Robust Auto-Weighted Multi-View Clustering

Pengzhen Ren, Yun Xiao*, Pengfei Xu, Jun Guo, Xiaojiang Chen, Xin Wang, Dingyi Fang

School of Information Science and Technology, Northwest University, Xi'an 710127, P.R. China

pengzhenr@stumail.nwu.edu.cn, yxiao@nwu.edu.cn, xpfu@nwu.edu.cn, guojun@nwu.edu.cn, xjchen@nwu.edu.cn, xinwang@nwu.edu.cn, dyf@nwu.edu.cn

Abstract

Multi-view clustering has played a vital role in real-world applications. It aims to cluster the data points into different groups by exploring complementary information of multi-view. A major challenge of this problem is how to learn the explicit cluster structure with multiple views when there is considerable noise. To solve this challenging problem, we propose a novel Robust Auto-weighted Multi-view Clustering (RAMC), which aims to learn an optimal graph with exactly k connected components, where k is the number of clusters. ℓ_1-norm is employed for robustness of the proposed algorithm. We have validated this in the later experiment. The new graph learned by the proposed model approximates the original graphs of each individual view but maintains an explicit cluster structure. With this optimal graph, we can immediately achieve the clustering results without any further post-processing. We conduct extensive experiments to confirm the superiority and robustness of the proposed algorithm.

1 Introduction

With the advent of the Internet and big data era, it is very common for many real-world applications to deal with multi-view data. The existence of such multi-view data raised the interest of multi-view learning [Xu *et al.*, 2013; Sun, 2013; Tao *et al.*, 2018]. To make full use of the information of multi-view data to improve clustering accuracy, multi-view clustering (MVC) has become an important research topic in the past two decades [Chao *et al.*, 2017]. MVC is a machine learning paradigm to classify the similar subjects into the same group and dissimilar subjects into different groups by combining the available multi-view information, which indicates that MVC searches for the consistent clusterings across different view [Chao *et al.*, 2017].

In the traditional multi-view data research [Hagen and Kahng, 1992; Blum and Mitchell, 1998], clustering is performed by building a weighted graph in which nodes correspond to data points and weights are related to the similarity between the points. The higher the similarity is, the greater the weight is, and vice versa. Each graph represents a view of the research object. Using the weighted graph, MVC can be solved with spectral clustering method [von Luxburg, 2007].

There is a straightforward way to convert multi-view into a single view so that the spectral clustering method can be used for multi-view clustering. However, this simple method ignores the connection between each view and lacks corresponding theoretical support.

Some researchers also seek breakthroughs in other directions of clustering [Chang *et al.*, 2016; Kang *et al.*, 2017a; 2017b]. [Cai *et al.*, 2011] proposed a multi-view spectral clustering model and [Kumar *et al.*, 2011] introduced the co-regularization technique into the spectral clustering model to perform clustering tasks. However, these methods are easily affected by the poor quality view. [Kumar *et al.*, 2011; Li *et al.*, 2015] approximated the weight of data nodes by combining prior knowledge. But this approach, due to the introduction of human intervention, is difficult to ensure that it still works when faced with big data tasks. In addition, in the face of specific clustering tasks, most researchers focus on spectral clustering to deal with it [Kumar *et al.*, 2011; Xia *et al.*, 2014; Li *et al.*, 2015; Chang *et al.*, 2015; Xue *et al.*, 2017]. They tried to optimize and improved on the basis of spectral clustering in order to obtain better clustering performance. However, there is an obvious uncertainty in these improved methods based on spectral clustering, which requires post-processing (e.g., K-means) to eventually generate a labeling matrix.

Recently, [Nie *et al.*, 2016] proposed a Constrained Laplacian Rank (CLR) method based on Laplacian matrix rank constraints. Combining two measurements, Frobenius norm and ℓ_1-norm, this method solves MVC problem by minimizing the distance between the eventually learned similarity matrix and the similarity matrix of the various view. The adjacency matrix with just connected components is obtained by the minimum distance, and the number of connected components is exactly equal to the number of clustering in the original multi-view data. The adjacency matrix can be used directly to complete the clustering task. It can avoid the uncertainty caused by the post-processing of the traditional spectral clustering algorithm. However, the CLR method is only applicable to a single view clustering. How to obtain the explicit cluster structure with multiple noisy views has not been

---

* Corresponding author.

solved yet.

In this paper, we propose a novel Robust Auto-weighted Multi-view Clustering (RAMC) that aims to learn a *consensus* graph with exactly k connected components, where k is the number of clusters. In this way, the obtained graph can be directly used for clustering without the need of further post-processing. To make sure the graph has exactly k connected components, we explicitly impose a low-rank constraint on the objective function. ℓ_l-norm has been employed in this paper for robustness. In addition, by solving each subproblem with an optimal solution through the alternating optimization steps, the proposed RAMC gives the concrete steps to solve the multi-view clustering. Furthermore, experiments are conducted on two synthetic datasets containing noise and three actual datasets, and the experimental results confirm the superiority of the proposed algorithm. The robustness performance of the proposed RAMC shows that as the ratio of random noise increases, although the performance of all the compared algorithms decreases, the performance gain of the proposed algorithm becomes more significant.

The contribution of this paper briefly summarized as follows:

- The proposed algorithm is able to learn a *consensus* graph with exactly k connected components (k is the number of clusters) by exploring the mutual information among multiple views. The obtained graph can be directly used for clustering without any further post-processing.

- The proposed algorithm is robust to the outliers since we employ the ℓ_1-norm. We have validated this in the later experiment.

- Since the algorithm is non-smooth and difficult to solve, we propose an efficient iterative method with guaranteed convergence to optimize it.

- We conduct extensive experiments to verify the performance of the proposed algorithm. The experimental results confirm its superiority compared to state-of-the-art alternatives.

Notation. In the whole article, every matrix is written in capital letters. In a matrix A, a_i represents the i-th row, and a_{ij} represents the element corresponding to the i-th row and the j-th column. The trajectory is expressed as $Tr(A)$. The matrix representation corresponding to the v-th view is written as $A^{(v)}$. The ℓ_1-norm of the matrix is expressed as $\|A\|_1$. Especially, $\mathbf{1}_n$ is used to represent a n dimensional column vector, each of which is 1.

2 The Proposed Methodology

2.1 Graph-based Clustering Description

Suppose there are n samples which belong to c clusters, a Similarity Matrix (SM) is firstly constructed to represent the affinities of all the samples in the graph-based clustering methods. Many previous works have studied to design an SM with high quality, such as [Cai *et al.*, 2005]. Later, [Nie *et al.*, 2014; Chen and Dy, 2016; Nie *et al.*, 2016] proposed an ideal SM $S \in R^{n \times n}$, and introduced the CLR method [Nie *et al.*,

2016; 2017]. which is supposed to exactly have c connected components. So S can be directly used for the clustering task.

Based on the $\ell2$-norm and the $\ell1$-norm, between the given affinity matrix A and the learned similarity matrix S, Nie et al. defined the CLR for graph-based clustering [Nie *et al.*, 2016]. Then they applied Frobenius norm and proposed the Parameter-weighted Multi-view Clustering (PwMC) [Nie *et al.*, 2017]. However, Frobenius norm is sensitive to outliers and noise which exist in many real situations [Pang *et al.*, 2010]. In order to reduce the negative effect of outliers and noise in graph-based clustering, Frobenius norm is replaced by $\ell1$-norm. Then the target function can be written as

$$\min_{s_i \mathbf{1}_n = 1, s_{ij} \geq 0, S \in C} \|S - A\|_1, \tag{1}$$

where S is a nonnegative matrix, and each row in S sums up to 1. C stands for the set of n by n square matrices with c connected components. Then based on $\ell1$-norm and graph theory in [Fan, 1997], the connectivity constraint can be replaced with a rank constraint, problem (1) becomes

$$\min_{s_i \mathbf{1}_n = 1, s_{ij} \geq 0, rank(L_S) = n-c} \|S - A\|_1, \tag{2}$$

where L_S is a Laplacian matrix, and $L_S = D_S - \frac{(S^T + S)}{2}$. $rank(L_S)$ is the rank of L_S. The degree matrix $D_S \in R^{n \times n}$ is defined as a diagonal matrix with i-th diagonal element $\frac{\sum_j (s_{ij} + s_{ji})}{2}$. Now the target SM can be solved and be directly used for clustering.

2.2 Robust Auto-Weighted Multi-view Clustering (RAMC)

In this paper, a Robust Auto-weighted Multi-view Clustering (RAMC) is proposed to deal with the multi-view clustering problem. For multi-view data, m denotes the number of view and $A^{(1)}, A^{(2)}, ..., A^{(m)}$ denotes the corresponding input SMs, where $A^{(v)} \in R^{n \times n} (1 \leq v \leq m)$. Assigning the same weight to each graph and calculating an average SM is a simple way to deal this problem. However, this method has an obvious disadvantage that it ignores the differences between different view and is easily disturbed by poor quality views, which ultimately leads to the insufficient precision of clustering. Co-training methods also have the similar disadvantage [Blum and Mitchell, 1998]. So an effective strategy to solve this problem is to use a group of meaningful weight to measure the importance of each view. Thus, the formulated objective becomes

$$\min_{\alpha^{(v)}, S} \sum_{v=1}^{m} \alpha^{(v)} \left\| S - A^{(v)} \right\|_1,$$

$$s.t. \quad \alpha^T \mathbf{1}_n = 1, s_{ij} \geq 0, s_i \mathbf{1}_n = 1, rank(L_S) = n - c. \tag{3}$$

To facilitate parameter adjustment, a regularization item is added to the problem (3). Thus, solving the problem (3) is equivalent to finding out the target SM S, which is constrained as in Eq. (3) but can approximate each original input

SM $A^{(v)}$. Then, the formulated objective becomes

$$\min_{\alpha^{(v)},S} \sum_{v=1}^{m} \alpha^{(v)} \left\| S - A^{(v)} \right\|_1 + \gamma \left\| \alpha \right\|_2^2,$$

$$s.t. \quad \alpha^T \mathbf{1}_n = 1, s_{ij} \geq 0, s_i \mathbf{1}_n = 1, rank(L_S) = n - c. \tag{4}$$

where $\alpha = [\alpha^{(1)}, \alpha^{(2)}, ..., \alpha^{(m)}]^T$ and $\gamma > 0$. $\gamma \left\| \alpha \right\|_2^2$ is used to smoothen the weight distribution. Obviously, If $\gamma \to 0$, then the weight of best view will be 1 and other weights will be 0, so the trivial solution will be obtained. Conversely, if $\gamma \to \infty$, then the equal weights will be obtained. Since solving the problem is based on the parameter γ and robust $\ell 1$-norm, so this method is named as Robust Auto-Weighted Multi-view Clustering (RAMC).

2.3 Optimization of RAMC

When the parameter γ is set to a proper value, the alternating iterative strategy is used to optimize the problem (4).

When α is fixed, the following subproblem needs to be solved

$$\min_{s_{ij} \geq 0, s_i \mathbf{1}_n = 1, rank(L_S) = n-c} \sum_{v=1}^{m} \alpha^{(v)} \left\| S - A^{(v)} \right\|_1. \tag{5}$$

$\sigma_i(L_S)$ means the i-th smallest eigenvalue of L_S. Because L_S is a positive semidefinite matrix, so $\sigma_i(L_S) \geq 0$. Given a λ which is large enough, so the rank constraint in Eq. (5) can be removed. Thus the problem (5) can be changed into the following form

$$\min_{s_{ij} \geq 0, s_i \mathbf{1}_n = 1} \sum_{v=1}^{m} \alpha^{(v)} \left\| S - A^{(v)} \right\|_1 + 2\lambda \sum_{i=1}^{c} \sigma_i(L_S). \tag{6}$$

Because λ is large enough, and $\sigma_i(L_S) > 0$ for each i, so for the problem (6), the optimal solution S will let the second term $\sum_{i=1}^{c} \sigma_i(L_S)$ equal to 0 and the constraint $rank(L_S) = n - c$ will be satisfied. Furthermore, according to Ky Fan's Theory [Fan, 1949], the following equation can be written

$$\sum_{i=1}^{c} \sigma_i(L_S) = \min_{F \in R^{n \times c}, F^T F = I} Tr(F^T L_S F). \tag{7}$$

Then, utilizing Eq. (7), the problem (6) is changed into the following problem

$$\min_{S,F} \sum_{v=1}^{m} \alpha^{(v)} \left\| S - A^{(v)} \right\|_1 + 2\lambda Tr(F^T L_S F),$$

$$s.t. \quad s_{ij} \geq 0, s_i \mathbf{1}_n = 1, F \in R^{n \times c}, F^T F = I. \tag{8}$$

This problem can be solved iteratively by optimizing variables F and S.

i. **Solving F When S is fixed**, the problem (8) can be written

$$\min_{F \in R^{n \times c}, F^T F = I} Tr(F^T L_S F). \tag{9}$$

Algorithm 1 The algorithm of Robust Auto-Weighted multi-view Clustering (RAMC) in Eq. (4)

Input:
SMs for m views $A^{(1)}, A^{(2)}, ..., A^{(m)}$ and $A^{(v)} \in R^{n \times n}$, number of clusters c, parameter γ. Initialize the weight for each view (e.g.,$\alpha^{(v)} = \frac{1}{m}$). Let $A = \sum_{v=1}^{m} \alpha^{(v)} A^{(v)}$, and compute $F \in R^{n \times c}$, which is formed by the c eigenvectors of $L_A = D_A - \frac{A^T + A}{2}$ corresponding to the c smallest eigenvalues.
repeat
  repeat
    i. For each i , update the i-th row of S by solving the problem (15). where $U^{(v)}$ is a diagonal matrix with the j-th diagonal element as $\frac{1}{2|\tilde{s}_{ij} - a_{ij}|}$ and $p_i = \sum_{v=1}^{m} \alpha^{(v)} U^{(v)} a_i^{(v)} - \frac{\lambda}{2} v_i$, the j-th element of v_i is $v_{ij} = \|f_i - f_j\|_2^2$.
    ii. Update F, which is formed by the c eigenvectors of $L_S = D_S - \frac{S^T + S}{2}$ corresponding to the c smallest eigenvalues.
  until converge
  Update the weight $\alpha^{(v)}$ by solving the problem (23).
until converge
Output: $S \in R^{n \times n}$ with exactly c connected components and the view weight α.

By the c eigenvectors of L_S corresponding to the c smallest eigenvalues, the optimal solution of F is well composed in this method.

ii. **Solving S When F is fixed**. For fixed F, the problem (8) is as follows

$$\min_{\sum_j s_{ij} > 0} \sum_{v=1}^{m} \alpha^{(v)} \sum_{i,j} |s_{ij} - a_{ij}^{(v)}| + \lambda \sum_{i,j} \|f_i - f_j\|_2^2 s_{ij}. \tag{10}$$

Since for different i, the problem (10) is independent, then the following problem can be solved separately

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

The problem (11) can be changed into vector form as

$$\min_{s_i \mathbf{1}_n = 1, s_i \geq 0} \sum_{v=1}^{m} \alpha^{(v)} \left\| s_i - a_i^{(v)} \right\|_1 + \lambda s_i v_i^T, \tag{12}$$

where v_i is a vector with the j-th element $v_{ij} = \|f_i - f_j\|_2^2$. With the iterative reweighted way, the problem (12) can be solved by iteratively solving the following problem

$$\min_{s_i \mathbf{1}_n = 1, s_i \geq 0} \sum_{v=1}^{m} \alpha^{(v)} Tr(s_i - a_i^{(v)}) U^{(v)} (s_i - a_i^{(v)})^T + \lambda s_i v_i^T, \tag{13}$$
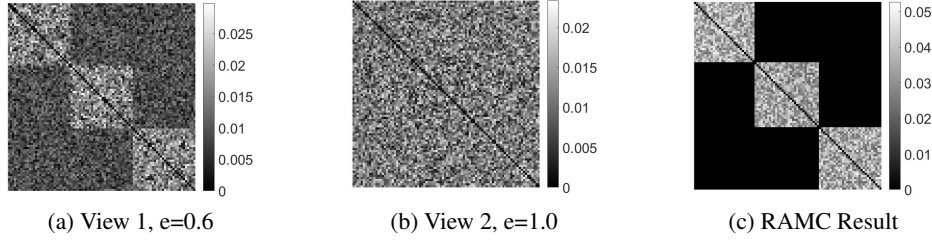
(a) View 1, e=0.6   (b) View 2, e=1.0   (c) RAMC Result

Figure 1: Toy-1's original dataset and the clustering result with RAMC.

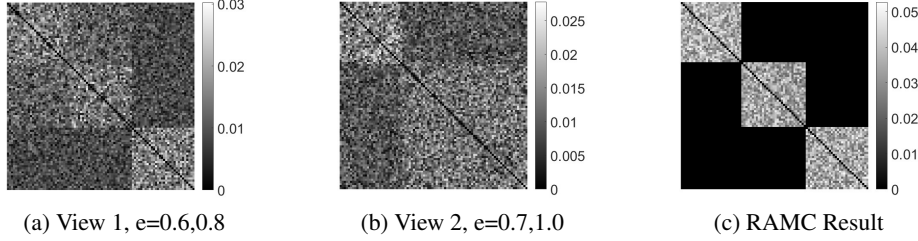(a) View 1, e=0.6,0.8   (b) View 2, e=0.7,1.0   (c) RAMC Result

Figure 2: Toy-2's original dataset and the clustering result with RAMC.

where $U^{(v)}$ is a diagonal matrix with the j-th diagonal element $u_{jj}^{(v)} = \frac{1}{2\left|\widetilde{s}_{ij} - a_{ij}^{(v)}\right|}$, and $\widetilde{s}_{ij}$ is the current solution. Nie et al have been proved that the iterative method decreases the objective of the problem (12) in each iteration and will converge to the optimal solution to the problem (12) .

The problem (13) becomes

$$\min_{s_i \mathbf{1}_n = 1, s_i \geq \mathbf{0}} \sum_{v=1}^{m} \frac{1}{2} \alpha^{(v)} s_i U^{(v)} s_i^T + s_i \left(\sum_{v=1}^{m} \alpha^{(v)} U^{(v)} a_i^{(v)} - \frac{\lambda}{2} v_i \right)^T .$$

(14)

For each i, let $p_i = \sum_{v=1}^{m} \alpha^{(v)} U^{(v)} a_i^{(v)} - \frac{\lambda}{2} v_i$, the following problem can be written

$$\min_{s_i \mathbf{1}_n = 1, s_i \geq \mathbf{0}} \frac{1}{2} \sum_{v=1}^{m} \alpha^{(v)} s_i U^{(v)} s_i^T - s_i p_i^T .$$

(15)

The Lagrangian function of problem (15) is

$$\mathcal{L}\left(s_i, \eta, \beta_i\right) = \frac{1}{2} \sum_{v=1}^{m} \alpha^{(v)} s_i U^{(v)} s_i^T - s_i p_i^T - \eta\left(s_i \mathbf{1} - 1\right) - s_i \beta_i^T ,$$

(16)

where η and β_i are the Lagrangian multipliers, $\eta \neq 0$ and $\beta_i \geq 0$. Let the derivative of Eq. (16) $w.r.t.$ be zero, that is

$$\sum_{v=1}^{m} \alpha^{(v)} U^{(v)} s_i - p_i - \eta \mathbf{1} - \beta_i = \mathbf{0}.$$

(17)

For the j-th element of s_i, it then becomes

$$\sum_{v=1}^{m} \alpha^{(v)} u_{ii}^{(v)} s_{ij} - p_{ij} - \eta - \beta_{ij} = 0.$$

(18)

According to the KKT condition, then $s_{ij}\beta_{ij} = 0$. So from Eq. (18) there is

$$s_{ij} = \frac{1}{\sum\limits_{v=1}^{m} \alpha^{(v)}} \left(\frac{1}{u_{ii}^{(v)}} \eta + \frac{1}{u_{ii}^{(v)}} p_{ij} \right)_+ ,$$

(19)

where $(v)_+ = max(0, v)$. The following function $w.r.t.$ η is defined as

$$g_i(\eta) = \frac{1}{\sum\limits_{v=1}^{m} \alpha^{(v)}} \left[\sum_v \left(\frac{1}{u_{ii}^{(v)}} \eta + \frac{1}{u_{ii}^{(v)}} p_{ij} \right)_+ - 1 \right].$$

(20)

Based on Eqs. (19)-(20), and the constraint, $s_i \mathbf{1}_n = 1$, there is

$$g_i(\eta) = 0.$$

(21)

Obviously, the root of a function $g_i(x)$ is the value of η. Because $g_i(x)$ is a piecewise linear and monotonically increasing function, so η can be easily calculated by Newton's method. After calculating η, the optimal solution to the problem (15) can be obtained by Eq. (19).

When S is fixed, the problem (4) is equivalent to minimization to the following problem:

$$\min_{\alpha^v \geq 0, \alpha^T \mathbf{1}_m = 1} \sum_{v=1}^{m} \alpha^{(v)} e^{(v)} + \gamma \|\alpha\|_2^2 ,$$

(22)

where $e^{(v)} = \left\| S - A^{(v)} \right\|_1$. Then the problem (22) becomes

$$\min_{\alpha \geq 0_m^T, \alpha^T \mathbf{1}_m = 1} \left\| \frac{e}{2\gamma} + \alpha \right\|_2^2 .$$

(23)

The problem can be solved by a valid iterative algorithm which is put forward by [Duchi *et al.*, 2008]. By solving the problem (23), α can be updated. Therefore, the solving process of the problem (4) can be summarized in Algorithm 1. According to [Nie *et al.*, 2017], Algorithm 1 will converge, because each subproblem is obtained a optimal solution with the alternating optimization steps.

3 Experiments

In this section, some experiments are designed to verify the effectiveness of the proposed algorithm. Following the CLR

(a) MSRCv1

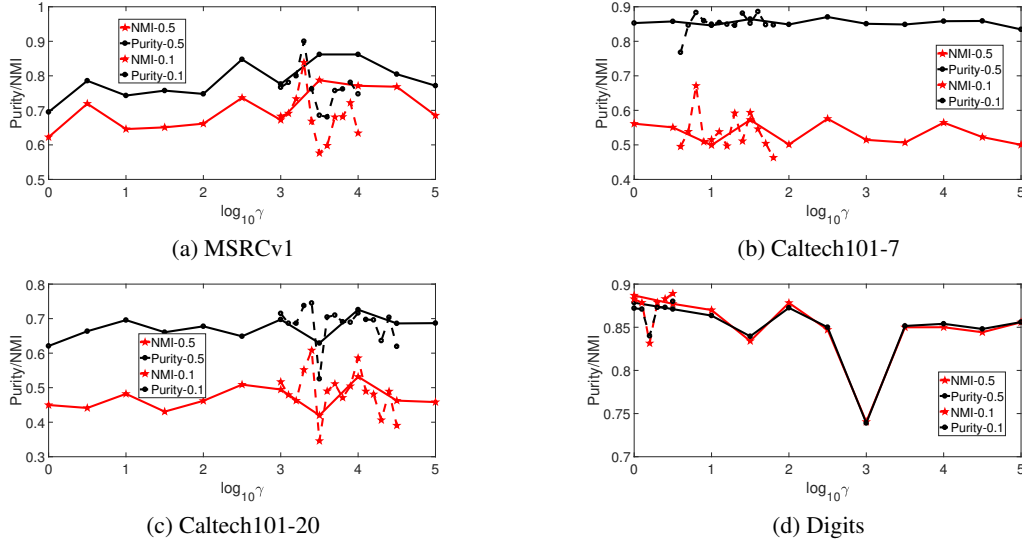(b) Caltech101-7

(c) Caltech101-20

(d) Digits

Figure 3: The clustering performance of the RAMC algorithm on on MSRCv1, Caltech101-7, Caltech101-20 and Digits datasets. The clustering Purity and NMI are shown in solid line and dotted line with step size 0.5 and 0.1 respectively.

method, a graph is constructed for each view of the multi-view datasets as the initial input matrix $A^{(v)}$ with k-nearest neighbor method. Only one parameter κ needs to be set in this method, here the κ represents the number of neighbor nodes. Too large and too small values of κ can lead to poor clustering results. Take the MSRCv1 data set as an example, when κ is set between 5 and 25, it is found that the purity and NMI increase as κ increases. When $\kappa = 12$, the Purity and NMI start to oscillate and don't obviously be improved with increasing the κ value and the Purity and NMI reach the optimal values when $\kappa = 16$. The κ values setting are similar on other data sets. In this paper, κ is fixed to 16. Two metrics are used to evaluate the clustering results, one is the standard clustering Purity, and the other is Normalized Mutual Information (NMI).

3.1 Toy Example

In this part, two toy experiments on the Toy-1 dataset and Toy-2 dataset are conducted to verify the effectiveness of RAMC algorithm. Toy-1 and Toy-2 datasets include two views respectively. Each view is made up of 90×90 matrix. There are three diagonally arranged block matrices in each view of the Toy-1 dataset and the Toy-2 dataset. The elements in the three-block matrices of each view are randomly set from 0 to 1. For the Toy-1 dataset, other data in each matrix are noise data, which is randomly set from 0 to e with $e = 0.6$ in the first view and $e = 1$ in the second view. For the Toy-2 dataset, it has the same scale with the Toy-1 dataset while with different noise settings. With the initial noise $e = 0.6$ and $e = 0.7$ in view 1 and view 2, the input matrix for view 1 is obtained with $e = 0.8$ for the first and second block data, while the input matrix for view 2 is obtained with $e = 1.0$ for the second block and third data. Figure 1 and Figure 2 show the normalized original input graphs and the gray-scales of clustering results processed by the RAMC algorithm. Experimental results show that the RAMC algorithm can remove

View	MSRCv1	Caltech101-7(20)	Digits
1	CM(24)	Gabor(48)	FOU(76)
2	HOG(576)	WM(40)	FAC(216)
3	GIST(512)	CENT(254)	KAR(64)
4	LBP(256)	HOG(1984)	PIX(240)
5	CENT(254)	GIST(512)	ZER(47)
6	-	LBP(928)	MOR(6)
#Size	210	1474(2386)	2000
#Class	7	7(20)	10

Table 1: Statistics of four datasets

the noise successfully and get the pure block matrices. That is, the proposed algorithm shows the optimal clustering performance on multi-view data with noise.

3.2 Clustering Effects on Different Real Datasets

Following [Li *et al.*, 2015], three widely used real-world multi-view datasets, MSRCv1 [Winn and Jojic, 2005], Caltech101 [Li *et al.*, 2007] (following [Nie *et al.*, 2017], two regular subsets Caltech101-7 and Caltech101-20 are used in our experiments) and Digits [Asuncion and Newman, 2007] are considered in the experiments. The summarization of these datasets is shown in Table 1. The proposed RAMC is compared with the methods: Co-regularized spectral clustering [Kumar *et al.*, 2011] (Co-reg), Multi-View Spectral Clustering [Cai *et al.*, 2011] (MVSC), Robust Multi-view Spectral Clustering [Xia *et al.*, 2014] (RMSC), Parameter-weighted Multi-view Clustering (PwMC) [Nie *et al.*, 2017], Self-weighted Multi-view Clustering (SwMC) [Nie *et al.*, 2017]. The result of CLR is also compared as a baseline in the experiment. The parameter in each compared method is set as optimum. In order to increase the generality of the proposed algorithm, $\alpha^{(v)}$ for each view, is initialized to a random weight between 0 and 1, and all weights are normalized. For each compared method, we repeat the experiment 20 times on each dataset and report the average performance with the standard deviation (*std*).

	MSRCv1		Caltech 101-7		Caltech101-20		Digits	
	Purity	NMI	Purity	NMI	Purity	NMI	Purity	NMI
CLR	0.6133±0.0163	0.6005±0.0165	0.8429±0.0209	0.5221±0.0126	0.6099±0.0150	0.3761±0.0184	0.8719±0.0201	0.8759±0.0185
Co-reg	0.6227±0.0154	0.5924±0.0174	0.6671±0.0203	0.3227±0.0198	0.5598±0.0154	0.4879±0.0264	0.8198±0.0198	0.8068±0.0184
MVSC	0.7304±0.0252	0.6152±0.0178	0.8449±0.0149	0.5972±0.0127	0.7105±0.0159	0.5025±0.0184	0.8609±0.0163	0.8532±0.0125
RMSC	0.7369±0.0203	0.6243±0.0348	0.8021±0.0010	0.4790±0.0103	0.7318±0.0158	0.4923±0.0146	0.8001±0.0490	0.7431±0.0257
PwMC	0.8798±0.0346	0.8015±0.0094	0.8549±0.0013	0.5684±0.0009	0.7137±0.0165	0.5119±0.0184	0.8789±0.0135	0.8889±0.0176
SwMC	0.8714±0.0179	0.7894±0.0142	0.8367±0.0157	0.5310±0.0137	0.7137±0.0184	0.5122±0.0194	0.8820±0.0123	0.8935±0.0190
RAMC	**0.8840±0.0087**	**0.8145±0.0241**	**0.8839±0.0033**	**0.6378±0.0052**	**0.7368 ±0.0105**	**0.6180±0.0169**	**0.8950±0.0324**	**0.8960±0.0297**

Table 2: Performance comparison (with *std*) of all the compared algorithms on MSRCv1, Caltech101-7, Caltech101-20 and Digits datasets. Purity and NMI are used as evaluation metrics. The best performance is marked in bold. From the experimental results, we observe that the proposed algorithm generally performs better than the other compared algorithms.

Following [Nie *et al.*, 2017], the clustering results of all the methods are shown in Table 2, and the best results are marked in boldface. According to the Table 2, the proposed RAMC achieves the best clustering purity and NMI on the three real-world datasets compared with other six methods.

It can also be seen that the standard deviation of clustering performance obtained by the RAMC method is quite small, which means that the proposed RAMC method has stable clustering performance. It's worth mentioning that the average clustering NMI of the proposed RAMC on the dataset Caltech101-20 is even 10 percentage points higher than the best performance on other approaches. And the clustering NMI is also improved on the Caltech 101-7 dataset.

For space limitation, only one random running result is shown in Figure 3. It can be seen that the solid lines show the clustering results with parameter γ searched in logarithmic form (log_{10}^{γ}) from 0 to 5 with step size 0.5. In theory, the best clustering results would be obtained if all possible γ is searched. With an acceptable way, the step size is reduced from 0.5 to 0.1, thus the more accurate clustering results can be obtained. For clarity, the dotted line exhibits some clustering results near the optimal value with step size 0.1.

3.3 Robustness Performance Evaluation

To prove the robustness of the proposed algorithm, the following experiments are designed. First, a set of noisy dataset needs to be constructed based on an original dataset. Suppose r is the ratio of random noise, and n is the number of the original datasets, we randomly pick out $n \times r$ data points from the original dataset. The selected data is added to a normal distribution with an average of 300 and a standard deviation of 30, thus a set of the noisy dataset are formed with different r from 0 to 0.5 with a step size 0.05. For space limitation, we only take the MSRCv1 dataset as an example. To make the comparison more clear, we only compare with the second best and the third best algorithms. We use the clustering purity and NMI as evaluation metrics in this section. The experimental results are reported in Figure 4. From the experimental results, we have the following observation. Although as the ratio of random noise increases, the performance of all the compared algorithms decreases, the performance gain of the proposed algorithm becomes more significant. For example, compared with PwMC, the performance gain of RAMC on Purity increases from 0.59% to 67.00% when the ratio of random noise increases from 0 to 0.5. Meanwhile, compared with PwMC, the performance gain of RAMC on NMI increases from 1.22% to 498.55% when the ratio of random
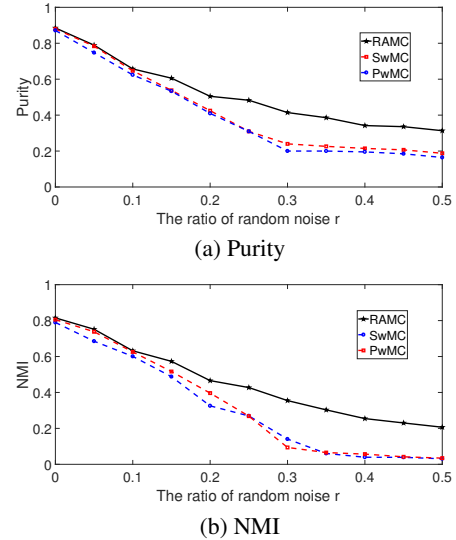

(a) Purity

(b) NMI

Figure 4: We conduct experiments to test the robustness of the proposed algorithm. From the experimental results, we can see that as the noisy percentage increases, the performance gain of the proposed algorithm becomes more significant.

noise increases from 0 to 0.5. This phenomenon confirms the robustness of the proposed algorithm.

4 Conclusion and Future Work

In this paper, we have proposed a novel Robust Auto-Weighted multi-view Clustering (RAMC) that aims to learn a *consensus* graph with exactly k connected components, where k is the number of clusters. The structure of the obtained graph makes it suitable to be directly used for clustering without any further post-processing. To make sure the graph has exactly k connected components, we explicitly impose a low-rank constraint on the objective function and ℓ_l-norm has been employed in this paper for robustness. With the extensive experiments, the proposed method shows the superiority and robustness on two synthetic datasets and three real-world datasets. In future work, this framework will be extended to semi-supervised context.

Acknowledgments

References

[Asuncion and Newman, 2007] Arthur Asuncion and David Newman. UCI machine learning repository, 2007.

[Blum and Mitchell, 1998] Avrim Blum and Tom M. Mitchell. Combining labeled and unlabeled data with co-training. In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory*, pages 92–100, 1998.

[Cai *et al.*, 2005] Deng Cai, Xiaofei He, and Jiawei Han. Document clustering using locality preserving indexing. *IEEE Trans. Knowl. Data Eng.*, 17(12):1624–1637, 2005.

[Cai *et al.*, 2011] Xiao Cai, Feiping Nie, Heng Huang, and Farhad Kamangar. Heterogeneous image feature integration via multi-modal spectral clustering. In *CVPR*, pages 1977–1984, 2011.

[Chang *et al.*, 2015] Xiaojun Chang, Feiping Nie, Zhigang Ma, Yi Yang, and Xiaofang Zhou. A convex formulation for spectral shrunk clustering. In *AAAI*, pages 2532–2538, 2015.

[Chang *et al.*, 2016] Xiaojun Chang, Feiping Nie, Yi Yang, Chengqi Zhang, and Heng Huang. Convex sparse PCA for unsupervised feature learning. *TKDD*, 11(1):3:1–3:16, 2016.

[Chao *et al.*, 2017] Guoqing Chao, Shiliang Sun, and Jinbo Bi. A survey on multi-view clustering. *Computing Research Repository*, December 2017.

[Chen and Dy, 2016] Junxiang Chen and Jennifer G. Dy. A generative block-diagonal model for clustering. In *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence*. Association for Uncertainty in Artificial Intelligence, 2016.

[Duchi *et al.*, 2008] John C. Duchi, Shai Shalev-Shwartz, Yoram Singer, and Tushar Chandra. Efficient projections onto the l_1-ball for learning in high dimensions. In *ICML*, pages 272–279. ACM, 2008.

[Fan, 1949] Ky Fan. On a theorem of weyl concerning eigenvalues of linear transformations ii. *Proceedings of the National Academy of Sciences of the United States of America*, 35(11):652–655, January 1949.

[Fan, 1997] R. K. Chung Fan. *Spectral graph theory*. Published for the Conference Board of the mathematical sciences by the American Mathematical Society, 1997.

[Hagen and Kahng, 1992] Lars W. Hagen and Andrew B. Kahng. New spectral methods for ratio cut partitioning and clustering. *IEEE Trans. on CAD of Integrated Circuits and Systems*, 11(9):1074–1085, September 1992.

[Kang *et al.*, 2017a] Zhao Kang, Chong Peng, and Qiang Cheng. Twin learning for similarity and clustering: A unified kernel approach. 2017.

[Kang *et al.*, 2017b] Zhao Kang, Chong Peng, Qiang Cheng, and Zenglin Xu. Unified spectral clustering with optimal graph. 2017.

[Kumar *et al.*, 2011] Abhishek Kumar, Piyush Rai, and Hal Daumé III. Co-regularized multi-view spectral clustering. In *NIPS*, pages 1413–1421, 2011.

[Li *et al.*, 2007] Fei-Fei Li, Robert Fergus, and Pietro Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding*, 106(1), 2007.

[Li *et al.*, 2015] Yeqing Li, Feiping Nie, Heng Huang, and Junzhou Huang. Large-scale multi-view spectral clustering via bipartite graph. In *AAAI*, pages 2750–2756, 2015.

[Nie *et al.*, 2014] Feiping Nie, Xiaoqian Wang, and Heng Huang. Clustering and projected clustering with adaptive neighbors. In *ACM SIGKDD*, pages 977–986, 2014.

[Nie *et al.*, 2016] Feiping Nie, Xiaoqian Wang, Michael I. Jordan, and Heng Huang. The constrained laplacian rank algorithm for graph-based clustering. In *AAAI*, pages 1969–1976, 2016.

[Nie *et al.*, 2017] Feiping Nie, Jing Li, and Xuelong Li. Self-weighted multiview clustering with multiple graphs. In *IJCAI*, pages 2564–2570, 2017.

[Pang *et al.*, 2010] Yanwei Pang, Xuelong Li, and Yuan Yuan. Robust tensor analysis with l1-norm. *IEEE Trans. Circuits Syst. Video Techn.*, 20(2):172–178, February 2010.

[Sun, 2013] Shiliang Sun. A survey of multi-view machine learning. *Neural Computing and Applications*, 23(7-8):2031–2038, February 2013.

[Tao *et al.*, 2018] Hong Tao, Chenping Hou, Xinwang Liu, Dongyun Yi, and Jubo Zhu. Reliable multi-view clustering. 2018.

[von Luxburg, 2007] Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, December 2007.

[Winn and Jojic, 2005] John M. Winn and Nebojsa Jojic. LOCUS: learning object classes with unsupervised segmentation. In *ICCV*, pages 756–763, 2005.

[Xia *et al.*, 2014] Rongkai Xia, Yan Pan, Lei Du, and Jian Yin. Robust multi-view spectral clustering via low-rank and sparse decomposition. In *AAAI*, pages 2149–2155. AAAI, 2014.

[Xu *et al.*, 2013] Chang Xu, Dacheng Tao, and Chao Xu. A survey on multi-view learning. *CoRR*, April 2013.

[Xue *et al.*, 2017] Xiaowei Xue, Feiping Nie, Sen Wang, Xiaojun Chang, Bela Stantic, and Min Yao. Multi-view correlated feature learning by uncovering shared component. In *AAAI*, pages 2810–2816, 2017.