Multi-view Spectral Clustering via Multi-view Weighted Consensus and Matrix-Decomposition Based Discretization

Man-Sheng Chen[1,2], Ling Huang[1,2], Chang-Dong Wang[1,2(✉)],
and Dong Huang[3]

[1] School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China
chenmsh27@mail2.sysu.edu.cn, huanglinghl@hotmail.com,
changdongwang@hotmail.com
[2] Guangdong Province Key Laboratory of Computational Science,
Guangzhou, China
[3] College of Mathematics and Informatics, South China Agricultural University,
Guangzhou, China
huangdonghere@gmail.com

Abstract. In recent years, multi-view clustering has been widely used in many areas. As an important category of multi-view clustering, multi-view spectral clustering has recently shown promising advantages in partitioning clusters of arbitrary shapes. Despite significant success, there are still two challenging issues in multi-view spectral clustering, i.e., (i) how to learn a similarity matrix for multiple weighted views and (ii) how to learn a robust discrete clustering result from the (continuous) eigenvector domain. To simultaneously tackle these two issues, this paper proposes a unified spectral clustering approach based on multi-view weighted consensus and matrix-decomposition based discretization. In particular, a multi-view consensus similarity matrix is first learned with the different views weighted w.r.t. their confidence. Then the eigen-decomposition is performed on the similarity matrix and a set of c eigenvectors are obtained. From the eigenvectors, we first learn a continuous cluster label and then discretize it to build the final clustering label, which avoids the potential instability of the conventional k-means discretization. Extensive experiments have been conducted on multiple multi-view datasets to validate the superiority of our proposed approach.

Keywords: Multi-view spectral clustering · Weighted consensus · Matrix-decomposition · Discretization

1 Introduction

With the development of the information technology [1], a huge amount of multi-view data have emerged from various kinds of real-world applications [2–12].

Multi-view data can be captured from heterogenous views or sources, and these different views or sources reveal the distinct information of the same object. For instance, a YouTube video consists of text features, auditory features and visual features. A text news can be translated into different languages. In traditional multi-view clustering, a straightforward idea to deal with multi-view data is to concatenate all the features into a new feature vector, and then perform single-view clustering method on the new feature vector to obtain the clustering result. However, this simple strategy ignores the different characteristics as well as the correlation among multiple views. The features for multiple views are able to provide complementary information between views. To capture the diversity and correlation in multi-view data, many multi-view clustering algorithms have been developed to improve the robustness of the clustering by making full use of the information from multiple views [13–18].

In the past few years, many multi-view clustering algorithms have been proposed by considering the rich information of multiple views [19–24]. For example, Cai et al. [22] developed a multi-view spectral clustering framework to integrate heterogeneous image features. Kumar et al. [21] introduced the co-regularization technique in multi-view spectral clustering. These methods, however, may be affected by weak or poor views, and thereby result in degraded clustering performances. In multi-view clustering, different views may be associated with very different reliability and should be weighted accordingly. Inspired by the co-training technique [19], Kumar and Daumé III [20] exploited prior knowledge to decide the view weights, and designed a consensus cluster label matrix for multi-view spectral clustering. However, besides the view-weighting issue, another limitation to these existing multi-view spectral clustering methods [21,25,26] is that they mostly rely on the k-means algorithm to perform discretization on the continuous eigenvector domain, where the inherent instability of k-means may significantly affect the final clustering result after discretization.

To simultaneously deal with the issue of view weighting and the issue of potentially unstable discretization of k-means, in this paper, we propose a unified multi-view spectral clustering framework based on multi-view weighted consensus similarity and matrix-decomposition based discretization. Specifically, a consensus similarity matrix is first built with the multiple views evaluated and weighted. Then, a continuous cluster label is learned, from which the final discrete clustering label can be obtained in an optimization model. In the optimization model, we exploit an alternative iteration scheme to achieve an approximate solution. Extensive experiments have been conducted on multiple multi-view datasets, which demonstrate the superiority of our proposed method.

The following sections are organized as follows. In Sect. 2 we describe the proposed model in detail, and present an optimization algorithm to solve the model. Next in Sect. 3, extensive experiments are conducted on four real-world datasets to show the superiority of our method. Finally in Sect. 4, we conclude the whole paper.

Notations. In this paper, uppercase letters are used to represent the matrices. For a matrix M, its i-th row can be written as m_i whose j-th entry is denoted as m_{ij}. $Tr(M)$ stands for the trace of the matrix M. The v-th view of the matrix M is expressed as $M^{(v)}$. We use $\|M\|_2$ and $\|M\|_F$ to respectively represent the l_2-norm and the Frobenius norm of the matrix M. In addition, 1_n means the column vector whose length is n and the entries are all one.

2   The Proposed Algorithm

In this section, we introduce in detail the proposed Multi-view Spectral Clustering via Multi-view Weighted Consensus and Matrix-decomposition based Discretization (MvWCMD) algorithm. First of all, we will briefly introduce the preliminary knowledge. And then we will describe in detail the proposed model, the optimization problem of which will be solved by the alternative iteration scheme. Finally, we will summarize the entire algorithm and provide time complexity analysis.

2.1   Preliminary Knowledge

Graph-Based Clustering Description. Suppose there are n samples which can be partitioned into c categories. To well represent the affinities between these samples, a similarity matrix is supposed to be constructed in a graph-based clustering method. A decent graph plays a vital role therein, therefore it has been studied in many works [27]. When a similarity matrix is ideal, the number of its connected components must be c the same as the number of the final clusters, and it can be directly applied for clustering. Inspired by the idea above, Nie et al. [28] proposed a Constraint Laplacian Rank (CLR) method which aims to learn an ideal graph from the given similarity matrix. Given an arbitrary similarity matrix $A \in \mathbb{R}^{n \times n}$, the target graph can be solved by the following model

$$\min_{s_i 1_n = 1, s_{ij} \geq 0, S \in C} \|S - A\|_F^2, \tag{1}$$

where S is non-negative, and the entries of each row sum up to 1. C indicates a set of n by n square matrices whose connected components are c. In the light of the graph theory in [29], the connectivity constraint can be substituted for a rank constraint, and thus the problem (1) can be rewritten as

$$\min_{s_i 1_n = 1, s_{ij} \geq 0, rank(L) = n - c} \|S - A\|_F^2, \tag{2}$$

where $rank(L)$ stands for the rank of the Laplacian matrix L, and $L = D - \frac{(S^T + S)}{2}$. The n by n degree matrix D is a diagonal matrix, and $D(ii) = \frac{\sum_j (s_{ij} + s_{ji})}{2}$. In this way, the ideal similarity matrix S can be obtained, and thus it can be directly used in clustering. However, the CLR method is just applicable for single-view clustering.

Spectral Clustering Revisit. Looking back on the spectral clustering method [30], data points can be partitioned into different groups according to their similarities. Not requiring data is linearly separable, the method can explore the non-convex pattern. For spectral clustering, Laplacian matrix $L \in \mathbb{R}^{n \times n}$ is required as an input. To obtain the Laplacian matrix L, the similarity matrix $S \in \mathbb{R}^{n \times n}$ is firstly needed to be constructed in traditional spectral clustering methods by one of the three common strategies, such as the k-nearest-neighborhood (knn). Suppose in data X there are c clusters, the spectral clustering problem can be written as

$$\min_{Y} Tr\left(Y^{T} L Y\right), \qquad \text{s.t. } Y \in Ind, \qquad (3)$$

where $Y = [y_1, y_2, ..., y_n]^{T} \in \mathbb{R}^{n \times c}$ is the cluster indicator matrix whose labels are discrete, and $Y \in Ind$ indicates that the cluster label vector of each point $y_i \in \{0, 1\}^{c \times 1}$ only comprises one and only one element "1" to reveal the cluster membership of x_i. Actually, the problem (3) is an NP-hard problem according to the discrete constraint on Y. Thus, the matrix Y is usually relaxed to allow continuous values, and finally the problem becomes

$$\min_{F} Tr\left(F^{T} L F\right), \qquad \text{s.t. } F^{T} F = I, \qquad (4)$$

where $F \in \mathbb{R}^{n \times c}$ is the relaxed continuous cluster label matrix, and the trivial solution can be avoided by the orthogonal constraint therein. And then the approximate solution of F can be achieved by the c eigenvectors of L corresponding to the c smallest eigenvalues. Subsequently, traditional clustering method such as k-means is applied to compute on F to get the final discrete cluster labels [31]. Nevertheless, there still exists potential instability. Due to the uncertainty of the post-processing step, the final solution may deviate from the real discrete labels unpredictably [32].

2.2   The Proposed Model

Motivated by the idea that the spectral embedding matrix F is spanned by the column vectors of the cluster indicator matrix $Y \in Ind$ [31] when the similarity matrix is ideal, we extend the CLR method mentioned above to the multi-view clustering. Despite of this idea, the spectral embedding matrix F is actually not equal to the cluster indicator matrix $Y \in Ind$. Thus, in this paper, not only the spectral embedding matrix can be focused on, but also the cluster indicator matrix can be solved finally without k-means discretization.

In multi-view clustering, the same object represented in different views is expected to be partitioned into the same group. Thus, the ground truth similarity matrix of each view is supposed to be the same. That is to say, there is a consensus similarity matrix among all the views. For multi-view data, suppose that there are m views, and $A^{(1)}$, $A^{(2)}$, ..., $A^{(m)}$ corresponding to the similarity matrix of each view, we aim to get the multi-view consensus similarity matrix S that can well approximate the original input similarity matrix $A^{(v)} \in \mathbb{R}^{n \times n}$ ($1 \leq v \leq m$). A straight-forward solution is to assign the same

weight to every input similarity matrix and achieve an average similarity matrix by the equation $\overline{A} = \frac{1}{m} \sum_{v=1}^{m} A^{(v)}$. However, this simple way ignores the different contributions among views, leading to bad clustering performance when there are poor quality views. Accordingly, a group of meaningful weights are needed to be introduced to measure the importance of different views. In this paper, a trick idea [24] is followed by our algorithm to adaptively measure the weights of the views. Consequently, the target multi-view weighted consensus similarity matrix S with rank constraint is learned to approximate the similarity matrix of each view with different weights. To solve this problem, a linear combination of the reconstruction error $\|S - A^{(v)}\|_F^2$ for each view will be minimized [24]. Thus, the problem can be written as

$$\min_{S} \sum_{v=1}^{m} w^{(v)} \|S - A^{(v)}\|_F^2, \tag{5}$$

$$\text{s.t. } s_i 1_n = 1, s_{ij} \geq 0, rank(L) = n - c,$$

where the constant $w^{(v)}$ is the optimal target function value of the following problem:

$$w^{(v)} \stackrel{def}{=} \min_{S} \frac{1}{\|S - A^{(v)}\|_F}. \tag{6}$$

We can obviously find that $w^{(v)}$ depends on S. If the view v is good, the value of $\|S - A^{(v)}\|_F$ should be small, and therefore $w^{(v)}$ is supposed to be large. Otherwise, a small weight is required to be assigned to a weak view.

Problem (5) is not easy to be solved, due to the rank constraint where $L = D - \frac{(S^T + S)}{2}$ and D is an n by n diagonal matrix whose diagonal elements $D(ii) = \frac{\sum_j (s_{ij} + s_{ji})}{2}$ also depend on the similarity matrix S. Here L is a positive semi-definite matrix, and thus $\sigma_i(L) \geq 0$, where $\sigma_i(L)$ corresponds to the i-th smallest eigenvalue of the Laplacian matrix L. Inspired by [29], $rank(L) = n - c$ is tantamount to $\sum_{i=1}^{c} \sigma_i(L) = 0$. To cope with the optimization question with rank constraint whose complexity analysis is combinatorial, the rank constraint is incorporated into the objective function as a regularizer term [28,33]. Therefore, the constraint is relaxed and our model is reformulated as

$$\min_{S} \sum_{v=1}^{m} w^{(v)} \|S - A^{(v)}\|_F^2 + \alpha \sum_{i=1}^{c} \sigma_i(L), \tag{7}$$

$$\text{s.t. } s_i 1_n = 1, s_{ij} \geq 0.$$

If α is enough large, the minimization of Eq. (7) will make the regularizer term $\sum_{i=1}^{c} \sigma_i(L) \rightarrow 0$. And then the rank constraint $rank(L) = n - c$ will be solved.

Despite all this, problem (7) still remains a challenging problem as a result of the last term. Fortunately, the Ky Fan's Theorem [34] can be applied to solve the problem above, that is to say

$$\sum_{i=1}^{c} \sigma_i(L) = \min_{F^T F = I} Tr\left(F^T L F\right), \tag{8}$$

where $F \in \mathbb{R}^{n \times c}$ is a spectral embedding matrix, and the spectral embedding matrix F is actually not equal to the cluster indicator matrix $Y \in Ind$. To better achieve our clustering task, our multi-view spectral clustering via multi-view weighted consensus and matrix-decomposition based discretization (MvWCMD) model is proposed as follows:

$$
\min_{S,F,Y,Q} \quad \underbrace{\sum_{v=1}^{m} w^{(v)} \|S - A^{(v)}\|_F^2}_{\text{multi-view weighted consensus similarity learning}} \quad +
$$

$$
\underbrace{\alpha Tr\left(F^T L F\right)}_{\text{continuous cluster label learning}} \quad + \quad \underbrace{\beta \|Y - FQ\|_F^2}_{\text{discrete cluster label learning}} \quad ,
$$

$$
\text{s.t. } s_i 1_n = 1, s_{ij} \geq 0, F^T F = I, Q^T Q = I, Y \in Ind, \tag{9}
$$

where α and β are the penalty parameters, and Q is a rotation matrix. Due to the invariance property of spectral solution [35], FQ is another solution for any solution F [36]. The last term expects to find an appropriate orthogonal rotation matrix Q so that the result of FQ is closely approaching to the ground truth discrete cluster label matrix Y. From Eq. (9), the multi-view weighted consensus similarity matrix S, the continuous cluster label matrix F and the final discrete cluster label matrix Y can be automatically learned from the data. Ideally, we must have $s_{ij} = 0$ if data point i and j belong to different groups and vice versa. That is to say, if and only if data point i and j belong to different groups, we have $s_{ij} = 0$ or $f_i \neq f_j$. Therefore, the correlation between the learned similarity matrix and the cluster labels can be exploited in our unified framework Eq. (9). In fact, there is a self-taught property in our clustering model because of the feedback of cluster labels to induce the ideal similarity matrix and vice versa.

2.3   Optimization

In this subsection, an alternative iteration scheme is utilized to solve the problem (9). When updating one variable, the remaining variables will be fixed in the alternative iteration scheme.

Computation of S. With F, Q and Y fixed, the problem is reduced to

$$
\min_{S} \sum_{v=1}^{m} w^{(v)} \|S - A^{(v)}\|_F^2 + \alpha Tr\left(F^T L F\right),
$$
$$
\text{s.t. } s_i 1_n = 1, s_{ij} \geq 0. \tag{10}
$$

In particular, the problem (10) can be further written as

$$
\min_{s_i 1_n = 1, s_{ij} \geq 0} \sum_{v=1}^{m} w^{(v)} \sum_{i,j=1}^{n} \left(s_{ij} - a_{ij}^{(v)}\right)^2 + \alpha \sum_{i,j=1}^{n} \|f_i - f_j\|_2^2 s_{ij}. \tag{11}
$$

Due to the independence of the problem (11) for different i, it is equivalent to separately solving the following problem for each i

$$\min_{s_i 1_n = 1, s_{ij} \geq 0} \sum_{j=1}^{n} \sum_{v=1}^{m} w^{(v)} \left(s_{ij} - a_{ij}^{(v)} \right)^2 + \alpha \sum_{j=1}^{n} \|f_i - f_j\|_2^2 s_{ij}. \tag{12}$$

For briefness, $v_{ij} = \|f_i - f_j\|_2^2$ is used, and v_i is a vector whose j-th entry is v_{ij}. s_i and a_i are in like manner. Thus, the problem (12) becomes

$$\min_{s_i 1_n = 1, s_i \geq 0_n^T} \|s_i - \frac{\sum_{v=1}^{m} w^{(v)} a_i^{(v)} - \frac{\alpha}{2} v_i}{\sum_{v=1}^{m} w^{(v)}}\|_2^2. \tag{13}$$

The problem above can be addressed by an efficient iterative algorithm proposed in [37]. To rapidly obtain the totally sparse multi-view consensus similarity matrix S, the neighbors of the i-th data can be chosen to be updated, and exactly the neighbors can be set as a const, like 10 in our algorithm.

Computation of F. With S, Q and Y fixed, we have

$$\min_{F} \alpha Tr \left(F^T L F \right) + \beta \|Y - FQ\|_F^2, \qquad \text{s.t. } F^T F = I. \tag{14}$$

The problem (14) which is constrained by the orthogonal condition can be settled efficiently by the algorithm proposed by [38].

Computation of Q. With S, F and Y fixed, the problem becomes

$$\min_{Q} \|Y - FQ\|_F^2, \qquad \text{s.t. } Q^T Q = I. \tag{15}$$

This is an orthogonal Procrustes problem [39], which allows a closed-form solution, and the solution is as follows

$$Q = UV^T, \tag{16}$$

where U and V are the left and right components of the SVD decomposition of $Y^T F$.

Computation of Y. With S, F and Q fixed, it is equivalent to solving

$$\min_{Y} \|Y - FQ\|_F^2, \qquad \text{s.t. } Y \in Ind. \tag{17}$$

Knowing that $Tr \left(Y^T Y \right) = n$, the problem above can be reformulated as

$$\max_{Y} Tr \left(Y^T FQ \right), \qquad \text{s.t. } Y \in Ind. \tag{18}$$

Consequently, the optimal solution can be achieved from the following equation

$$Y_{ij} = \begin{cases} 1, & j = \arg\max_{k} (FQ)_{ik} \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

The variables S, F, Q and Y are separately initialized at first. And then they are updated iteratively in an interplay manner until convergence. In this way, an overall optimal solution can be achieved.

2.4   Algorithm Summary and Time Complexity Analysis

For clarity, the main procedure of the proposed MvWCMD method is summarized in Algorithm 1. In what follows, we will provide the time computational complexity analysis. With our optimization strategy, the computation of S requires $\mathcal{O}\left(n^3 + nv\right)$ complexity where $v \ll n$, since it needs to perform eigenvalue decomposition in every iterative step. SVD is involved in the updating of Q, and its computational complexity is $\mathcal{O}\left(nc^2 + c^3\right)$. The complexity for F is $\mathcal{O}\left(nc^2 + c^3\right)$. To update Y, $\mathcal{O}\left(nc^2\right)$ is needed. The number of clusters c is usually a small digit. Therefore, the main computational load of the model in Eq. (9) relies on obtaining the multi-view consensus similarity matrix S.

Algorithm 1. Multi-view Spectral Clustering via Multi-view Weighted Consensus and Matrix-decomposition based Discretization

---

Input: Similarity matrices for m views $A^{(1)}, A^{(2)}, ..., A^{(m)}$ and $A^{(v)} \in \mathbb{R}^{n \times n}$, number of clusters c, parameter $\alpha > 0$, $\beta > 0$
 1: Initialize the weight of each view $w^{(v)} = \frac{1}{m}$, random matrices $F \in \mathbb{R}^{n \times c}$ and $Q \in \mathbb{R}^{n \times n}$, zero matrix $Y \in \mathbb{R}^{n \times c}$.
 2: Let $A = \sum_{v=1}^{m} w^{(v)} A^{(v)}$.
 3: Compute F, which is spanned by the c eigenvectors of $L = D - \frac{A^T + A}{2}$ corresponding to the c smallest eigenvalues.
 4: **repeat**
 5:     **repeat**
 6:         For each i, update the i-th row of S by solving the problem of Eq. (13).
 7:     **until** stopping criterion is met.
 8:     Update F according to Eq. (14).
 9:     Update Q by solving Eq. (16).
10:     $Y = 0$.
11:     Update Y by solving Eq. (19).
12: **until** stopping criterion is met.
Output: $S \in \mathbb{R}^{n \times n}$ with exactly c components, spectral embedding matrix $F \in \mathbb{R}^{n \times c}$, orthogonal rotation matrix $Q \in \mathbb{R}^{n \times n}$ and indicator matrix $Y \in Ind$.

---

3   Experiment

In this section, extensive experiments are conducted to verify the superiority of the proposed method on four real-world datasets. In our experiments, two common evaluation metrics, accuracy (ACC), and normalized mutual information (NMI) are used to estimate the clustering performance of our proposed method and baselines. For each measure, the value is higher, the clustering performance is better [40]. Readers can refer to [41] for further details of the two measures. In addition, parameter analysis, convergence analysis and comparison experiments are separately conducted on the four real-world datasets.

3.1   Real-World Datasets

In our experiment, the four benchmark datasets, UCI Handwritten digits, MSRCv1, Caltech101-7 and Caltech101-20 are used. In the following, we will introduce the details of these datasets.

1. **Handwritten digits dataset**
   Coming from UCI machine learning repository, multiple features (Mfeat) dataset is a handwritten digits dataset[1]. The dataset consists of 2000 samples in which there are 10 classes. In our experiment, three kinds of features, 216 profile correlations, 76 Fourier coefficients and 47 Zernike moments are used to represent images. Each type of features is considered as a view.

2. **MSRCv1 dataset**
   MSRCv1 dataset is an image dataset [42]. The dataset consists of 210 objects and 7 classes. In our experiment, four kinds of features, CM feature, GIST feature, LBP feature and GENT feature are used to represent images, and each type of features is regarded as a view.

3. **Caltech101 datasets**
   Consisting of 101 categories of images, caltech101 [43] is an image dataset. For experimental purpose, two subsets are chosen to represent two datasets following the previous work [25]. The one dataset is named Caltech101-7, and it has 1474 images and 7 widely used classes. The other dataset which is larger is called Caltech101-20, and it is made up of 2386 images and 20 classes. Three types of features, 1984-dimensional HOG feature, 512-dimensional GIST feature and 928-dimensional LBP feature from the images are selected to stand for three views.

The summarization of the four real-world datasets is shown in Table 1.

Table 1. Statistic of the four real-world datasets.

	Mfeat	MSRCv1	Caltech101-7	Caltech101-20
View1	fac(216)	cm(24)	hog(1984)	hog(1984)
View2	fou(76)	gist(512)	gist(512)	gist(512)
View3	zer(47)	lbp(256)	lbp(928)	lbp(928)
View4	-	gent(254)	-	-
# Size	2000	210	1474	2386
# Class	10	7	7	20

3.2   Parameter Analysis

There are two parameters in our model: α and β. In the following, parameter analysis is conducted to show the effect of the two parameters. There are

---

[1] http://archive.ics.uci.edu/ml/index.php.

different properties of different datasets, and thus different ranges of α and β are applied to different datasets. For example, the ranges of α and β are separately $10, 30, 50, 70, 90$ and $0.01, 0.03, 0.05, 0.07, 0.09$ in Mfeat dataset, while the ranges of α and β are separately $1, 3, 5, 7, 9$ and $0.001, 0.003, 0.005, 0.007, 0.009$ in Caltech101-7 dataset. The experimental results are respectively exhibited in Figs. 1, 2, 3 and 4. According to the figures, best results in different datasets can be obtained. For Mfeat dataset, there are the best results when α is 50 and β is 0.01. Similarly, when α is 1 and β is 0.009, best results are achieved for MSRCv1 dataset. In particular, when α is 7 and β is 0.003, the best ACC value can be obtained in Caltech101-7, but the NMI value is lower at this time. To be balanced, the comparatively better results are chosen when α is 9 and β is 0.007 for Caltech101-7. In Caltech101-20 dataset, α is 30 and β is 1 when there are the best results.

(a) ACC   (b) NMI
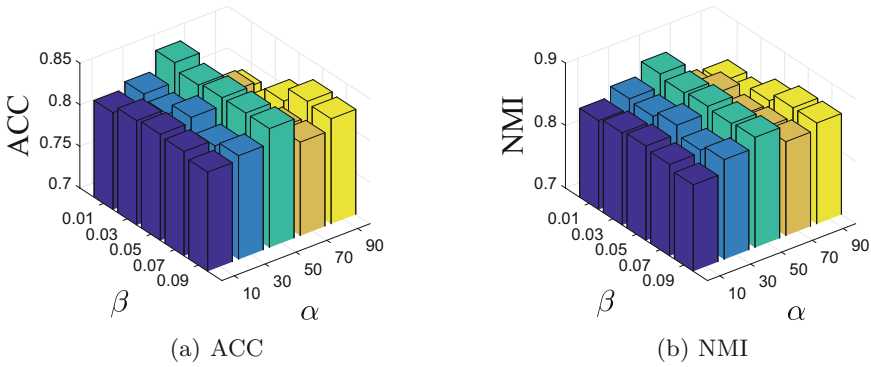
Fig. 1. Parameter analysis on α and β on Mfeat.

(a) ACC   (b) NMI
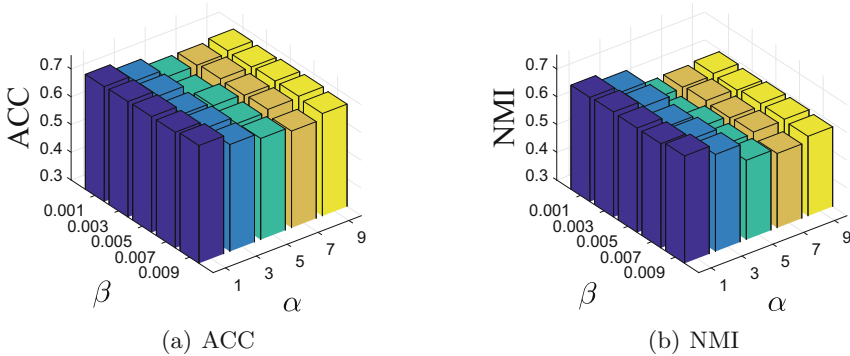
Fig. 2. Parameter analysis on α and β on MSRCv1.

(a) ACC

(b) NMI
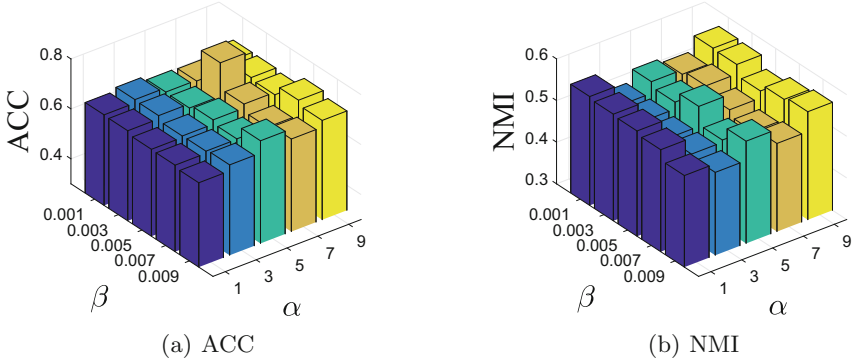
Fig. 3. Parameter analysis on α and β on Caltech101-7.
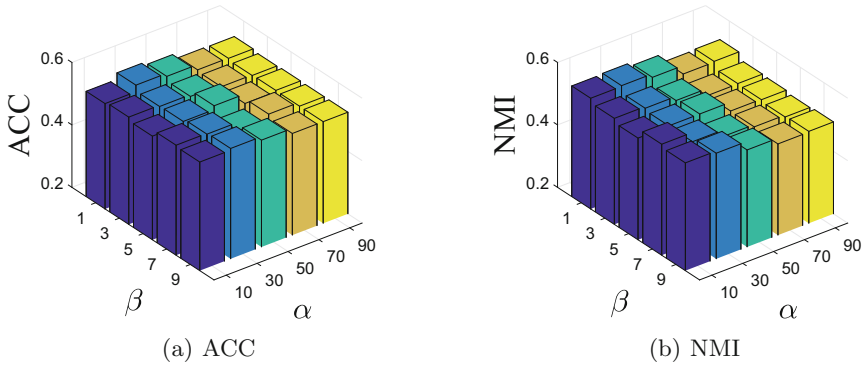

(a) ACC

(b) NMI

Fig. 4. Parameter analysis on α and β on Caltech101-20.

3.3   Convergence Analysis

To verify the convergence property of the proposed method, convergence analysis is conducted. With the best results, the values of α and β from different datasets are set according to the parameter analysis. The experimental results are showed in Fig. 5. Obviously, we can generally conclude that the method will converge during the 30 times of iterations from the subfigures.

3.4   Comparison Experiment

To validate the superiority of the proposed MvWCMD method, we compare our algorithm with the following methods: Constraint Laplacian Rank [28] (CLR), Co-Regularized Spectral Clustering [21] (CoReg), Co-Training Multi-view Clustering [20] (CoTrn), Self-weighted Multi-view Clustering [24] (SwMC), Multi-View Spectral Clustering [22] (MVSC), Robust Multi-view Spectral Clustering [26] (RMSC) and Multi-view Learning with Adaptive Neighbors [44] (MLAN). Following the CLR method, an initial input similarity matrix $A^{(v)}$
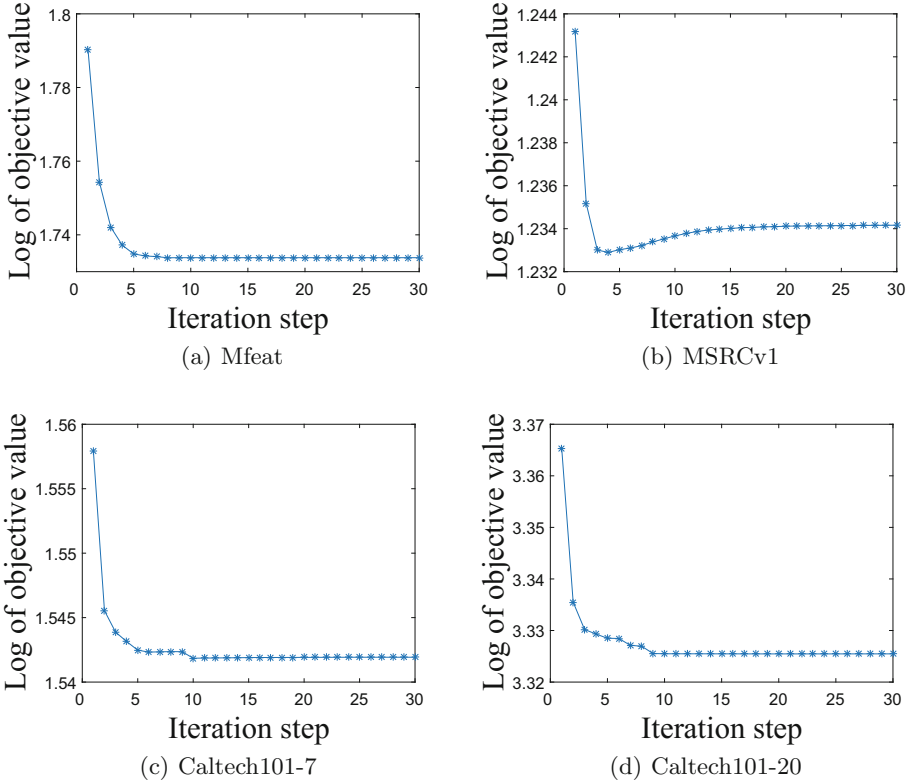
(a) Mfeat

(b) MSRCv1

(c) Caltech101-7

(d) Caltech101-20

Fig. 5. Convergence analysis of optimization.

can be constructed for each view. Only a parameter k that means the number of neighbors is needed to be set in the construction method. For the proposed method, the k is fixed as 10. With the advantage of this graph construction method, the neat normalized similarity matrix of each view is achieved. For all the compared methods, the corresponding parameters are tuned to achieve better performance suggested by the authors. The number of clusters c is set to be equal to the number of the ground truth cluster labels. At the same time, all the methods are conducted for 20 times to avoid the randomness, and the average performance and their standard deviation (std) are computed. The best experimental results will be remarked in bold face.

Tables 2 and 3 show the ACC and NMI results of all algorithms on the four real-world datasets. From the two tables, the proposed algorithm can be seen to obtain the best results among all the state-of-the-art methods in comparison. Thus, our proposed method MvWCMD which jointly learn the multi-view weighted consensus similarity matrix and the cluster label matrix in a unified framework is preferred.

Table 2. Clustering results in terms of ACC on all datasets.

Method	Mfeat	MSRCv1	Caltech101-7	Caltech101-20
CLR	0.7385(0.0000)	0.5524(0.0000)	0.6859(0.0000)	0.4933(0.0000)
CoTrn	0.8002(0.0128)	0.6898(0.0139)	0.4184(0.0082)	0.4283(0.0055)
CoReg	0.7387(0.0095)	0.6241(0.0078)	0.4101(0.0033)	0.3796(0.0033)
SwMC	0.8300(0.0000)	0.5619(0.0000)	0.6635(0.0000)	0.4434(0.0000)
MLAN	0.7640(0.0000)	0.7098(0.0066)	0.6255(0.0000)	0.5270(0.0223)
MVSC	0.8224(0.0511)	0.7205(0.0452)	0.6197(0.0159)	0.4316(0.0329)
RMSC	0.5789(0.0126)	0.3246(0.0098)	0.5369(0.0047)	0.4909(0.0066)
MvWCMD	**0.8479(0.0092)**	**0.7243(0.0332)**	**0.7169(0.0633)**	**0.5585(0.0485)**

Table 3. Clustering results in terms of NMI on all datasets.

Method	Mfeat	MSRCv1	Caltech101-7	Caltech101-20
CLR	0.7609(0.0000)	0.4857(0.0000)	0.5112(0.0000)	0.3795(0.0000)
CoTrn	0.7494(0.0051)	0.6142(0.0090)	0.4145(0.0022)	0.5380(0.0015)
CoReg	0.6949(0.0044)	0.5088(0.0058)	0.4026(0.0027)	0.4884(0.0023)
SwMC	0.8542(0.0000)	0.5639(0.0000)	0.5251(0.0000)	0.4123(0.0000)
MLAN	0.8110(0.0005)	0.6007(0.0134)	0.5482(0.0002)	0.5478(0.0295)
MVSC	0.8393(0.0270)	0.6162(0.0198)	0.5256(0.0206)	0.5505(0.0111)
RMSC	0.5759(0.0142)	0.3103(0.0077)	0.5333(0.0055)	0.5021(0.0064)
MvWCMD	**0.8792(0.0124)**	**0.6868(0.0310)**	**0.5521(0.0178)**	**0.5569(0.0439)**

4 Conclusion

In this work, to eliminate the potential instability from the conventional k-means discretization, we have proposed a novel Multi-view Spectral Clustering via Multi-view Weighted Consensus and Matrix-decomposition based Discretization (MvWCMD) method aiming to jointly learn the multi-view weighted consensus similarity matrix, the continuous cluster label matrix and the final discrete cluster label matrix without k-means discretization. With the help of this framework, variables are updated iteratively in an interplay manner until convergence, so that an overall optimal solution can be achieved. Extensive experiments have been conducted on several real-world datasets to show the superiority of our proposed method.

References

1. Bertino, E.: Introduction to data science and engineering. Data Sci. Eng. **1**(1), 1–3 (2016)
2. Cesa-Bianchi, N., Hardoon, D.R., Leen, G.: Guest editorial: learning from multiple sources. Mach. Learn. **79**(1–2), 1–3 (2010)
3. Chen, N., Zhu, J., Sun, F., Xing, E.P.: Large-margin predictive latent subspace learning for multiview data analysis. IEEE Trans. Pattern Anal. Mach. Intell. **34**(12), 2365–2378 (2012)
4. Gao, Y., Gu, S., Li, J., Liao, Z.: The multi-view information bottleneck clustering. In: Kotagiri, R., Krishna, P.R., Mohania, M., Nantajeewarawat, E. (eds.) DASFAA 2007. LNCS, vol. 4443, pp. 912–917. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-71703-4_78
5. Huang, L., Wang, C.D., Chao, H.Y.: A harmonic motif modularity approach for multi-layer network community detection. In: International Conference on Data Mining (ICDM 2018), pp. 1043–1048 (2018)
6. Zhang, H., Wang, C.D., Lai, J.H., Yu, P.S.: Community detection using multilayer edge mixture model. Knowl. Inf. Syst. (2018). (In press)
7. Li, J.H., Wang, C.D., Li, P.Z., Lai, J.H.: Discriminative metric learning for multi-view graph partitioning. Pattern Recognit. **75**, 199–213 (2018)
8. Sun, Z.R., et al.: Multi-view intact space learning for tinnitus classification in resting state EEG. Neural Process. Lett. **49**, 1–14 (2018)
9. Huang, L., Wang, C.D., Chao, H.Y.: Overlapping community detection in multi-view brain network. In: International Conference on Bioinformatics and Biomedicine (BIBM 2018), pp. 655–658 (2018)
10. Hu, Q.Y., Zhao, Z.L., Wang, C.D., Lai, J.H.: An item orientated recommendation algorithm from the multi-view perspective. Neurocomputing **269**, 261–272 (2017)
11. Hu, Q.Y., Huang, L., Wang, C.D., Chao, H.Y.: Item orientated recommendation by multi-view intact space learning with overlapping. Knowl. Based Syst. **164**, 358–370 (2018)
12. Huang, L., Wang, C.D., Chao, H.Y.: Higher-order multi-layer community detection. In: 33rd AAAI Conference on Artificial Intelligence (AAAI 2019) (2019)
13. Xu, C., Tao, D., Xu, C.: A survey on multi-view learning. CoRR abs/1304.5634 (2013)
14. Lin, K.Y., Wang, C.D., Meng, Y.Q., Zhao, Z.L.: Multi-view unit intact space learning. In: Proceedings of the 10th International Conference on Knowledge Science, Engineering and Management, pp. 211–223 (2017)
15. Zhang, G.Y., Wang, C.D., Huang, D., Zheng, W.S.: Multi-view collaborative locally adaptive clustering with Minkowski metric. Expert Syst. Appl. **86**, 307–320 (2017)
16. Tao, H., Hou, C., Liu, X., Liu, T., Yi, D., Zhu, J.: Reliable multi-view clustering. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-2018), the 30th Innovative Applications of Artificial Intelligence (IAAI-2018), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-2018), New Orleans, Louisiana, USA, 2–7 February 2018, pp. 4123–4130 (2018)
17. Zhang, G.Y., Wang, C.D., Huang, D., Zheng, W.S., Zhou, Y.R.: TW-Co-k-means: two-level weighted collaborative k-means for multi-view clustering. Knowl. Based Syst. **150**, 127–138 (2018)
18. Huang, L., Chao, H.Y., Wang, C.D.: Multi-view intact space clustering. Pattern Recognit. **86**, 344–353 (2019)

19. Blum, A., Mitchell, T.M.: Combining labeled and unlabeled data with co-training. In: Proceedings of the Eleventh Annual Conference on Computational Learning Theory, COLT 1998, Madison, Wisconsin, USA, 24–26 July 1998, pp. 92–100 (1998)
20. Kumar, A., Daumé III, H.: A co-training approach for multi-view spectral clustering. In: Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, 28 June–2 July 2011, pp. 393–400 (2011)
21. Kumar, A., Rai, P., Daumé III, H.: Co-regularized multi-view spectral clustering. In: 25th Annual Conference on Neural Information Processing Systems, Advances in Neural Information Processing Systems 24, Granada, Spain, 12–14 December 2011, pp. 1413–1421 (2011)
22. Cai, X., Nie, F., Huang, H., Kamangar, F.: Heterogeneous image feature integration via multi-modal spectral clustering. In: The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011, pp. 1977–1984 (2011)
23. Xu, Y.M., Wang, C.D., Lai, J.H.: Weighted multi-view clustering with feature selection. Pattern Recognit. **53**, 25–35 (2016)
24. Nie, F., Li, J., Li, X.: Self-weighted multiview clustering with multiple graphs. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, 19–25 August 2017, pp. 2564–2570 (2017)
25. Li, Y., Nie, F., Huang, H., Huang, J.: Large-scale multi-view spectral clustering via bipartite graph. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, Texas, USA, 25–30 January 2015, pp. 2750–2756 (2015)
26. Xia, R., Pan, Y., Du, L., Yin, J.: Robust multi-view spectral clustering via low-rank and sparse decomposition. In: Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, Québec City, Québec, Canada, 27–31 July 2014, pp. 2149–2155 (2014)
27. Zelnik-Manor, L., Perona, P.: Self-tuning spectral clustering. In: Advances in Neural Information Processing Systems 17, Neural Information Processing Systems, NIPS 2004, Vancouver, British Columbia, Canada, 13–18 December 2004, pp. 1601–1608 (2004)
28. Nie, F., Wang, X., Jordan, M.I., Huang, H.: The constrained Laplacian rank algorithm for graph-based clustering. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, Arizona, USA, 12–17 February 2016, pp. 1969–1976 (2016)
29. Mohar, B., Alavi, Y., Chartrand, G., Oellermann, O.: The Laplacian spectrum of graphs. Graph Theory Comb. Appl. **2**(871–898), 12 (1991)
30. von Luxburg, U.: A tutorial on spectral clustering. Stat. Comput. **17**(4), 395–416 (2007)
31. Huang, J., Nie, F., Huang, H.: Spectral rotation versus k-means in spectral clustering. In: Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence, Bellevue, Washington, USA, 14–18 July 2013 (2013)
32. Yang, Y., Shen, F., Huang, Z., Shen, H.T.: A unified framework for discrete spectral clustering. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9–15 July 2016, pp. 2273–2279 (2016)
33. Wang, X., Liu, Y., Nie, F., Huang, H.: Discriminative unsupervised dimensionality reduction. In: Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, 25–31 July 2015, pp. 3925–3931 (2015)

34. Fan, K.: On a theorem of Weyl concerning eigenvalues of linear transformations I. Proc. Nat. Acad. Sci. **35**(11), 652–655 (1949)
35. Yu, S.X., Shi, J.: Multiclass spectral clustering. In: 9th IEEE International Conference on Computer Vision (ICCV 2003), Nice, France, 14–17 October 2003, pp. 313–319 (2003)
36. Kang, Z., Peng, C., Cheng, Q., Xu, Z.: Unified spectral clustering with optimal graph. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-2018), the 30th Innovative Applications of Artificial Intelligence (IAAI-2018), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-2018), New Orleans, Louisiana, USA, 2–7 February 2018, pp. 3366–3373 (2018)
37. Duchi, J.C., Shalev-Shwartz, S., Singer, Y., Chandra, T.: Efficient projections onto the l_1-ball for learning in high dimensions. In: Proceedings of the Twenty-Fifth International Conference on Machine Learning (ICML 2008), Helsinki, Finland, 5–9 June 2008, pp. 272–279 (2008)
38. Wen, Z., Yin, W.: A feasible method for optimization with orthogonality constraints. Math. Program. **142**(1), 397–434 (2013)
39. Schönemann, P.H.: A generalized solution of the orthogonal procrustes problem. Psychometrika **31**(1), 1–10 (1966)
40. Lin, K.-Y., Huang, L., Wang, C.-D., Chao, H.-Y.: Multi-view proximity learning for clustering. In: Pei, J., Manolopoulos, Y., Sadiq, S., Li, J. (eds.) DASFAA 2018. LNCS, vol. 10828, pp. 407–423. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91458-9_25
41. Wang, C.D., Lai, J.H., Yu, P.S.: Multi-view clustering based on belief propagation. IEEE Trans. Knowl. Data Eng. **28**(4), 1007–1021 (2016)
42. Winn, J.M., Jojic, N.: LOCUS: learning object classes with unsupervised segmentation. In: 10th IEEE International Conference on Computer Vision (ICCV 2005), Beijing, China, 17–20 October 2005, pp. 756–763 (2005)
43. Li, F.F., Fergus, R., Perona, P.: Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. Comput. Vis. Image Underst. **106**(1), 59–70 (2007)
44. Nie, F., Cai, G., Li, X.: Multi-view clustering and semi-supervised classification with adaptive neighbours. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, California, USA, 4–9 February 2017, pp. 2408–2414 (2017)