

# Research Article Weighted Nuclear Norm Minimization on Multimodality Clustering

## Lei Du, Songsong Dai, Haifeng Song D, Yuelong Chuang, and Yingying Xu

School of Electronics and Information Engineering, Taizhou University, Taizhou, China

Correspondence should be addressed to Haifeng Song; isshf@126.com

Received 7 December 2020; Revised 24 December 2020; Accepted 6 January 2021; Published 28 January 2021

Academic Editor: Liguo Zhang

Copyright © 2021 Lei Du et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Generally, multimodality data contain different potential information available and are capable of providing an enhanced analytical result compared to monosource data. The way to combine the data plays a crucial role in multimodality data analysis which is worth investigating. Multimodality clustering, which seeks a partition of the data in multiple views, has attracted considerable attention, for example, robust multiview spectral clustering (RMSC) explicitly handles the possible noise in the transition probability matrices associated with different views. Spectral clustering algorithm embeds the input data into a low-dimensional representation by dividing the clustering problem into *k* subproblems, and the corresponding eigenvalue reflects the loss of each subproblem. So, the eigenvalues of the Laplacian matrix should be treated differently, while RMSC regularizes each singular value equally when recovering the low-rank matrix. In this paper, we propose a multimodality clustering algorithm which recovers the low-rank matrix by weighted nuclear norm minimization. We also propose a method to evaluate the weight vector by learning a shared low-rank matrix. In our experiments, we use several real-world datasets to test our method, and experimental results show that the proposed method has a better performance than other baselines.

## 1. Introduction

Clustering, a task of partitioning data points into multiple clusters, is a fundamental research problem in data mining and machine intelligence. A series of algorithms have been proposed over the past decades [1-7]. One of the representative methods is spectral clustering, which has a lot of applications [8-11]. With the development of information and communication technologies, which led to data production in most areas, it is relatively easy to capture features from a given subject. So, it is necessary to design new pattern recognition methods to deal with views of the same subjects. For example, in multilingual information retrieval, the same document can be represented by different languages, and each language can be regarded as a view. These individual views can provide complementary information to each other which can lead to improved performance on the learning task. In the context of multimodality clustering, it seeks to get a better clustering performance by leveraging the information from multiple views.

Many multimodality clustering methods have been proposed in recent years. In general, there are three steps when clustering multiple data  $X = [X^{(1)}, X^{(2)}, \dots, X^{(m)}]$  [12]:

- (1) Obtain a similarity matrix  $S^i$  from each view  $X^i$ , (i = 1, 2, ..., m)
- (2) Compute a projection of each similarity matrix into a space suitable for clustering
- (3) Produce a clustering assignment (i.e., K-means)

The main difference between the multimodality clustering methods lies in the step where the information is collapsed to produce a single new representation. The first category (information merges in Step 1) merges  $S = [S^{(1)}, S^{(2)}, \ldots, S^{(m)}]$  to get a new similarity matrix. The method presented in [13] is a Markov chain method for the generalized normalized cut on multimodality data; the method described in [14] uses the philosophy of co-regularization to make the clustering in different views agree with each other, as described in [15]; RMSC is a Markov-chainbased multimodality spectral clustering method via lowrank and sparse decomposition. The second category (information merges in Step 2) of methods merges the information to generate a compatible projection for all views. In [16], the authors used canonical correlation analysis to maximize the correlation of subjects across the projected views. In the third step, spectral clustering produces the assignment by *K*-means. The assignment is not stable for the randomness of *K*-means, so the third category learns a stable assignment. For example, ensemble clustering [17, 18] methods are designed to find a stable assignment.

The standard nuclear norm minimization regularizes each singular value equally to pursue the convexity of the loss function, while the singular values have different meanings and should be treated differently. Gu [19] proposed a weighted nuclear norm method and applied it to image denoising. The weight vector is evaluated by the singular values of image patches and the noise variance, but it is not useful for multimodality clustering.

In [15], we presented a Markov-chain-based multimodality spectral clustering method via low-rank and sparse decomposition. In this paper, as shown in Figure 1, we extend our previous study by applying the weighted nuclear norm to multimodality clustering and propose a method to evaluate the weight vector. The difference between them is that RMSC recovers the low-rank matrix  $\hat{P}$  by solving a nuclear norm minimization (NNM) problem, while the proposed method recovers that by solving a weighted nuclear norm minimization (WNNM) problem. For the experiments, we use several real-world datasets to test our method. Experimental results show that the proposed method has a better performance than other baselines.

This paper is organized as follows. Section 2 briefly describes the related work from which our method is based on. Section 3 describes the reason for using WNNM on multimodality clustering, defines our algorithm, and presents the optimization procedure. Section 4 presents the results of our method and other multimodality clustering methods. Section 5 outlines the main contributions of the work presented in this paper.

## 2. Related Work

To make this paper clear, Table 1 summarizes the symbols used in this paper.

2.1. Spectral Clustering. Finding good clusters has been a focus of considerable research in pattern recognition. Spectral clustering applies the spectral graph theory [20] which gives the conditions where a graph can be divided into several non-connected subgraphs. The method embeds the input data into a low-dimensional representation and then applies *K*-means.

Here we give the framework of the spectral clustering algorithm [8, 21, 22] (Algorithm 1).



FIGURE 1: Framework of the robust multimodality on spectral clustering via low-rank and sparse decomposition; the proposed method is similar to RMSC [15]; the difference between them is that RMSC recovers the low-rank matrix  $\hat{P}$  by solving a nuclear norm minimization (NNM) problem, while the proposed method recovers that by solving a weighted nuclear norm minimization (WNNM) problem.

TABLE 1: Symbols used in this paper.

XThe whole dataset $X^{(\nu)}$ The $\nu$ -th view dataset $S^{(\nu)}$ The $\nu$ -th similarity matrix $\hat{P}$ Low-rank matrix $P^{(i)}$ The $\nu$ -th transition matrix $E^{(i)}$ The $\nu$ -th noise matrix $m$ Number of views $n$ Number of views $\lambda_i$ The $i$ -th eigenvalue $\sigma_i$ The $i$ -th singular value $w_i$ The $i$ -th weight of $\sigma_i$ $\lambda$ Trade-off parameter	Symbol	Meaning			
$ \begin{array}{llllllllllllllllllllllllllllllllllll$	X	The whole dataset			
$\begin{array}{ccc} S^{(\nu)} & & \text{The $\nu$-th similarity matrix} \\ \widehat{P} & & \text{Low-rank matrix} \\ P^{(i)} & & \text{The $\nu$-th transition matrix} \\ E^{(i)} & & \text{The $\nu$-th noise matrix} \\ m & & \text{Number of views} \\ n & & \text{Number of views} \\ n & & \text{Number of views} \\ \lambda_i & & \text{The $i$-th eigenvalue} \\ \sigma_i & & \text{The $i$-th singular value} \\ w_i & & \text{The $i$-th weight of $\sigma_i$} \\ \lambda & & & \text{Trade-off parameter} \end{array}$	$X^{(\nu)}$	The $\nu$ -th view dataset			
$ \begin{array}{cccc} \widehat{P} & & \text{Low-rank matrix} \\ P^{(i)} & & \text{The $v$-th transition matrix} \\ E^{(i)} & & \text{The $v$-th noise matrix} \\ m & & \text{Number of views} \\ n & & \text{Number of views} \\ n & & \text{Number of instances} \\ \lambda_i & & \text{The $i$-th eigenvalue} \\ \sigma_i & & \text{The $i$-th singular value} \\ w_i & & \text{The $i$-th weight of $\sigma_i$} \\ \lambda & & & \text{Trade-off parameter} \\ \end{array} $	$S^{(\nu)}$	The <i>v</i> -th similarity matrix			
$P^{(i)}$ The v-th transition matrix $E^{(i)}$ The v-th noise matrix $m$ Number of views $n$ Number of instances $\lambda_i$ The i-th eigenvalue $\sigma_i$ The i-th singular value $w_i$ The i-th weight of $\sigma_i$ $\lambda$ Trade-off parameter	$\widehat{P}$	Low-rank matrix			
$E^{(i)}$ The v-th noise matrix $m$ Number of views $n$ Number of instances $\lambda_i$ The i-th eigenvalue $\sigma_i$ The i-th singular value $w_i$ The i-th weight of $\sigma_i$ $\lambda$ Trade-off parameter	$P^{(i)}$	The $\nu$ -th transition matrix			
mNumber of viewsnNumber of instances $\lambda_i$ The <i>i</i> -th eigenvalue $\sigma_i$ The <i>i</i> -th singular value $w_i$ The <i>i</i> -th weight of $\sigma_i$ $\lambda$ Trade-off parameter	$E^{(i)}$	The <i>v</i> -th noise matrix			
$ \begin{array}{ll} n & & \text{Number of instances} \\ \lambda_i & & \text{The $i$-th eigenvalue} \\ \sigma_i & & \text{The $i$-th singular value} \\ w_i & & \text{The $i$-th weight of $\sigma_i$} \\ \lambda & & & \text{Trade-off parameter} \end{array} $	m	Number of views			
$ \begin{array}{ll} \lambda_i & \text{The $i$-th eigenvalue} \\ \sigma_i & \text{The $i$-th singular value} \\ w_i & \text{The $i$-th weight of $\sigma_i$} \\ \lambda & \text{Trade-off parameter} \end{array} $	п	Number of instances			
$\sigma_i$ The <i>i</i> -th singular value $w_i$ The <i>i</i> -th weight of $\sigma_i$ $\lambda$ Trade-off parameter	$\lambda_i$	The <i>i</i> -th eigenvalue			
$w_i$ The <i>i</i> -th weight of $\sigma_i$ $\lambda$ Trade-off parameter	$\sigma_i$	The <i>i</i> -th singular value			
$\lambda$ Trade-off parameter	$w_i$	The <i>i</i> -th weight of $\sigma_i$			
	λ	Trade-off parameter			

2.2. Robust Multimodality Spectral Clustering via Low-Rank and Sparse Decomposition (RMSC). Consider a set of multimodality data  $X = [X^{(1)}, X^{(2)}, ..., X^{(m)}]$  with  $X^{(i)} \in$  $\mathbb{R}^{d^{(i)} \times n}$ , where *m* is the number of views, *n* is the number of data points,  $d^{(i)}$  represents the feature dimension of the *i*-th view, and the *j*-th column  $X_{i}^{(i)}$  in  $X^{(i)}$  represents the features *j*-th data point in the *i*-th view of the (j = 1, 2, ..., n; i = 1, 2, ..., m). The first step of RMSC is using Gaussian kernels to define the similarity matrix, i.e.,  $S_{ii} = \exp(-\|x_i - x_i\|_2^2/\sigma^2)$  where  $\|.\|_2$  denotes the  $\ell_2$  norm and  $\sigma^2$  denotes the standard deviation (e.g., one can set  $\sigma^2$  to be the average Euclidean distance over all pairs of data points). The second step is to construct the transition matrix *P* by  $P = D^{-1}S$  where *D* is a diagonal matrix with  $D_{ii} = d_i = \sum_{i=1}^n S_{ii}$ . Under the low-rank and sparse assumptions, they formulate the transition matrix construction problem as

**Input**:  $X \in \mathbb{R}^{n \times d}$ 

- (1) Construct the similarity matrix S by Gaussian kernel, where  $S_{ij}$  represents the similarity of the *i*-th sample and the *j*-th sample. (2) Compute the normalized symmetrical Laplacian  $L = I - D^{-1/2}SD^{-1/2}$ , where D is a diagonal matrix with  $D_{ii} = \sum_j S_{ij}$ .
- (2) Compute the normalized symmetrical Laplacian  $L = I D^{-11/2}SD^{-11/2}$ , where D is a diagonal matrix with  $D_{ii} = \sum_j S_{ij}$ . (3) Let U be a matrix with columns representing the top k eigenvectors of L.
- (4) Normalize each row of U.
- (5) Run the k-means algorithm on U.
- **Output**: the result of *k*-means.

ALGORITHM 1: Spectral clustering.

$$\min_{\widehat{P}, E^{(i)}} \operatorname{rank} (\widehat{P}) + \lambda \sum_{i=1}^{m} \left\| E^{(i)} \right\|_{0},$$
s.t.  $i = 1, 2, \dots, m, \quad P^{(i)} = \widehat{P} + E^{(i)}, \widehat{P} \ge 0, \widehat{P}1 = 1,$ 

$$(1)$$

where the  $\ell_0$  norm  $||E^{(i)}||_0$  is the number of nonzero elements in  $E^{(i)}$ , rank  $(\hat{P})$  represents the rank of  $\hat{P}$ , 1 is a vector with all ones, and  $\lambda$  is a trade-off parameter. Note that the constraints  $\hat{P} \ge 0$ ,  $\hat{P}1 = 1$  enforce  $\hat{P}$  to be a transition probability matrix, i.e., each of its rows is a probability distribution.

As the problem is nonconvex, they replace rank  $(\hat{P})$  with the trace norm  $\|\hat{P}\|_*$ , and  $\|E^{(i)}\|_0$  with the  $\ell_1$  norm  $\|E^{(i)}\|_1$ , resulting in the following convex optimization problem:

$$\min_{\widehat{P}, E^{(i)}} \|\widehat{P}\|_* + \lambda \sum_{i=1}^m \|E^{(i)}\|_1$$
s.t.  $i = 1, 2, ..., m, \quad P^{(i)} = \widehat{P} + E^{(i)}, \widehat{P} \ge 0, \widehat{P}1 = 1.$ 

$$(2)$$

The  $\ell_1$  norm  $||E^{(i)}||_1 = \sum_{(i,j)} |E_{ij}|$  is well known to be a convex surrogate of  $||E||_0$ . Then, they propose an optimization procedure to solve this problem via the augmented Lagrangian multiplier (ALM) scheme, which has shown its good balance between efficiency and accuracy in many matrix learning problems.

Let  $\sigma_i$  ( $\Sigma$  are in a nonascending order) represent the *i*-th singular value ( $\Sigma = (\sigma_1, \sigma_2, ...)$ ). When updating Q, the subproblem is

$$\min_{Q} \|Q\|_{*} + \frac{\mu}{2} \left\| \widehat{P} - Q + \frac{Z}{\mu} \right\|_{F}^{2}.$$
 (3)

Let  $U\Sigma V^T$  be the SVD form of  $(\hat{P} + t(Z/\mu))$ , and the solution is as follows:

$$Q = U \mathscr{S}_{1/\mu}(\Sigma) V^T, \tag{4}$$

where  $\mathscr{S}_{\delta}(\sigma_i) = \max(\sigma_i - \delta, 0) + \min(\sigma_i + \delta, 0)$  is the shrinkage operator.

In the optimization procedure, each single value adds or subtracts the same value. So, RMSC treats each singular value equally, which may degrade the performance of the result of clustering.

2.3. Weighted Nuclear Norm. Gu et al. [19] studied the weighted nuclear norm minimization (WNNM) problem, where the singular values are assigned different weights. The definition of weighted nuclear norm of a matrix X is as follows:

$$\|X\|_{w,*} = \Sigma_i |w_i \sigma_i(X)|_1.$$
(5)

They analyzed the solutions of the WNNM problem under different weight conditions and proposed a method to evaluate the weight vector according to many image patches when applied the WNNM algorithm to image denoising. The difference between WNNM and our method is as follows: (1) we extend the weighted nuclear norm to multimodality clustering; (2) the methods which evaluated the weight vector were different; the former evaluates the weight vector according to image patches, while our method evaluates that by matrix decomposition.

#### 3. The Proposed Method: Weighted RMSC

In this section, we present how to apply the weighted nuclear norm to multimodality clustering.

As described in Section 2.2, RMSC treats each singular value when updating  $Q(\hat{P} = tQ)$ , while for spectral clustering, different eigenvalues of *L* have different meaning.

According to [23], the RatioCut object function is defined as

RatioCut
$$(A_1, A_2, \dots, A_k) = \sum_{i=1}^k h_i^T L h_i$$
  
$$= \sum_{i=1}^k (H^T L H)_{ii} = \operatorname{Tr}(H^T L H)$$
s.t.  $H^T H = I$ , (6)

where Tr (.) denotes the trace of a matrix and  $h_i$  represents the *i*-th column of *H*.

So, the RatioCut object function is defined as

$$\min_{H} \operatorname{Tr}(H^{T}LH)$$
s.t.  $H^{T}H = I.$ 
(7)

According to Rayleigh–Ritz theorem [24] the problem has a fixed solution and H is constructed by the top k eigenvectors of L. From equation (6), we can find that the normalized spectral clustering divides the problem into ksubproblems. Each subproblem partitions the points into 2 clusters, and  $h_i$  (i = 2, ..., k) is the solution of the subproblem ( $h_1$  is an all-one vector, which represents it dividing all the points into the same cluster; this partition is useless). As  $H^T H = I$ , the RatioCut object function can be rewritten as

RatioCut
$$(A_1, A_2, \dots, A_k) = \sum_{i=1}^k h_i^T L h_i = \sum_{i=1}^k \lambda_i$$
 (8)  
s.t.  $H^T H = I$ .

So, we can find that the loss of each subproblem has a relation with the corresponding eigenvalue; the small eigenvalue reflects the little loss of the subproblem, so the smaller the eigenvalue  $\lambda$  is, the larger the weight of corresponding eigenvector should be assigned. It is also known that

$$L = I - \widehat{P}.$$
 (9)

So, the larger eigenvalues (or singular values,  $\sigma_i = \lambda_i$ ) of  $\hat{P}$  are more important than the smaller ones when updating  $\hat{P}$  in RMSC; the larger the eigenvalues, the less they should be shrunk. Therefore, the weight assigned to  $\sigma_i$  should be inversely proportional to  $\sigma_i$ . We let

$$w_i = \frac{c}{\left(\sigma_i + \epsilon\right)},\tag{10}$$

where c > 0 is a constant;  $\epsilon = 10^{-16}$  is to avoid dividing by zero.

For multimodality clustering, we can construct *m* Laplacian matrices, leading to *m* groups of  $\sigma_i$ , which are not equal. So evaluating accurate  $\sigma_i$  is a challenging procedure, leading to the difficulty to make sure the weight of singular value. As we know, the output of RMSC is a shared low-rank matrix, and all the views share the same singular values. So, one way to evaluate the singular values is making use of other multimodality clustering algorithms, such as RMSC, to get the shared Laplacian matrix and then evaluating the final  $\Sigma$  by the singular values of the shared Laplacian matrix.

Following RMSC, under the low-rank and sparse assumptions, we formulate the transition matrix construction problem as follows:

$$\min_{\widehat{P}, E^{(i)}} \|\widehat{P}\|_{w,*} + \lambda \sum_{i=1}^{m} \|E^{(i)}\|_{1}$$
s.t.  $i = 1, 2, ..., m, P^{(i)} = \widehat{P} + E^{(i)}, \widehat{P} \ge 0, \widehat{P}1 = 1.$ 

$$(11)$$

The optimization problem (11) is still challenging because the matrix  $\hat{P}$  has two constraints. We introduce an auxiliary variable Q to solve this problem. The optimization problem (11) becomes as follows:

$$\min_{\widehat{P},Q,E^{(i)}} \|Q\|_{w,*} + \lambda \sum_{i=1}^{m} \|E^{(i)}\|_{1}$$
s.t.  $i = 1, 2, ..., m, P^{(i)} = \widehat{P} + E^{(i)},$ 

$$\widehat{P} \ge 0, \quad \widehat{P}1 = 1, \widehat{P} = Q.$$

$$(12)$$

The corresponding augmented Lagrange function of (12) is

$$\mathscr{L}(\hat{P}, tQn, qE^{(i)}) = \|Q\|_{w,*} + \lambda \sum_{i=1}^{m} \|E^{(i)}\|_{1} + \sum_{i=1}^{m} \langle Y^{(i)}, \hat{P} + E^{(i)} - P^{(i)} \rangle + \frac{\mu}{2} \sum_{i=1}^{m} \|\hat{P} + E^{(i)} - P^{(i)}\|_{F}^{2} + \langle Z, \hat{P} - Q \rangle + \frac{\mu}{2} \|\hat{P} - Q\|_{F}^{2} \quad \text{s.t. } \hat{P} \ge 0, \hat{P}1 = 1,$$
(13)

m

where Z,  $Y^{(i)}$  represent the Lagrange multipliers,  $\langle \cdot, \cdot \rangle$  denotes the inner product of matrices (i.e., for two matrices A and B,  $\langle A, B \rangle = A^T B$ ), and  $\mu > 0$  is an adaptive penalty parameter.

The sketch of the proposed algorithm is shown in Algorithm 2. Next we will present the update rules for each of  $\hat{P}$ , Q, and  $E^{(i)}$ .

When other variables are fixed, the subproblem with respect to Q is

$$\min_{Q} \|Q\|_{w,*} + \frac{\mu}{2} \|\widehat{P} - Q + \frac{Z}{\mu}\|_{F}^{2}.$$
 (14)

More specifically, let  $U\Sigma V^T$  be the SVD form of  $(\hat{P} + t(Z/\mu))$ . We use RMSC to evaluate the final  $\Sigma$  and use it to evaluate W via Equation (10). According to [19], the solution to (14) is as follows:

$$Q = U \mathscr{S}_W(\Sigma) V^T. \tag{15}$$

The subproblem with respect to  $E^{(i)}$  (i = 1, 2, ..., m) can be simplified as

$$\min_{E^{(i)}} \lambda \left\| E^{(i)} \right\|_{1} + \frac{\mu}{2} \left\| E^{(i)} - \left( P^{(i)} - \widehat{P} - \frac{Y^{(i)}}{\mu} \right) \right\|_{F}^{2}, \quad (16)$$

which has a closed form solution  $E^{(i)} = S_{\lambda/\mu} (P^{(i)} - \hat{P} - (Y^{(i)}/\mu)).$ 

With other variables being fixed, we update  $\hat{P}$  by solving

$$\widehat{P} = \operatorname{argmin}_{\widehat{P}2} \sum_{i=1}^{m} \left\| \widehat{P} + E^{(i)} - P^{(i)} + \frac{Y^{(i)}}{\mu} \right\|_{F}^{2} + \frac{\mu}{2} \left\| \widehat{P} - Q + \frac{Z}{\mu} \right\|_{F}^{2} \quad \text{s.t. } \widehat{P} \ge 0, \widehat{P}1 = 1.$$
(17)

The solution is given by RMSC, which can be decomposed into n independent subproblems. Each subproblem is a proximal operator problem with probabilistic simplex constraint, which can be efficiently solved by the projection algorithm.

#### 4. Experimental Setup

The proposed method was tested on several real-world datasets; the details are shown in Table 2.

In all the experiments, we use six metrics to measure the clustering performances: *F*-score, precision,

```
Input: \lambda, P^{(i)} \in \mathbb{R}^{n \times n} (i = 1, 2, ..., m).
     Initialize: \widehat{P} = 0, Q = 0, Z = 0, Y^{(i)} = 0, E^{(i)} = 0, \mu = 10^{-6}, \rho = 1.9, \max_{\mu} = 10^{10}, \epsilon = 10^{-8}.
     Evaluate W by running RMSC.
     Repeat
(1) Let C \leftarrow (1/(m+1))(Q - (Z/\mu) + \sum_{i=1}^{m} (P^{(i)} - E^{(i)} - (Y^{(i)}/\mu))).
(2) For j = 1, 2, ..., n
         Update \hat{P}_i
(3) For i = 1, 2, \dots, m
         Update E^{(i)} via equation (16).
(4) Update Q via equation (15).
(5) Set Z \leftarrow Z + \mu (\widehat{P} - tQ).
(6) For i = 1, 2, ..., m
         Set Y^{(i)} \longleftarrow Y^{(i)} + \mu (\widehat{P} + E^{(i)} - P^{(i)}).
(7) Set \mu \leftarrow \min(\rho\mu, \max_{\mu}).
     Until min (\|\widehat{P} + E^{(i)} - P^{(i)}\|_{\infty}, \|\widehat{P} - Q\|_{\infty}) \le \epsilon.
     Output: \hat{P}, E^{(i)} (i = 1, 2, ..., m).
```

ALGORITHM 2: Weighted nuclear norm on robust multimodality clustering.

TABLE 2: Statistics of the real-world datasets.

Dataset	Views	Instances	Clusters
BBCSports	2	544	5
WebKb	2	1051	2
Reuters	5	600	6
UCI	3	2000	10

recall, normalized mutual information (NMI) [25], entropy, and adjusted rand index (Adj-RI) [26]. Note that higher values indicate better performance except for entropy.

When evaluating the weight vector, there is a constant parameter *c*. We set c = 0.00001 in all the experiments. Similarity matrices are constructed by Gaussian kernels.  $\sigma^2$  is set to the median of the Euclidean distance between every pair of data points for all of the datasets except BBCSports ( $\sigma^2 = 100$ ).  $\lambda$  is set to be 0.005.

#### **5. Experimental Results**

We chose the following six multimodality clustering algorithms as baselines:

- (1) Single view: performing spectral clustering on a single view.
- (2) Feature concatenation: concatenating all the features of each view and then performing spectral clustering on the new representation.
- (3) Kernel addition: constructing the similarity matrices from each view and then averaging all the matrices to obtain a new similarity matrix.
- (4) Mixture of Markov chains (MMC): a mixture of Markov chains defined on each view [13].
- (5) Co-regularized spectral clustering (Co-Reg): making use of the philosophy of co-regularization to make the clustering in different views agree with each other [14].

(6) Robust multiview spectral clustering via low-rank and sparse decomposition (RMSC): a Markov-chainbased multimodality spectral clustering method via low-rank and sparse decomposition [15].

Following the settings in [14], we use the Gaussian kernel to construct similarity matrix for each view if needed in all algorithms.

Table 3 shows the results of the proposed method and the baselines on BBCSports. As can be seen, the proposed method shows superior performance gains over the baselines with respect to all the six metrics. Here are some statistics: the results of our method indicate a relative increase of 4.82%, 2.88%, 2.10%, and 6.58% with respect to *F*-score, precision, NMI, and Adj-RI, respectively, compared to the corresponding second best baseline.

Table 4 shows the results of the proposed method and the baselines on UCI. As can be seen, the proposed method shows superior performance gains over the baselines with respect to all the six metrics. Here are some statistics: the results of our method indicate a relative increase of 1.35%, 1.62%, 1.32%, and 1.647.59% with respect to *F*-score, precision, NMI, and Adj-RI, respectively, compared to the corresponding second best baseline.

Table 5 shows the results of the proposed method and the baselines on WebKb. As can be seen, the proposed method shows superior performance gains over the baselines with respect to most of the six metrics. Here are some statistics: the results of our method indicate a relative increase of 0.32%, 2.46%, and 0.82% with respect to *F*-score, NMI, and Adj-RI, respectively, compared to the corresponding second best baseline. Although the precision value of the proposed method is lower than that of kernel addition, the difference is small.

Table 6 shows the results of the proposed method and the baselines on Reuters. As can be seen, the proposed method shows superior performance gains over the baselines with respect to all the six metrics. Here are some statistics: the results of our method indicate a relative increase of 1.89%, 3.56%, 5.85%, and 4.74% with respect to

TABLE 3: Comparison results on BBCSports.

Method	<i>F</i> -score	Precision	Recall	NMI	Entropy	Adj-RI
SC view 1	0.767 (0.003)	0.785 (0.011)	0.750 (0.017)	0.717 (0.004)	0.609 (0.021)	0.696 (0.002)
SC view 2	0.410 (0.008)	0.319 (0.028)	0.589 (0.066)	0.218 (0.020)	1.772 (0.060)	0.146 (0.038)
Feature concat	0.669 (0.019)	0.654 (0.017)	0.688 (0.049)	0.618 (0.023)	0.853 (0.039)	0.563 (0.020)
Kernel addition	0.672 (0.021)	0.657 (0.022)	0.691 (0.049)	0.621 (0.029)	0.847 (0.063)	0.566 (0.024)
MMC	0.788 (0.000)	0.825 (0.000)	0.755 (0.000)	0.728 (0.000)	0.571 (0.000)	0.726 (0.000)
Co-Reg	0.766 (0.763)	0.788(0.780)	0.745 (0.747)	0.718 (0.712)	0.603 (0.621)	0.695 (0.691)
RMSC	0.851 (0.056)	0.868 (0.027)	0.836 (0.081)	0.810 (0.026)	0.408 (0.034)	0.806 (0.070)
Ours	0.892 (0.003)	0.893 (0.005)	0.891 (0.002)	0.827 (0.004)	0.380 (0.011)	0.859 (0.004)

TABLE 4: Comparison results on UCI.

Method	<i>F</i> -score	Precision	Recall	NMI	Entropy	Adj-RI
SC view 1	0.574 (0.029)	0.563 (0.028)	0.585 (0.032)	0.632 (0.019)	1.231 (0.063)	0.526 (0.033)
SC view 2	0.583 (0.023)	0.574 (0.028)	0.591 (0.018)	0.643 (0.013)	1.191 (0.045)	0.536 (0.026)
SC view 3	0.377 (0.004)	0.364 (0.011)	0.392 (0.009)	0.483 (0.005)	1.732 (0.024)	0.306 (0.006)
Feature concat	0.448 (0.014)	0.433 (0.020)	0.466 (0.013)	0.552 (0.016)	1.506 (0.058)	0.385 (0.017)
Kernel addition	0.746 (0.024)	0.723 (0.040)	0.773 (0.015)	0.784 (0.014)	0.736 (0.055)	0.717 (0.027)
MMC	0.729 (0.052)	0.708 (0.066)	0.752 (0.040)	0.771 (0.032)	0.778 (0.116)	0.697 (0.059)
Co-Reg	0.594 (0.650)	0.585 (0.634)	0.603 (0.668)	0.644 (0.687)	1.189 (1.051)	0.548 (0.610)
RMSC	0.816 (0.061)	0.805 (0.073)	0.828 (0.049)	0.832 (0.038)	0.565 (0.136)	0.795 (0.069)
Ours	0.827 (0.041)	0.818 (0.052)	0.836 (0.029)	0.843 (0.022)	0.528 (0.082)	0.808 (0.046)

TABLE 5: Comparison results on WebKb.

Method	<i>F</i> -score	Precision	Recall	NMI	Entropy	Adj-RI
SC view 1	0.592 (0.000)	0.645 (0.000)	0.546 (0.000)	0.167 (0.000)	0.618 (0.000)	0.028 (0.000)
SC view 2	0.889 (0.000)	0.824 (0.000)	0.965 (0.000)	0.532 (0.000)	0.406 (0.000)	0.618 (0.000)
Feature concat	0.896 (0.000)	0.836 (0.000)	0.966 (0.000)	0.559 (0.000)	0.384 (0.000)	0.648 (0.000)
Kernel addition	0.947 (0.000)	0.947 (0.000)	0.947 (0.000)	0.718 (0.000)	0.214 (0.000)	0.845 (0.000)
MMC	0.372 (0.017)	0.338 (0.017)	0.416 (0.029)	0.314 (0.025)	1.801 (0.063)	0.232 (0.020)
Co-Reg	0.365 (0.365)	0.334 (0.332)	0.405 (0.406)	0.312 (0.312)	1.804 (1.803)	0.225 (0.223)
RMSC	0.951 (0.000)	0.946 (0.000)	0.957 (0.000)	0.734 (0.000)	0.206 (0.000)	0.856 (0.000)
Ours	0.954 (0.000)	0.940 (0.000)	0.970 (0.000)	0.752 (0.000)	0.201 (0.000)	0.863 (0.000)

TABLE 6: Comparison results on Reuters.

Method	<i>F</i> -score	Precision	Recall	NMI	Entropy	Adj-RI
SC view 1	0.368 (0.017)	0.337 (0.009)	0.409 (0.042)	0.318 (0.017)	1.792 (0.031)	0.229 (0.014)
SC view 2	0.361 (0.016)	0.315 (0.019)	0.426 (0.030)	0.315 (0.028)	1.811 (0.070)	0.211 (0.021)
SC view 3	0.341 (0.005)	0.306 (0.008)	0.385 (0.012)	0.279 (0.015)	1.888 (0.039)	0.192 (0.007)
SC view 4	0.345 (0.014)	0.307 (0.014)	0.394 (0.014)	0.271 (0.016)	1.910 (0.041)	0.196 (0.018)
SC view 5	0.346 (0.014)	0.308 (0.016)	0.397 (0.022)	0.269 (0.017)	1.916 (0.042)	0.197 (0.018)
Feature concat	0.367 (0.014)	0.327 (0.013)	0.419 (0.028)	0.316 (0.023)	1.799 (0.055)	0.222 (0.016)
Kernel addition	0.371 (0.019)	0.333 (0.017)	0.419 (0.027)	0.313 (0.028)	1.806 (0.070)	0.228 (0.023)
MMC	0.367 (0.016)	0.334 (0.018)	0.408 (0.021)	0.308 (0.019)	1.812 (0.048)	0.227 (0.020)
Co-Reg	0.367 (0.016)	0.334 (0.018)	0.408 (0.021)	0.308 (0.019)	1.812 (0.048)	0.227 (0.020)
RMSC	0.372 (0.019)	0.338 (0.017)	0.415 (0.022)	0.325 (0.021)	1.770 (0.053)	0.232 (0.023)
Ours	0.379 (0.021)	0.350 (0.009)	0.413 (0.011)	0.344 (0.021)	1.719 (0.015)	0.243 (0.016)

*F*-score, precision, NMI, and Adj-RI, respectively, compared to the corresponding second best baseline. Although the recall value of the proposed method is lower than that of kernel addition and feature concatenation, the difference is small.

# 6. Conclusion

With the development of information and communication technologies, it is necessary to design new pattern recognition methods to deal with views of the same subjects. It is a challenge task to deal with multimodality problems. Inspired by the previous work, we proposed a method applying the weighted nuclear norm to RMSC and gave a method to evaluate the weight vector, which distinguishes different single values. To solve the optimization problem, we designed a procedure based on ALM. To evaluate the proposed method, we apply it to four real-world datasets. Experimental results show that the proposed method has a superior performance than other baselines. In the future, we will continue the studies in multimodality clustering, including evaluating the weight vector more accurately and clustering on the large-scale datasets.

## **Data Availability**

The BBCSports dataset used to support the findings of this study has been deposited in the UCD repository (http://mlg.ucd.ie/datasets/bbc.html) and the other datasets have been deposited in the UCI repository (http://archive.ics.uci.edu/ml/index.php).

## **Conflicts of Interest**

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research was supported by the Cultivating Science Foundation of Taizhou University (2019PY014), the Agricultural Science and Technology Project of Taizhou (20ny13), the Zhejiang Provincial Natural Science Foundation of China (LQ21F020001 and LQ21A010001), and Taizhou Science and Technology Project (1901gy20).

## References

- X.-H. Kuang, L. Liu, Q. Liu, and X. Li, "A clustering approach based on convergence degree chain for wireless sensor networks," *Security and Communication Networks*, vol. 8, no. 10, pp. 1878–1889, 2015.
- [2] M. A. Azad, R. Morla, J. Arshad, and K. Salah, "Clustering voip caller for spit identification," *Security and Communication Networks*, vol. 9, no. 18, pp. 4827–4838, 2016.
- [3] "Clustering and splitting charging algorithms for large scaled wireless rechargeable sensor networks," *Journal of Systems* and Software, vol. 113, pp. 381–394, 2016.
- [4] M.-C. Su and C.-H. Chou, "A modified version of the k-means algorithm with a distance based on cluster symmetry," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 674–680, 2001.
- [5] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation," *IEEE Transactions* on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 881–892, 2002.
- [6] L. Kaufman and P. J. Rousseeuw, Finding groups in data: an introduction to cluster analysis, John Wiley & Sons, Hoboken, NJ, USA, 2009.
- [7] A. K. Jain, "Data clustering: 50 years beyond k-means," Pattern Recognition Letters, vol. 31, no. 8, pp. 651–666, 2010.

- [8] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.
- [9] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR 2009, pp. 2790–2797, Miami, FL, USA, June 2009.
- [10] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10*, pp. 663–670, Haifa, Israel, June 2010.
- [11] J. Feng, Z. Lin, H. Xu, and S. Yan, "Robust subspace segmentation with block-diagonal prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3818–3825, Seattle, WA, USA, June 2014.
- [12] S. Kanaan-Izquierdo, A. Ziyatdinov, and A. Perera-Lluna, "Multiview and multifeature spectral clustering using common eigenvectors," *Pattern Recognition Letters*, vol. 102, pp. 30–36, 2018.
- [13] D. Zhou and C. J. Burges, "Spectral clustering and transductive learning with multiple views," in *Proceedings of the* 24th International Conference on Machine Learning, ACM, pp. 1159–1166, 2007.
- [14] A. Kumar, P. Rai, and H. Daume, "Co-regularized multi-view spectral clustering," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 1413–1421, Lake Tahoe, NV, USA, December 2011.
- [15] R. Xia, Y. Pan, L. Du, and J. Yin, "Robust multi-view spectral clustering via low-rank and sparse decomposition," in *Proceedings of the AAAI*, pp. 2149–2155, New York, NY, USA, February 2014.
- [16] K. Chaudhuri, S. M. Kakade, K. Livescu, and K. Sridharan, "Multi-view clustering via canonical correlation analysis," in *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 129–136, ACM, Montreal, Quebec, Canada, June 2009.
- [17] S. Vega-Pons and J. Ruiz-Shulcloper, "A survey of clustering ensemble algorithms," *International Journal of Pattern Rec*ognition and Artificial Intelligence, vol. 25, no. 3, pp. 337–372, 2011.
- [18] X. Cai, F. Nie, and H. Huang, "Multi-view k-means clustering on big data," in *Proceedings of the IJCAI*, pp. 2598–2604, New York, NY, USA, August 2013.
- [19] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2862–2869, Seattle, WA, USA, June 2014.
- [20] S. Butler and F. Chung, "Spectral graph theory," *Handbook of linear algebra*, p. 47, CRC Press, Boca Raton, FL, USA, 2006.
- [21] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.
- [22] A. Y. Ng, M. I. Jordan, Y. Weiss et al., "On spectral clustering: analysis and an algorithm," *NIPS*, vol. 14, pp. 849–856, 2001.
- [23] U. Luxburg, "A tutorial on spectral clustering," Statistics and Computing, vol. 17, no. 4, pp. 395–416, 2007.
- [24] H. Lutkepohl, "Handbook of matrices," Computational Statistics and Data Analysis, vol. 2, no. 25, p. 243, 1997.
- [25] H.-C. Huang, Y.-Y. Chuang, and C.-S. Chen, "Multiple kernel fuzzy clustering," *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 1, pp. 120–134, 2012.
- [26] Y. Zhao and G. Karypis, "Empirical and theoretical comparisons of selected criterion functions for document clustering," *Machine Learning*, vol. 55, no. 3, pp. 311–331, 2004.