



Spectral clustering via half-quadratic optimization

Xiaofeng Zhu^{1,2} · Jiangzhang Gan² · Guangquan Lu³ · Jiaye Li³ · Shichao Zhang⁴

Received: 2 June 2019 / Revised: 6 August 2019 / Accepted: 9 September 2019 /
Published online: 07 November 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Spectral clustering has been demonstrated to often outperform K-means clustering in real applications because it improves the similarity measurement of K-means clustering. However, previous spectral clustering method still suffers from the following issues: 1) easily being affected by outliers; 2) constructing the affinity matrix from original data which often contains redundant features and outliers; and 3) unable to automatically specify the cluster number. This paper focuses on address these issues by proposing a new clustering algorithm along with the technique of half-quadratic optimization. Specifically, the proposed method learns the affinity matrix from low-dimensional space of original data, which is obtained by using a robust estimator to remove the influence of outliers as well as a sparsity regularization to remove redundant features. Moreover, the proposed method employs the $\ell_{2,1}$ -norm regularization to automatically learn the cluster number according to the data distribution. Experimental results on both synthetic and real data sets demonstrated that the proposed method outperforms the state-of-the-art methods in terms of clustering performance.

Keywords Spectral clustering · Subspace learning · Feature selection · M-estimation · Half-quadratic optimization

1 Introduction

Cluster analysis is playing an important role in real applications, such as stock data analysis, market segmentation, production supervision, and anomaly detection [28, 39]. In recent years, spectral clustering has been attracting wide attention due to conducting K-mean clustering on the spectral representation rather than original representation [11, 31, 32]. To do this, spectral clustering first obtains the spectral representation of original data by conducting the eigenvalue decomposition of the affinity matrix of original data, and then conducts K-means clustering on the resulted spectral representation [30].

This article belongs to the Topical Collection: *Computational Social Science as the Ultimate Web Intelligence*

Guest Editors: Xiaohui Tao, Juan D. Velasquez, Jiming Liu, and Ning Zhong

✉ Shichao Zhang
zhangsc@csu.edu.cn

Extended author information available on the last page of the article.

Spectral clustering generates the new representation (*i.e.*, the spectral representation) of original data through matrix spectral analysis theory so that it is simple to implement, and does not easily fall into local optimal solution, compared to other clustering methods [30]. The key step of spectral clustering is the learning of spectral representation, *i.e.*, affinity matrix learning, which is also the key difference between spectral clustering and K-means clustering. Usually, the construction of the affinity matrix includes local representation methods and global representation methods. The global representation methods represent each data point by all of data points [34, 46], while the local representation methods use the local information of original data to construct the affinity matrix between samples, such as local subspace affinity [33] and locally linear manifold clustering [6]. However, previous spectral clustering still has issues to be addressed.

First, cluster performance always suffers from the influences of redundant features and outliers. However, it is very often to find them on original data. Hence, the quality of both the cluster models and the affinity matrix constructed on original data are not guaranteed [25, 37, 40]. In the literature, some studies only focused on avoiding the influence of either redundant features or outliers for constructing clustering model. For example, the literature [16, 38] employed the sparse technique to avoid the influence of redundant features for clustering, while the studies in [5, 19, 45, 46] used the low-rank method to avoid the influence of outliers. A few literature has focused on taking the two issues into account in a unified framework [42]. For example, some studies in [17, 25] proposed avoiding the influence of redundant features by an $\ell_{2,1}$ -norm regularization as well as reducing the effect of outliers by the robust estimator.

Second, the affinity matrix learning is also affected by the redundant features and outliers. A number of previous literature focus on constructing the affinity matrix from the low-dimensional space of original data rather than the original space used in traditional spectral clustering [20, 27]. To the best of our knowledge, no literature has focused on simultaneously learning clustering models and the affinity matrix from the low-dimensional space of original data by reducing the influence of outliers and redundant features.

Third, many clustering methods (such as K-means clustering and spectral clustering) need to specify the cluster number, which requires prior knowledge. Recently, the literature [25] focuses on automatically learning the cluster number by designing a regularization, but does not taking into account the influence of outliers and redundant features on the affinity matrix learning.

In this paper, we proposed a new clustering algorithm to address above three issues in a unified framework by employing the half-quadratic optimization technique. Specifically, our proposed method employs the sparse regularization and the half-quadratic technique, respectively, to reduce the influence of redundant features and outliers so that our method is available to learn a clustering model and the affinity matrix from the low-dimensional space of high-dimensional original data. Moreover, we employ an $\ell_{2,1}$ -norm regularization to automatically determine the cluster number. Furthermore, we propose a new alternating optimization strategy to solve the proposed objective function.

Compared to the previous clustering methods, we conclude the contributions of our proposed method as follows:

- The proposed method jointly conducts the clustering task and feature selection from low-dimensional space of original data in a unified framework. This makes our method reduce the influence of outliers and redundant features. To the best of our knowledge, a few literature has focused on these two issues simultaneously.

- The proposed method does not need to predefine the cluster number, which is an open issue in the clustering study. Although some studies focused on solving this issue, they did not take into account the influence of outliers and redundant features.

We summarize the rest of this paper as follows. We briefly review the literature related to this paper in Section 2. In Section 3, we give the detailed description of our proposed method as well as list our optimization strategy to the resulted objective function. After that, we present experiment results in Section 4 and finally conclude this paper in Section 5.

2 Related work

In this section, we briefly review existing methods for clustering analysis by the information, such as the clustering definition, the clustering categories, and previous clustering methods for dealing with the noise and redundancy inherent data points.

2.1 Clustering analysis

Cluster analysis pushes data points to be grouped in clusters, helping us to analyze and describe the real-world. Thus it is an important branch in the field of machine learning as well as plays an important role in the real applications, such as psychology, biology, pattern recognition, and data mining [1, 26]. According to the calculation criteria of the similarity between two objects, previous clustering algorithms can be divided into the subcategories, such as partition-based methods, hierarchical method, and density-based method. The partition-based clustering method (*e.g.*, K-Means algorithm and K-modes algorithm) simply divides data points to several non-overlapped subsets such that every data point belongs to only one subset [4]. The hierarchical clustering method is to create a hierarchical tree by calculating the similarity between different types of data points. Moreover, in the clustering tree, original data points with different categories are located in the lowest layer of the tree while the top layer of the tree is located in the root node of a cluster. Traditional methods for a clustering tree generation include the bottom-up merge methods and the top-down split methods [44]. The density-based clustering method is designed to start from the distribution density of data objects, and then connects adjacent regions with enough density to detect clusters with different shapes. Recently, the density-based clustering methods are mainly used for clustering spatial data in the domain of database [36].

2.2 Clustering methods for feature selection

Clustering analysis usually involves high-dimensional data, which contains noise and redundancy to degrade the clustering performance [5, 14]. However, most existing clustering algorithms did not pay attention to this issue. A naive solution to noise and redundancy is to first conduct dimensionality reduction and then conduct clustering analysis on the reduced data. However, such a two-step strategy can not guarantee to output significant clustering performance as the goal of the first step (*i.e.*, dimensionality reduction [38, 40]) is focused on reducing the feature number rather than improving the clustering performance [42, 45, 46]. Recently, Volker et al. proposed to combine clustering analysis with feature selection in a unified framework [24]. Martin et al. proposed the concept of feature saliency and introduces the Expectation–Maximization (EM) algorithm for clustering analysis [15]. Zhu et al. proposed a novel one-step spectral clustering algorithm to combine the similarity learning

with feature selection in a unified one-step clustering framework to simultaneously conduct feature selection and spectral clustering [41].

2.3 Clustering methods for similarity measurement

Cluster analysis is an unsupervised learning method due to the lack of class labels, so the goal of clustering is to maximize the similarity between two data points within the same category as well as minimize the similarity between two data points in different categories. Hence, the similarity measurement is the key step for clustering analysis [14]. In the literature, Zelnik-Manor proposed to self-adjust the number of the nearest neighbors of data points for conducting spectral clustering [35]. Wang et al. proposed determining the similarity by the density and distance of the data points [23]. Liu et al. proposed adjusting the similarity between data points by the number of shared neighbors between two data points [18].

2.4 Clustering methods for avoiding the influence of outliers

Although many methods have been proposed to improve the robustness of clustering analysis, how to improve the robustness of clustering algorithms is still a challenging problem due to the diversity and complexity of noise. Recently, Wang et al. proposed to assign a scale value to each data point according to the distance between the data point and the corresponding centroid, so as to distinguish normal data points from outliers in the clustering process [29]. Liu et al. proposed to seek the lowest rank representation that can represent the data points as linear combinations of the bases for spectral clustering [7]. Zhu et al. proposed a new robust multi-view clustering method to solve the initialization sensitivity of existing multi-view clustering, and introduce the correlation entropy loss function to estimate the reconstruction error for effectively avoiding the impact of outliers [48].

3 Approach

In this section, we list the notations used in this paper in Section 3.1, and then report the details of our proposed method in Section 3.2. Finally, we describe our optimization method for the proposed objective function in Section 3.3.

3.1 Notations

In this paper, matrices and vectors are denoted as boldface uppercase letters and boldface lowercase letters, respectively. More specifically, given a matrix $\mathbf{X} = [x_{ij}]$, its i -th row and j -th column are denoted as \mathbf{x}_i and \mathbf{x}_j , respectively. We denote the Frobenius norm, $\ell_{2,p}$ -norm of a matrix \mathbf{X} , respectively, as $\|\mathbf{X}\|_F = \sqrt{\sum_i \|\mathbf{x}_i\|_2^2} = \sqrt{\sum_j \|\mathbf{x}_j\|_2^2}$ and $\|\mathbf{X}\|_{2,p} = \left(\sum_i \sqrt{\sum_j x_{ij}^{\frac{p}{2}}} \right)^{\frac{1}{p}}$. We further denote the transpose operator and the inverse of a matrix \mathbf{X} as \mathbf{X}^T and \mathbf{X}^{-1} , respectively.

3.2 Proposed method

Finding the internal structure of data is a challenging task in clustering task [13, 37]. Given a data set $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{n \times d}$ as the data matrix with n samples, where d represents the number of the features, we can use either global representation methods or local representation methods to construct the affinity matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$. Since the local representation method linearly represents the relationship between the data point and its nearest neighbors [16, 40], it could effectively avoid the influence of outliers, so that it has been proved better than the global representation method [45, 46]. In this paper, the local representation method is used to construct the affinity matrix, using Euclidean distance to measure the similarity between the data points, with the assumption that every data point can be connected to other data in the data set. Specially, letting $s_{i,j}$ to represent the similarity between the i -th data point and the j -th data point, a small distance $\|\mathbf{x}_i - \mathbf{x}_j\|_2^2$ should be assigned a large probability $s_{i,j}$ while a large distance $\|\mathbf{x}_i - \mathbf{x}_j\|_2^2$ should be assigned a small probability $s_{i,j}$. Hence, we can obtain the following objective function based on previous literature [38, 43]

$$\min \sum_{j=1}^n s_{i,j} \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 \quad (1)$$

Previous methods (e.g., [11, 39, 43]) assume that the affinity matrix constructed in the original feature space can accurately represent the true relationship among data points, so the affinity matrix can be used to guide the original feature space for conduct the prediction task. However, such prediction is often inaccurate because both the redundant features and outliers in the data set make the actual distribution of the data set complex. Thus, it is difficult to guarantee that the affinity matrix learnt from the original feature space can effectively guide the clustering process in the intrinsic low-dimensional subspace. This makes that it is very challenging to conduct clustering on high-dimensional data. To address this issue, we explore the optimal subspace in the original feature space and extract the inherent affinity matrix from the optimal low-dimensional subspace, based on the assumption that original data structure actually lies on a low-dimensional space.

More specifically, by denoting $\mathbf{W} \in \mathbb{R}^{d \times c}$ (where $c \leq d$ is the dimensions of intrinsic space of original high-dimensional data) as the transformation matrix that maps the original data \mathbf{X} to its intrinsic subspace spanned by \mathbf{XW} , we design to learn an intrinsic affinity matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ in the intrinsic low-dimensional space via the following objective function

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{S}} \sum_{j=1}^n s_{i,j} \|\mathbf{x}_i \mathbf{W} - \mathbf{x}_j \mathbf{W}\|_2^2 + \beta \|\mathbf{S}\|_2^2 + \lambda \|\mathbf{W}\|_{2,1} \\ \text{s.t.}, \mathbf{W}^T \mathbf{X}^T \mathbf{X} \mathbf{W} = \mathbf{I}_c, \forall i, \mathbf{s}_i^T \mathbf{1} = 1, 0 \leq s_{i,j} \leq 1, \end{aligned} \quad (2)$$

where λ is a tuning parameter. The penalty $\|\mathbf{W}\|_{2,1}$ conducts feature selection by outputting the row sparsity on \mathbf{W} to remove redundant features of \mathbf{X} . While the orthogonal constraint on the scatter matrix $\mathbf{W}^T \mathbf{X}^T \mathbf{X} \mathbf{W} = \mathbf{I}_c \in \mathbb{R}^{c \times c}$ actually conducts subspace learning to transfer original d -dimensional feature space into c -dimensional space. In this way, (2) reduces the influence of redundant features by two kinds of dimensionality reduction methods, i.e., subspace learning and feature selection.

In spectral clustering, the quality of the affinity matrix will affect the effect of clustering. Moreover, the ℓ_2 -norm regularization is only effective for Gaussian noise processing with independent distribution, and can not effectively handle the abnormal edge of the affinity matrix \mathbf{S} , i.e., outliers. In other words, in the process of optimizing \mathbf{W} , the abnormal edges will still appear. Given noisy real-world data, heavy contamination of the connectivity structure by connections across different underlying clusters is inevitable. To address

this issue, in this paper, our method uses robust estimators to automatically prune spurious inter-cluster connections while maintaining veridical intra-cluster correspondences by the following objective function

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{S}} \sum_{i,j=1}^n s_{i,j} \rho(\|\mathbf{x}_i \mathbf{W} - \mathbf{x}_j \mathbf{W}\|_2) + \beta \|\mathbf{S}\|_2^2 + \lambda \|\mathbf{W}\|_{2,1} \\ \text{s.t.}, \mathbf{W}^T \mathbf{X}^T \mathbf{X} \mathbf{W} = \mathbf{I}_c, \forall i, \mathbf{s}_i^T \mathbf{1} = 1, 0 \leq s_{i,j} \leq 1, \end{aligned} \quad (3)$$

where $\rho(\cdot)$ is a robust estimator designed by the half-quadratic theory [9, 12, 22] and the $\ell_{2,1}$ -norm regularization on $(\mathbf{x}_i \mathbf{W} - \mathbf{x}_j \mathbf{W})$ could automatically output the number of the clusters based on the data distribution [25].

Robust estimators could produce robust estimations that are not unduly influenced by outliers and output significant performance with reducing influence of outliers [12]. Based on the half-quadratic theory, the researchers designed a number of robust estimators, each of which could theoretically reduce the influence of outliers. In this paper, for simplicity, we use the well-known German and Reynolds estimator [3] to replace $\rho(\cdot)$, *i.e.*,

$$\rho(\mathbf{z}) = \frac{-1}{1+|\mathbf{z}|} \quad \text{with} \quad \varphi(\mathbf{z}) = \frac{1}{(1+\mathbf{z})^2} \quad (4)$$

where \mathbf{z} is a vector or a matrix, $\varphi(\mathbf{z})$ is a minimization function.

According to the literature [22], the half-quadratic theory usually transfers the robust estimator to its equivalent formulation by introducing an auxiliary variable p , *e.g.*, $\rho(z_i) = \min_{p_i} \{p_i z_i^2 + \varphi(p_i)\}$, where $p_i z_i^2$ is a quadratic function and $\varphi(p_i)$ is the dual function of the robust estimator. Usually, each robust estimator has a corresponding minimization function, we can assign different weights to each z_i by minimization function. Specifically, if z_i is an outlier, the value of p_i is small or even 0, and vice versa. In this way, the influence of outliers will be reduced.

Based on above analysis, we transfer our objective function in (3) to our final objective function as follows.

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{S}, \mathbf{P}} \sum_{i,j=1}^n s_{i,j} p_{i,j} \|\mathbf{x}_i \mathbf{W} - \mathbf{x}_j \mathbf{W}\|_2^2 + \varphi(\mathbf{P}) + \beta \|\mathbf{S}\|_2^2 + \lambda \|\mathbf{W}\|_{2,1} \\ \text{s.t.}, \mathbf{W}^T \mathbf{X}^T \mathbf{X} \mathbf{W} = \mathbf{I}_c, \forall i, \mathbf{s}_i^T \mathbf{1} = 1, 0 \leq s_{i,j} \leq 1. \end{aligned} \quad (5)$$

Algorithm 1: Pseudo code of solving (5).

Input: $\mathbf{X} \in \mathbb{R}^{n \times d}$, β , and λ ;

Output: $\mathbf{S} \in \mathbb{R}^{n \times n}$;

1 Initialize $\mathbf{W} \in \mathbb{R}^{d \times c}$;

2 Initialize $t = 1$;

3 **repeat**

4 Update \mathbf{P} via (6);

5 Update \mathbf{S} via (9);

6 Update \mathbf{W} via (13);

7 $t = t + 1$;

8 **until** *Convergence*;

3.3 Optimization

Equation (5) is not jointly convex to all the variables (i.e., \mathbf{S} , \mathbf{W} and \mathbf{P}), but is convex for each variable while fixing the others. In this paper, we employ an alternating optimization strategy to optimize (5) and list the details in Algorithm 1.

(i) Update \mathbf{P} by fixing \mathbf{S} and \mathbf{W}

According to the literature [2, 22], if the variables \mathbf{S} and \mathbf{W} are fixed, \mathbf{P} can be updated by the following formula:

$$p_{i,j} = \frac{1}{(1 + \|\mathbf{x}_i \mathbf{W} - \mathbf{x}_j \mathbf{W}\|_2)^2}, i, j = 1, \dots, n. \tag{6}$$

(ii) Update \mathbf{S} by fixing \mathbf{W} and \mathbf{P}

Given \mathbf{P} and \mathbf{W} , we obtain the optimization problem with respect to \mathbf{S} as follows:

$$\begin{aligned} \min_{\mathbf{S}} \quad & \sum_{i,j} s_{i,j} p_{i,j} \|\mathbf{x}_i \mathbf{W} - \mathbf{x}_j \mathbf{W}\|_2^2 + \beta \|\mathbf{s}_i\|_2^2, \\ \text{s.t.} \quad & \forall i, \mathbf{s}_i^T \mathbf{1} = 1, 0 \leq s_{i,j} \leq 1 \end{aligned} \tag{7}$$

The optimization of \mathbf{S} is independent on each row, so we further change (7) to individually optimize \mathbf{s}_i ($i = 1, \dots, n$) as follow:

$$\min_{\mathbf{s}_i^T \mathbf{1}=1, s_{i,j} \geq 0} \sum_{j=1}^n s_{i,j} e_{i,j} + \beta s_{i,j}^2, \tag{8}$$

where $\mathbf{e}_i = \{e_{i,1}, \dots, e_{i,n}\}$, $e_{i,j} = p_{i,j} \|\mathbf{x}_i \mathbf{W}^T - \mathbf{x}_j \mathbf{W}^T\|_2^2$.

We first calculate k nearest neighbors of each data point, and then set the value of $s_{i,j}$ as 0 if the j -th data point is not one of k nearest neighbors of the i -th data point. Otherwise, the value of $s_{i,j}$ can be solved by Karush–Kuhn–Tucker (KKT) [47] conditions, i.e.,

$$s_{i,j} = \begin{cases} \frac{\hat{e}_{i,k+1} - e_{i,j}}{k \hat{e}_{i,k+1} - \sum_{j=1}^k e_{i,j}} & j \leq k \\ 0 & j > k \end{cases} \tag{9}$$

The number of nearest neighbors k can be tuned by cross-validation methods. Moreover, different data points have different numbers of nearest neighbors, base on the constraint “ $\hat{e}_{i,k+1} = \hat{e}_{i,j}$ ” for specific k and j .

(iii) Update \mathbf{W} by fixing \mathbf{S} and \mathbf{P}

While fixing \mathbf{P} and \mathbf{S} , the objective function with respect to \mathbf{W} becomes:

$$\begin{aligned} \min_{\mathbf{W}} \quad & \sum_{i,j=1}^n s_{i,j} p_{i,j} \|\mathbf{x}_i \mathbf{W} - \mathbf{x}_j \mathbf{W}\|_2^2 + \lambda \|\mathbf{W}\|_{2,1} \\ \text{s.t.}, \quad & \mathbf{W}^T \mathbf{X}^T \mathbf{X} \mathbf{W} = \mathbf{I}_c, \forall i, \mathbf{s}_i^T \mathbf{1} = 1, 0 \leq s_{i,j} \leq 1 \end{aligned} \tag{10}$$

By setting $\tilde{s}_{i,j} = s_{i,j} p_{i,j}$ and $\mathbf{L} = \mathbf{D} - \tilde{\mathbf{S}} \in \mathbb{R}^{n \times n}$ as Laplacian matrix, where $\mathbf{D} = \text{diag}(\tilde{\mathbf{S}}\mathbf{1})$, (10) is equivalent to

$$\min_{\mathbf{W}} \text{tr}(\mathbf{W}^T \mathbf{X}^T \mathbf{L} \mathbf{X} \mathbf{W}) + \lambda \|\mathbf{W}\|_{2,1}, \text{s.t.}, \mathbf{W}^T \mathbf{X}^T \mathbf{X} \mathbf{W} = \mathbf{I}_c \tag{11}$$

Since the $\ell_{2,1}$ -norm regularization is convex and non-smooth, we employ the framework of iteratively reweighted least squares (IRLS) to optimize \mathbf{S} , via changing (11) to the following objective function

$$\min_{\mathbf{W}} \operatorname{tr}(\mathbf{W}^T \mathbf{X}^T \mathbf{L} \mathbf{X} \mathbf{W} + \lambda \mathbf{W}^T \mathbf{Q} \mathbf{W}), \text{ s.t.}, \mathbf{W}^T \mathbf{X}^T \mathbf{X} \mathbf{W} = \mathbf{I}_c \quad (12)$$

where \mathbf{Q} is a diagonal matrix with the i -th diagonal element $q_{i,i} = \frac{1}{2\|\mathbf{w}^i\|_2}$, $i = 1, \dots, d$.

In (12), both \mathbf{Q} and \mathbf{W} are unknown. Moreover, \mathbf{Q} depends on \mathbf{W} . According to the IRLS framework, we design an iterative algorithm to solve problem (12) by two sequential steps until the algorithm converges: 1) By fixing the value of \mathbf{W} , we can obtain the value of \mathbf{Q} ; 2) By fixing the value of \mathbf{Q} , (12) is changed to an eigen-decomposition problem with respect to \mathbf{W} .

$$\min_{\mathbf{W}} \operatorname{tr}(\mathbf{W}^T (\mathbf{X}^T \mathbf{L} \mathbf{X} + \lambda \mathbf{Q}) \mathbf{W}), \text{ s.t.}, \mathbf{W}^T \mathbf{X}^T \mathbf{X} \mathbf{W} = \mathbf{I}_c \quad (13)$$

The optimal solution \mathbf{W} in (13) is the eigenvectors of $(\mathbf{X}^T \mathbf{X} + \chi \mathbf{I}_d)^{-1} (\mathbf{X}^T \mathbf{L} \mathbf{X} + \lambda \mathbf{Q})$ since $\mathbf{X}^T \mathbf{X} + \chi \mathbf{I}_d$ is invertible, where χ is a very small positive value.

3.4 Convergence analysis

This section proves that the objective function value in (6) monotonically decrease in each iteration. Denoting the objective function in (6) as $J(\mathbf{W}, \mathbf{S}, \mathbf{P})$, according to the properties of half-quadratic theory in [9], if the values of both \mathbf{W} and \mathbf{S} are fixed, the optimization of \mathbf{P} has a closed-form solution. Hence, we have

$$J(\mathbf{W}^{t+1}, \mathbf{S}^{t+1}, \mathbf{P}^{t+1}) \leq J(\mathbf{W}^{t+1}, \mathbf{S}^{t+1}, \mathbf{P}^t) \quad (14)$$

When the values of \mathbf{W}^{t+1} and \mathbf{P}^t are fixed, the optimization of \mathbf{S}^{t+1} takes a closed-form solution, so we have

$$J(\mathbf{W}^{t+1}, \mathbf{S}^{t+1}, \mathbf{P}^t) \leq J(\mathbf{W}^{t+1}, \mathbf{S}^t, \mathbf{P}^t) \quad (15)$$

Since we conduct the optimization of \mathbf{W} by the eigenvalue decomposition method, so we have

$$J(\mathbf{W}^{t+1}, \mathbf{S}^t, \mathbf{P}^t) \leq J(\mathbf{W}^t, \mathbf{S}^t, \mathbf{P}^t) \quad (16)$$

By integrating above three results together, *i.e.*, (14)-(16), we obtain

$$J(\mathbf{W}^{t+1}, \mathbf{S}^t, \mathbf{P}^t) \leq J(\mathbf{W}^t, \mathbf{S}^t, \mathbf{P}^t) \quad (17)$$

By combining (14), (15) with (17), we have:

$$J(\mathbf{W}^{t+1}, \mathbf{S}^{t+1}, \mathbf{P}^{t+1}) \leq J(\mathbf{W}^t, \mathbf{S}^t, \mathbf{P}^t) \quad (18)$$

4 Experiments

In this section, we evaluated our proposed clustering method, by comparing to seven state-of-the-art clustering methods, on two synthetic data sets and 12 public data sets, in terms of clustering performance.

4.1 Experimental analysis on synthetic data sets

In this subsection, we evaluated the effectiveness of our proposed method on two synthetic data sets with different data distributions.

4.1.1 Experimental analysis on the two-moon data set

The bimonthly data was randomly generated for the first synthesized data set. Specifically, there were two clusters in the data set and they were distributed in the two-moon shape, where each cluster contains 100 data points and the noise ratio was set as 0.13. The aim of this experiment was set to use our proposed method to calculate the affinity matrix by dividing all of the data points into two clusters accurately. In our experiment, we marked these two clusters in the data set as red and blue colors respectively, as shown in Figure 1a. We also denoted the width of the line as the similarity of between two data points, as shown in Figure 1b-c.

As shown in Figure 1b, in the original data set, there are connection points from different clusters connected to each other. However, there are no data points connected between the two clusters for our proposed method in Figure 1c. This indicates that our proposed method has successfully divided the data into two clusters by avoiding the influence of outliers.

4.1.2 Experimental analysis on the three-ring data set

In the second synthetic data set, we randomly generated a three-ring data set with five dimensions, where the first two dimensions were distributed in three circles and the left three features are redundant features. Since only the first two dimensions of the data set contain important information, finding a subspace containing important information is very important for clustering. We visualized the original data sets represented by the first two features in Figure 2a.

In our experiment, we compared the clustering result of our proposed method with two commonly used dimensionality reduction methods, such as principal component analysis (PCA) and local retention prediction (LPP), and then visualized all the results in Figure 2.

From Figure 2, we knew that both LPP and PCA methods (*i.e.*, Figure 2b and c, respectively) did not find an effective low-dimensional subspace to output reasonable clustering results. However, our proposed method finds subspaces correctly as it is almost identical to the subspaces formed by the first two important dimensions. As a result, our proposed

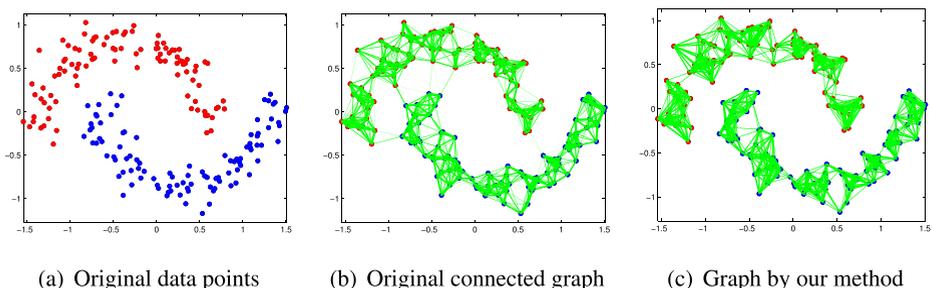
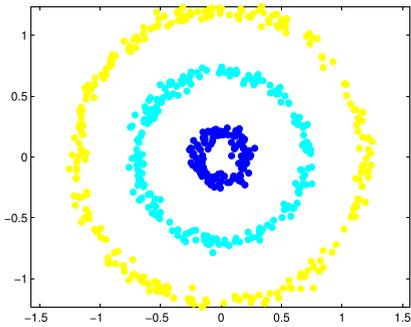
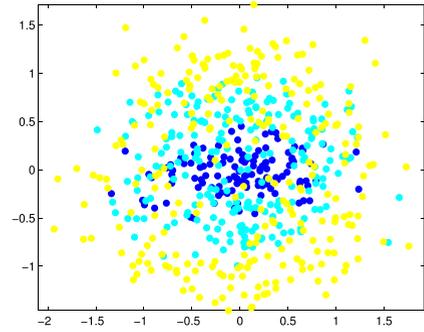


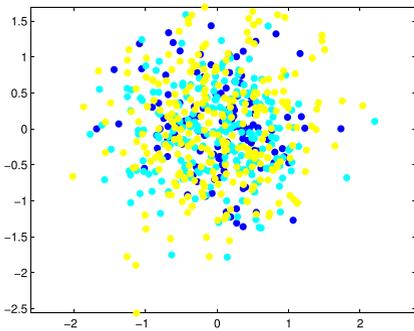
Figure 1 Visualization of the clustering results of our proposed method on the two-moon synthetic data



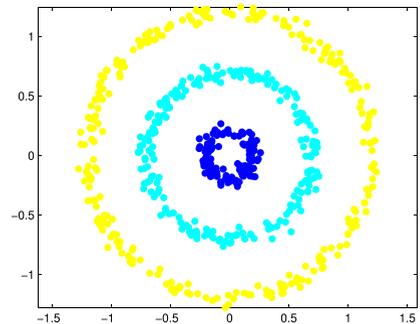
(a) Original data points



(b) Subspace learnt by LPP



(c) Subspace learnt by PCA



(d) Subspace learnt by our method

Figure 2 Visualization of the clustering results of our proposed method and two dimensionality reduction methods (*e.g.*, LPP and PCA) on the three-ring data set. We only visualized the first two features of the original data set in Figure 2a

method outputs feasible significant clustering results, as shown in Figure 2d. Hence, our proposed is able to deal with the influence of redundant features.

4.1.3 Summarization on two synthetic data sets

Based on the experimental results in Figures 1–2, our proposed method has been shown to be available individually dealing with either redundant features or outliers. In Section 4.2, we showed that our proposed method could deal with both redundant features and outliers on 12 real-world data sets, compared to the state-of-the-art clustering methods.

4.2 Experimental analysis on real benchmark data sets

4.2.1 Data sets

We evaluated our proposed clustering methods and the comparison methods on twelve benchmark data sets, whose details were summarized in Table 1.

Table 1 The details of all data sets used in our experiment

Dataset	# (samples)	# (features)	# (classes)
Wine	178	13	2
Ecoil	336	343	8
Control	600	60	6
Parkinsons	195	22	10
Movements	360	90	15
Vehicle	846	18	20
Jaffe	213	1024	15
German	1000	20	2
Usps	1000	256	10
Isolet	1560	617	26
Statlog	6435	36	6
Mfeatures	2000	649	10

4.2.2 Comparison methods

The comparison methods include two classic clustering methods (*i.e.*, NCut and k-means [8]), five state-of-the-art clustering methods (*i.e.*, Low-Rank Representation (LRR) [19], Constrained Laplacian Rank (CLR) [21], Sparse subspace clustering (SSC) [5]), and SMOoth Representation (SMR) [10]), and Structured Sparse Subspace Clustering (SSSC) [17].

We listed the details of the comparison methods as follows.

- **k-means** partitions all the data points into k groups so that the data points in the same group are with minimal mean.
- **NCut** partitions all the data points into k disjoint groups, where the weights of the edges between the groups are minimum.
- **LRR** seeks the lowest rank representation to linearly represent the data points.
- **SSC** first assumes that each point can be sparsely represented by all other data points, and then groups the data points drawn from multiple linear or affine subspaces of high-dimensional data.
- **SMR** makes use of the least square function and the trace norm regularization to enforce the grouping effect among representation coefficients, and then achieves the subspace segmentation.
- **CLR** learns the affinity matrix to have exactly k connected components where k is the number of clusters.
- **CSSC** integrate the affinity matrix learning with spectral clustering in a unified framework, as well as uses spectral clustering to correct errors in the affinity matrix.

4.2.3 Experimental setting

In the experiment, we employed four evaluation metrics (such as clustering ACCuracy (ACC), Normalized Mutual Information (NMI), F-measure and Adjusted Rand Index (ARI)) to evaluate the clustering performance of all the methods. We set the range of parameters (*i.e.*, λ and β) in our methods as $\lambda \in \{10^{-3}, 10^{-2}, \dots, 10^3\}$ and $\beta \in \{10^{-3}, 10^{-2}, \dots, 10^3\}$, as well as set the parameters in the comparison methods based on the

literature so that all comparison methods achieved their best clustering performance in our experiments.

We listed the details of four evaluation metrics as follows:

$$ACC = \frac{N_{correct}}{N}, \quad (19)$$

where N represents the total sample number and $N_{correct}$ represents the sample number accurately clustered.

$$ARI = \frac{TP+TN}{TP+FP+FN+TN} \quad (20)$$

where TP means that two samples clustered in a class are correctly classified, TN means that two samples that should not be clustered in a class are correctly separated, FP means that samples that should not be clustered in a class are incorrectly placed in a class, and FN means that samples that should not be clustered in a class are incorrectly separated.

$$F - measure = \frac{(\beta^2+1)PR}{\beta^2P+R} \quad (21)$$

where $P = \frac{TP}{TP+FP}$ and $R = \frac{TP}{TP+FN}$.

$$NMI = \frac{H(A)+H(B)}{H(A,B)} \quad (22)$$

where A and B are clustering results, and $H(A, B)$ is the joint entropy of A and B .

4.2.4 Experimental analysis of real-world data sets

We listed the clustering results of all the methods on these real data sets in Tables 2, 3, 4 and 5 and had the following observations.

Our proposed method achieved the best performance, followed by CSSC, CLR, SMR, SSC, LRR, NCut, and k-means. For example, the accuracy of our method improved on average by 5.29%, compared to the best comparison methods (*i.e.*, CSSC) on all 12 data sets. It may be that our method uses subspace learning to learn the affinity matrix from the low-dimensional subspace so that our method removes the influence of outliers and redundant features as well as uses robust estimation to avoid the influence of outliers in the data. On the contrary, the comparison algorithm cannot take into account the effects of these two constraints at the same time, so that they could obtain reliable models.

4.3 Parameters sensitivity

Two parameters need to be adjusted in our proposed objective function. In this section, we changed the values of λ and β by setting their ranges as $\{10^{-3}, 10^{-2}, \dots, 10^3\}$ to investigate the variations of the clustering accuracy of our method. We listed the results on all data sets in Figure 3.

The results showed that our method is sensitive to parameters setting. Moreover, different ranges of the parameter values need to be set for different data sets to get the best results. From example, our proposed method easily achieves significant clustering results while setting $\beta \in \{10^1, 10^2, 10^3\}$. Moreover, the parameter β is more sensitive than the parameter λ in our objective function.

Table 2 Clustering results on three public data sets, *i.e.*, Wine, Ecoli, and Control

Data sets	Wine				Ecoli				Control			
	NMI	ACC	ARI	F-measure	NMI	ACC	ARI	F-measure	NMI	ACC	ARI	F-measure
k-means	0.4287	0.7022	0.3711	0.5835	0.4696	0.5327	0.3952	0.5318	0.7011	0.5683	0.5853	0.6678
NCut	0.4214	0.7303	0.3981	0.5999	0.5025	0.5268	0.3498	0.4635	0.7920	0.6850	0.6588	0.7228
LRR	0.6890	0.9045	0.7196	0.8167	0.3824	0.6310	0.4615	0.6259	0.2020	0.2517	0.0798	0.2667
SSC	0.4277	0.7135	0.3743	0.5850	0.4614	0.6282	0.3974	0.5157	0.6653	0.7050	0.5840	0.6593
CLR	0.3927	0.7247	0.4007	0.6008	0.6696	0.8006	0.7430	0.8187	0.7015	0.7400	0.5594	0.6370
SMR	0.7973	0.9326	0.8009	0.8688	0.6494	0.7351	0.6171	0.7351	0.7286	0.7400	0.6476	0.7101
CSSC	0.8333	0.9551	0.8621	0.9086	0.6254	0.7589	0.7005	0.7803	0.7133	0.7567	0.5768	0.6502
Proposed	0.9196	0.9831	0.9472	0.9650	0.6579	0.8095	0.7171	0.7932	0.7912	0.8450	0.6942	0.7470

Bold number represents the maximum in the whole column

Table 3 Clustering results on three public data sets, *i.e.*, Parkinsons, Movement, and Vehicle

Data sets	Parkinsons				Movement				Vehicle			
	NMI	ACC	ARI	F-measure	NMI	ACC	ARI	F-measure	NMI	ACC	ARI	F-measure
k-means	0.0135	0.6564	0.0527	0.6256	0.5786	0.4556	0.3094	0.3592	0.1250	0.4078	0.1014	0.2832
NCut	0.0135	0.6564	0.0527	0.6256	0.6238	0.4917	0.3947	0.3958	0.1815	0.4492	0.1216	0.3575
LRR	0.0696	0.6154	0.0488	0.5774	0.5891	0.5056	0.3351	0.3789	0.1003	0.3842	0.0715	0.3141
SSC	0.0130	0.7590	0.0210	0.7725	0.6453	0.5169	0.3998	0.4412	0.1689	0.4669	0.1470	0.3709
CLR	0.1182	0.7435	0.2125	0.6743	0.6385	0.5083	0.3688	0.4139	0.2471	0.4574	0.1371	0.3681
SMR	0.0774	0.6308	0.0647	0.5837	0.6514	0.5333	0.4014	0.4444	0.2323	0.4704	0.1823	0.4223
CSSC	0.0191	0.7231	0.0841	0.9198	0.7130	0.6250	0.4978	0.5343	0.3204	0.5095	0.2129	0.4481
Proposed	0.1858	0.8508	0.3304	0.8029	0.7180	0.6306	0.5283	0.5630	0.3477	0.5378	0.2306	0.4517

Bold number represents the maximum in the whole column

Table 4 Clustering results on three public data sets, *i.e.*, Jaffe, German, and Usps

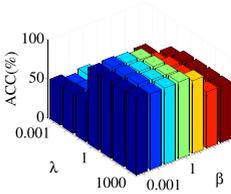
Data sets	Jaffe				German				Usps			
	NMI	ACC	ARI	F-measure	NMI	ACC	ARI	F-measure	NMI	ACC	ARI	F-measure
k-means	0.5786	0.4556	0.3094	0.3592	0.1250	0.4078	0.1014	0.2832	0.6035	0.6640	0.4821	0.5344
NCut	0.6238	0.4917	0.3947	0.3958	0.1815	0.4492	0.1216	0.3575	0.7913	0.7840	0.7196	0.7493
LRR	0.5891	0.5056	0.3351	0.3789	0.1003	0.3842	0.0715	0.3141	0.6909	0.8060	0.6254	0.6630
SSC	0.6453	0.5169	0.3998	0.4412	0.1689	0.4669	0.1470	0.3709	0.7973	0.7890	0.7155	0.7455
CLR	0.6385	0.5083	0.3688	0.4139	0.2471	0.4574	0.1371	0.3681	0.7849	0.8690	0.7420	0.7677
SMR	0.6514	0.5333	0.4014	0.4444	0.2323	0.4704	0.1823	0.4223	0.7991	0.7810	0.7136	0.7440
CSSC	0.7130	0.6250	0.4978	0.5343	0.3204	0.5095	0.2129	0.4481	0.7808	0.8427	0.7311	0.7421
Proposed	0.7180	0.6306	0.5283	0.5630	0.3477	0.5378	0.2306	0.4517	0.7927	0.8760	0.7542	0.7787

Bold number represents the maximum in the whole column

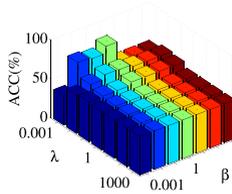
Table 5 Clustering results on three public data sets, *i.e.*, Isolet, Statlog, and Mfeatures

Data sets	Isolet				Statlog				Mfeatures			
	NMI	ACC	ARI	F-measure	NMI	ACC	ARI	F-measure	NMI	ACC	ARI	F-measure
k-means	0.7396	0.5583	0.5115	0.5323	0.5037	0.6289	0.3980	0.5144	0.5484	0.5070	0.3544	0.4505
NCut	0.7834	0.6737	0.5945	0.6105	0.6383	0.6281	0.5250	0.6126	0.7638	0.8025	0.6811	0.7312
LRR	0.7597	0.6481	0.5704	0.5876	0.5185	0.6134	0.4479	0.5883	0.7683	0.8575	0.7284	0.7556
SSC	0.8199	0.7327	0.6537	0.6678	0.6147	0.6692	0.5442	0.6345	0.9124	0.9605	0.9146	0.9231
CLR	0.8468	0.7776	0.7047	0.7166	0.6452	0.7082	0.5785	0.6601	0.8918	0.8680	0.8430	0.8598
SMR	0.7814	0.6571	0.5883	0.6046	0.4942	0.6609	0.4292	0.5491	0.7699	0.8370	0.7112	0.7404
CSSC	0.7233	0.7654	0.7155	0.6911	0.6552	0.7072	0.5965	0.6754	0.8853	0.8745	0.8337	0.8520
Proposed	0.7282	0.8350	0.7692	0.7193	0.6669	0.7263	0.6071	0.6837	0.9482	0.9780	0.9617	0.9565

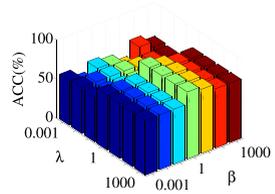
Bold number represents the maximum in the whole column



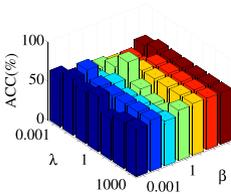
(a) Wine



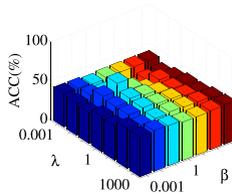
(b) Ecoli



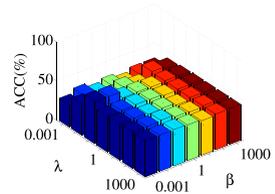
(c) Control



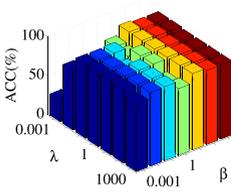
(d) Parkinsons



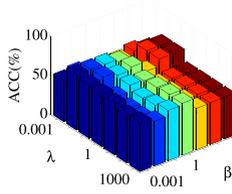
(e) Movements



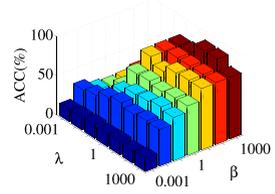
(f) Vehicle



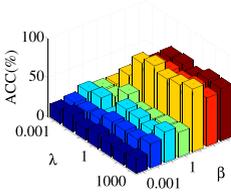
(g) Jaffe



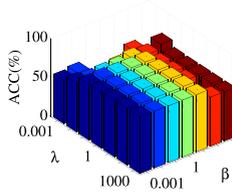
(h) German



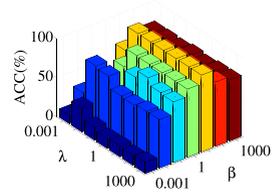
(i) Usps



(j) Isolet



(k) Statlog



(l) Mfeatures

Figure 3 ACC variations of our method at different parameter settings on 12 data sets

5 Conclusion

This paper has proposed a new clustering method, which learns the affinity matrix from the low-dimensional space of original data as well as uses the robust estimator to reduce the influence of outliers. Experimental results demonstrate the effectiveness of the method for clustering tasks.

Acknowledgments This work is partially supported by the China Key Research Program (Grant No: 2016YFB1000905), the Natural Science Foundation of China (Grants No: 61836016, 61876046, 61573270, and 61672177), the Project of Guangxi Science and Technology (GuiKeAD17195062), the Guangxi Collaborative Innovation Center of Multi-Source Information Integration and Intelligent Processing, the Guangxi High Institutions Program of Introducing 100 High-Level Overseas Talents, the Strategic Research Excellence Fund at Massey University, the Marsden Fund of New Zealand (Grant No: MAU1721), and the Research Fund of Guangxi Key Lab of Multisource Information Mining and Security (18-A-01-01).

References

- Alberto, P., Casbon, J.A., Saqi, M.A.S.: Spectral clustering of protein sequences. *Nucleic Acids Res.* **34**(5), 1571–1580 (2006)
- Boyd, V., Faybusovich, L.: Convex optimization. *IEEE Trans. Autom. Control* **51**(11), 1859–1859 (2006)
- Charbonnier, P., Blancfraud, L., Aubert, G., Barlaud, M.: Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. Image Process.* **6**(2), 298–311 (1997)
- Edgar, R.C.: Search and clustering orders of magnitude faster than blast. *Bioinformatics* **26**(19), 2460 (2010)
- Elhamifar, E., Vidal, R.: Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Trans Pattern Anal Mach Intell* **35**(11), 2765–2781 (2013)
- Goh, A., Vidal, R.: Locally linear manifold clustering. *Journal of Machine Learning Research* (2009)
- Guangcan, L., Zhouchen, L., Shuicheng, Y., Ju, S., Yu, Y., Yi, M.: Robust recovery of subspace structures by low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(1), 171–184 (2013)
- Hartigan, J.A., Wong, M.A.: A k-means clustering algorithm. In: *ECCV*, pp. 492–503 (2008)
- He, R., Hu, B., Yuan, X., Wang, L.: M-Estimators and Half-Quadratic minimization (2014)
- Hu, H., Lin, Z., Feng, J., Zhou, J.: Smooth representation clustering. In: *CVPR*, pp. 3834–3841 (2014)
- Hu, M., Yang, Y., Shen, F., Xie, N., Hong, R., Shen, H.T.: Collective reconstructive embeddings for cross-modal hashing. *IEEE Transactions on Image Processing*. <https://doi.org/10.1109/TIP.2018.2890144> (2019)
- Huber, P.J.: *Robust statistics*. Wiley, New York (2011)
- Ji, Z., Tao, X., Wang, H.: Outlier detection from large distributed databases. *World Wide Web* **17**(4), 539–568 (2014)
- Junye, G., Liu, M., Zhang, D.: Spectral clustering algorithm based on effective distance. *Journal of Frontiers of Computer Science & Technology* (2014)
- Law, M.H.C., Figueiredo, J., Jain, A.K.: Simultaneous feature selection and clustering using mixture models. *IEEE Trans Pattern Anal Mach Intell* **26**(9), 1154–1166 (2004)
- Lei, C., Zhu, X.: Unsupervised feature selection via local structure learning and sparse learning. *Multimed Tools Appl* **77**(22), 29605–29622 (2018)
- Li, C.G., You, C., Vidal, R.: Structured sparse subspace clustering: A joint affinity learning and subspace clustering framework. *IEEE Trans. Image Process.* **26**(6), 2988–3001 (2017)
- Liu, X.Y., Jing-Wei, L.I., Hong, Y.U., You, Q.Z., Lin, H.F.: Adaptive spectral clustering based on shared nearest neighbors. *J. Chin. Comput. Syst.* **32**(9), 1876–1880 (2011)
- Liu, G., Lin, Z., Yan, S., Ju, S., Yu, Y., Ma, Y.: Robust recovery of subspace structures by low-rank representation. *IEEE Trans Pattern Anal Mach Intell* **35**(1), 171 (2013)
- Ng, A.Y., Jordan, M.I., Weiss, Y.: On spectral clustering: analysis and an algorithm. In: *NIPS*, pp. 849–856 (2001)
- Nie, F., Wang, X., Jordan, M.I., Huang, H.: The constrained laplacian rank algorithm for graph-based clustering. In: *AAAI*, pp. 1969–1976 (2016)
- Nikolova, M., Ng, M.K.: Analysis of Half-Quadratic minimization methods for signal and image recovery. *Society for Industrial and Applied Mathematics* (2005)
- Peng, Y., Zhu, Q., Huang, B.: Spectral clustering with density sensitive similarity function. *Knowl.-Based Syst.* **24**(5), 621–628 (2011)
- Roth, V., Lange, T.: Feature selection in clustering problems. In: *NIPS* (2003)
- Shah, S.A., Koltun, V.: Robust continuous clustering. *Proc Natl Acad Sci U SA* **114**(37), 9814–9819 (2017)
- Szu-Hao, H., Yi-Hong, C., Shang-Hong, L., Novak, C.L.: Learning-based vertebra detection and iterative normalized-cut segmentation for spinal mri. *IEEE Trans. Med. Imaging* **28**(10), 1595–1605 (2009)

27. Tao, X., Li, Y., Zhong, N.: A personalized ontology model for Web information gathering. *IEEE Trans Knowl Data Eng* **23**(4), 496–511 (2010)
28. Wang, R., Zong, M.: Joint self-representation and subspace learning for unsupervised feature selection. *World Wide Web* **21**(6), 1745–1758 (2018)
29. Wang, W.W., Xiao-Ping, L.I., Feng, X.C., Si, Q.i.W.: A survey on sparse subspace clustering. *Acta Autom. Sin.* **41**(8), 1373–1384 (2015)
30. Wang, L., Zheng, K., Tao, X., Han, X.: Affinity propagation clustering algorithm based on large-scale data-set. *International Journal of Computers & Applications* (3), 1–6 (2018)
31. Wang, R., Ji, W., Liu, M., Wang, X., Weng, J., Deng, S., Gao, S., Yuan, C.-a.: Review on mining data from multiple data sources. *Pattern Recogn. Lett.* **109**, 120–128 (2018)
32. Wang, R., Ji, W., Song, B.: Durable relationship prediction and description using a large dynamic graph. *World Wide Web* **21**(6), 1575–1600 (2018)
33. Yan, J., Pollefeys, M.: A general framework for motion segmentation Independent, articulated, rigid, non-rigid, degenerate and non-degenerate. In: *ECCV*, pp. 94–106 (2006)
34. Yi, B., Yang, Y., Shen, F., Xie, N., Shen, H.T., Li, X.: Describing video with attention based bidirectional lstm. *IEEE Transactions on Cybernetics* (2018)
35. Zelnik-Manor, L., Perona, P.: Self-tuning spectral clustering. In: *NIPS*, pp. 1601–1608 (2005)
36. Zhang, S., Li, X., Ming, Z., Zhu, X., Cheng, D.: Learning k for knn classification. *Acm Trans Intell Syst Technol* **8**(3), 43 (2017)
37. Zhang, S., Li, X., Zong, M., Zhu, X., Wang, R.: Efficient knn classification with different numbers of nearest neighbors. *IEEE Trans Neural Netw Learn Syst* **29**(5), 1774–1785 (2018)
38. Zheng, W., Zhu, X., Zhu, Y., Hu, R., Lei, C.ong.: Dynamic graph learning for spectral feature selection. *Multimed Tools Appl* **77**(22), 29739–29755 (2018)
39. Zhou, X., Shen, F., Liu, L.i., Liu, W., Nie, L., Yang, Y., Shen, H.T.: Graph convolutional network hashing (2018)
40. Zhu, X., Li, X., Zhang, S.: Block-row sparse multiview multilabel learning for image classification. *IEEE Trans Cybern* **46**(2), 450–461 (2016)
41. Zhu, X., He, W., Li, Y., Yang, Y., Zhang, S., Hu, R., Xhu, Y.: One-step spectral clustering via dynamically learning affinity matrix and subspace, pp. 2963–2969 (2017)
42. Zhu, X., Li, X., Zhang, S., Ju, C., Wu, X.: Robust joint graph sparse coding for unsupervised spectral feature selection. *IEEE Trans Neural Netw Learn Syst* **28**(6), 1263–1275 (2017)
43. Zhu, X., Li, X., Zhang, S., Xu, Z., Yu, L., Wang, C.: Graph pca hashing for similarity search. *IEEE Trans Multimed* **19**(9), 2033–2044 (2017)
44. Zhu, X., Shichao, Z., Li, Y., Jilian, Z., Lifeng, Y., Yue, F.: Low-rank sparse subspace for spectral clustering. *IEEE Trans Knowl Data Eng* **31**(8), 1532–1543 (2019)
45. Zhu, X., Zhang, S., Hu, R., He, W., Lei, C., Zhu, P.: One-step multi-view spectral clustering. *IEEE Transactions on Knowledge and Data Engineering*. <https://doi.org/10.1109/TKDE.2018.2873378> (2018)
46. Zhu, X., Zhang, S., Li, Y., Zhang, J., Yang, L., Fang, Y.: Low-rank sparse subspace for spectral clustering. *IEEE Transactions on Knowledge and Data Engineering*. <https://doi.org/10.1109/TKDE.2018.2858782>
47. Zhu, X., Zhu, Y., Zhang, S., Hu, R., He, W.: Adaptive hypergraph learning for unsupervised feature selection, 3581–3587, pp. (2018)
48. Zhu, Y., Zhu, X., Zheng, W.: Robust multi-view learning via half-quadratic minimization, pp. 3278–3284 (2018)

Affiliations

Xiaofeng Zhu^{1,2} · Jiangzhang Gan² · Guangquan Lu³ · Jiaye Li³ · Shichao Zhang⁴

Xiaofeng Zhu
seanzhuxf@gmail.com

Jiangzhang Gan
j.gan@massey.ac.nz

Guangquan Lu
Guangquan85@gmail.com

Jiaye Li
Jia_yeligxnu@126.com

¹ University of Electronic Science and Technology of China, Chengdu 611731, China

² SNCS of Massey University Auckland Campus, Auckland 0745, New Zealand

³ Guangxi Key Lab of MIMS, Guangxi Normal University, Guilin, 541004, China

⁴ Central South University, Changsha 410083, China