

Gao, H.; Nie, F.; Li, X.; and Huang, H. 2015. Multi-view subspace clustering. In *ICCV*, 4238–4246.

Gao, H.; Peng, Y.; and Jian, S. 2016. Incomplete multi-view clustering. In *ICIIP*, 245–255. Springer.

Hu, M., and Chen, S. 2018. Doubly aligned incomplete multi-view clustering. In *IJCAI*, 2262–2268.

Huang, D.; Sun, J.; and Wang, Y. 2012. The buaa-visnir face database instructions. *School Comput. Sci. Eng., Beihang Univ., Beijing, China, Tech. Rep. IRIP-TR-12-FR-001*.

Li, S.; Jiang, Y.; and Zhou, H. 2014. Partial multi-view clustering. In *AAAI*, 1968–1974.

Mohar, B.; Alavi, Y.; Chartrand, G.; and Oellermann, O. 1991. The laplacian spectrum of graphs. *Graph theory, combinatorics, and applications* 2(871-898):12.

Nie, F.; Wang, X.; Jordan, M. I.; and Huang, H. 2016. The constrained laplacian rank algorithm for graph-based clustering. In *AAAI*, 1969–1976.

Nie, F.; Cai, G.; and Li, X. 2017. Multi-view clustering and semi-supervised classification with adaptive neighbours. In *AAAI*, 2408–2414.

Qian, M., and Zhai, C. 2014. Unsupervised feature selection for multi-view clustering on text-image web news data. In *ACM CIKM*, 1963–1966. ACM.

Rai, N.; Negi, S.; Chaudhury, S.; and Deshmukh, O. 2016. Partial multi-view clustering using graph regularized nmf. In *ICPR*, 2192–2197. IEEE.

Ren, P.; Xiao, Y.; Xu, P.; Guo, J.; Chen, X.; Wang, X.; and Fang, D. 2018. Robust auto-weighted multi-view clustering. In *IJCAI*, 2644–2650.

Shao, W.; He, L.; Lu, C.-t.; and Philip, S. Y. 2016. Online multi-view clustering with incomplete views. In *ICBD*, 1012–1017. IEEE.

Shao, W.; He, L.; and Philip, S. Y. 2015. Multiple incomplete views clustering via weighted nonnegative matrix factorization with $l_{2,1}$ regularization. In *ECML-KDD*, 318–334.

Shen, Y.; Wen, Z.; and Zhang, Y. 2014. Augmented lagrangian alternating direction method for matrix separation based on low-rank factorization. *Optimization Methods and Software* 29(2):239–263.

Trivedi, A.; Rai, P.; Daumé III, H.; and DuVall, S. L. 2010. Multiview clustering with incomplete views. In *NIPSW*, 1–7.

Tzortzis, G., and Likas, A. 2012. Kernel-based weighted multi-view clustering. In *ICDM*, 675–684. IEEE.

Wen, J.; Fang, X.; Xu, Y.; Tian, C.; and Fei, L. 2018a. Low-rank representation with adaptive graph regularization. *Neural Networks* 108:83–96.

Wen, J.; Zhang, B.; Xu, Y.; Yang, J.; and Han, N. 2018b. Adaptive weighted nonnegative low-rank representation. *Pattern Recogn.* 81:326–340.

Wen, J.; Zhang, Z.; Xu, Y.; and Zhong, Z. 2018c. Incomplete multi-view clustering via graph regularized matrix factorization. In *ECCVW*, 593–608.

Xia, R.; Pan, Y.; Du, L.; and Yin, J. 2014. Robust multi-view spectral clustering via low-rank and sparse decomposition. In *AAAI*, 2149–2155.

Xie, G.-S.; Zhang, X.-Y.; Yan, S.; and Liu, C.-L. 2017. Sde: A novel selective, discriminative and equalizing feature representation for visual recognition. *IJCV* 124(2):145–168.

Xu, Y.; Yin, W.; Wen, Z.; and Zhang, Y. 2012. An alternating direction algorithm for matrix completion with nonnegative factors. *Frontiers of Mathematics in China* 7(2):365–384.

Xu, C.; Tao, D.; and Xu, C. 2013. A survey on multi-view learning. *arXiv preprint arXiv:1304.5634*.

Zhan, K.; Zhang, C.; Guan, J.; and Wang, J. 2018. Graph learning for multiview clustering. *IEEE TCYB* 48(10):2887–2895.

Zhang, Z.; Liu, L.; Shen, F.; Shen, H. T.; and Shao, L. 2018. Binary multi-view clustering. *IEEE TPAMI*.

Zhao, H.; Ding, Z.; and Fu, Y. 2017. Multi-view clustering via deep matrix factorization. In *AAAI*, 2921–2927.

Zhao, H.; Liu, H.; and Fu, Y. 2016. Incomplete multi-modal visual data grouping. In *IJCAI*, 2392–2398.