Spectral Embedded Adaptive Neighbors Clustering

Qi Wang, *Senior Member, IEEE,* Zequn Qin, Feiping Nie, Xuelong Li, *Fellow, IEEE*

***Abstract*—Spectral clustering has been widely used in various aspects especially the machine learning fields. Clustering with similarity matrix and low-dimensional representation of data is the main reason of its promising performance shown in spectral clustering. However, such similarity matrix and low-dimensional representation directly derived from input data may not always hold when the data is high-dimensional and has complex distribution. First, the similarity matrix simply based on distance measurement might not be suitable for all kinds of data. Second, the low-dimensional representation might not be able to reflect the manifold structure of the original data. In this paper, we propose a novel linear space embedded clustering method which uses adaptive neighbors to address the above problems. Linearity regularization is used to make the data representation have a linear embedded spectral. We also use adaptive neighbors to optimize the similarity matrix and clustering results simultaneously. Extensive experimental results show promising performance compared with other state-of-the-art algorithms.**

***Index Terms*—machine learning, spectral clustering, spectral embedded clustering, adaptive neighbors.**

I. Introduction

CLUSTERING is a fundamental approach in the machine learning fields. Many clustering methods have been successfully applied in the data mining applications. Among them, the spectral clustering shows remarkable results because it can capture the structure information which conforms to manifold assumption. Another property of spectral clustering is that it is a graph based algorithm. The manifold assumption means that the nearby data points should have the same labels because the data should be formed in a certain manifold structure. This assumption can hold on many low-dimensional data while the high-dimensional or sparse data might violate this hypothesis. In [1], [2], rank constraint of Laplacian matrix is added to update the similarity graph and get a better result. In [3], a kernel based support vector clustering method is proposed to capture more complex data structure. In [4], [5] and [6], several methods which find the largest margin in the low-dimensional data space is proposed. In[7], a spectral embedded clustering method is proposed to address this problem. The main idea of this method is that they use a linear mapping to make the data fit the manifold assumption. In this way, the spectral clustering can be extended to high-dimensional data. Although many methods are proposed to fix the problem of high-dimensional and sparse data, the challenge in reasonably representing data still exists. In [8], a novel convex algorithm is proposed to address the problem of altered manifold properties in reduced-dimensional subspace. The starting point of low-dimensional representation is the same as this work, but we utilized a different way to solve this problem.

Another part of difficulty in clustering data is the backend process after the data representation. Many methods use K-means [9], [10], [11] as its backend process to obtain the discrete clustering labels. K-means is a simple and effective method which might be the most widely used one, but there is still one big problem, which is the initialization problem. Without good enough initialization, the K-means method might perform badly.

In order to address these problems, namely the difficulties in stably representing and clustering data, we propose the spectral embedded adaptive neighbors clustering (SEANC) algorithm. SEANC uses a two-stage framework containing spectral embedded representation of data and adaptive neighbors clustering. The embedded representation stage is used to process the high-dimensional data which might violate the manifold assumption. The adaptive neighbors clustering stage is used to optimize the similarity graph and obtain a stable clustering results. The novelty of our method can be summarized in three parts: 1. The proposed method can learn an effective linear embedded representation that can handle high-dimensional and sparse data. 2. The proposed method is proved to be a generalization of PCAN [12]. 3. The proposed method gains good performance when comparing with some state-of-the-art methods. Meanwhile, the effectiveness of our method when dealing high-dimensional and sparse data is examined.

A. Notation

All of the matrices and vectors are written in uppercase while the scalars are written in lowercase. For instance, M represents a matrix, M_i is the i-th column and M_{ij} is the ij-th element of M. $\|\cdot\|_F$ and $\|\cdot\|_2$ mean the Frobenius norm and the L2 norm respectively. $Tr(\cdot)$ demonstrates the trace operation.

This work was supported by the National Key R&D Program of China under Grant 2017YFB1002202, National Natural Science Foundation of China under Grant 61773316, Natural Science Foundation of Shaanxi Province under Grant 2018KJXX-024, Fundamental Research Funds for the Central Universities under Grant 3102017AX010, and the Open Research Fund of Key Laboratory of Spectral Imaging Technology,Chinese Academy of Sciences.

Q. Wang is with the School of Computer Science, with the Unmanned System Research Institute, and with the Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, China. E-mail: crabwq@gmail.com.

Z. Qin is with the School of Computer Science and the Center for Optical Imagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China. E-mail: qinzequn@mail.nwpu.edu.cn.

F. Nie is with the School of Computer Science and the Center for Optical Imagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China. E-mail: feipingnie@gmail.com.

X. Li is with the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, P. R. China and with the University of Chinese Academy of Sciences, Beijing 100049, P. R. China. Email: xuelong_li@opt.ac.cn

II. Related Work

A. Revisiting Spectral Clustering

The spectral clustering method can be seen as a graph theory based method. The main target of spectral clustering is to divide a dataset $X = \{x_1, \cdots, x_n\} \in R^{d\times n}$ into c parts, in which d and n are the dimension and the number of input data respectively. Based on the graph theory, the spectral clustering is to find the optimal graph partition and separate the data. Denote $G = (X, E)$ as an undirected graph. Let W be the weighted adjacency matrix which demonstrates the relationship of data. We have [13]:

$$W_{ij} = \exp(-\frac{\|x_i - x_j\|_2^2}{2\sigma^2}). \tag{1}$$

The parameter σ is used for adjusting the neighborhood size. The adjacency matrix can be changed with the largest k elements reserved. Define D as the diagonal degree matrix composed of $D_{ii} = \sum_{j=1}^{n} W_{ij}$. The Laplacian matrix is defined as:

$$L = D - W. \tag{2}$$

In some works, normalized Laplacian matrices are used:

$$L_{sym} = I - D^{-\frac{1}{2}} W D^{\frac{1}{2}} \tag{3}$$

and

$$L_{rw} = I - D^{-1} W. \tag{4}$$

We use L^* to demonstrate these forms of Laplacian matrix. The optimization of spectral clustering can be formulated as [14]:

$$\begin{aligned} &\min_F Tr(F^T L^* F), \\ &s.t.\ F^T F = I. \end{aligned} \tag{5}$$

The optimal solution to F is composed of eigen-decomposition. It is composed of eigenvectors corresponding to the Laplacian matrix, while the obtained F is the relaxed solution compared with discrete form of labels.

One way to convert the partition matrix F into labels is using K-means [15]. Another method is to utilize the orthonormal invariance of partition. Denote $\epsilon(\cdot)$ as the objective function of spectral clustering. The relation between partition and rotation matrices are:

$$\begin{aligned} &\epsilon(FR) = \epsilon(F), \\ &s.t.\ R^T R = I. \end{aligned} \tag{6}$$

In [14], the inverse mapping from partition matrix F to label Y_F is defined as:

$$Y_F = f^{-1}(F) = Diag(Diag(FF^T)^{-\frac{1}{2}})F. \tag{7}$$

Then, the optimization of discrete solution can be formulated as:

$$\begin{aligned} &\min_{Y,R} \|Y - Y_F R\|^2, \\ &s.t.\ Y \in \{0,1\}^{d\times n}, Y1 = 1, R^T R = I. \end{aligned} \tag{8}$$

B. Nonlinear Embedding

In [16], two kinds of spectral nonlinear embedding methods are proposed. The first one is the kernel-based method. The nonlinear embedding problem can be written as:

$$\begin{aligned} &\min_{F,\alpha} Tr(F^T L^* F) + \mu(\|K\alpha - F\|^2 + \gamma_g Tr(\alpha^T K \alpha)), \\ &s.t. F^T F = I, \end{aligned} \tag{9}$$

in which K is a symmetric kernel matrix and $\alpha = [\alpha_1, \cdots, \alpha_n]^T \in R^{n\times c}$.

The second one is the spectral nonlinearly embedded clustering model based on extreme learning machine (ELM). We have:

$$\begin{aligned} &\min_{F,\beta} Tr(F^T L^* F) + \mu(\|H_{elm}\beta - F\|^2 + \gamma_g Tr(\beta^T \beta)), \\ &s.t. F^T F = I, \end{aligned} \tag{10}$$

where H_{elm} represents the output of ELM.

In fact, both these two methods utilize a nonlinear embedding function to fit the manifold assumption of original data and gains the ability of handling the out-sample data.

III. General Framework for SEANC

In this section, we will describe the detailed derivation of SEANC. In order to deal with high-dimensional or sparse data which breaks the manifold assumption, we use linear embedded low-dimensional representation [7]. Different from the previous methods, we treat the low-dimensional representation as a kind of data itself and cluster with such representation instead of the original data. We use a two-stage method to address the above problems.

A. Spectral Embedded Representation

According to [7], the clustering results can be regarded as a linear mapping of the original data, which can be written as:

$$Y = X^T W + 1b^T, \tag{11}$$

in which Y is the one-hot assignment matrix, $W \in R^{d\times c}$ and $b \in R^{c\times 1}$ are the mapping matrices. The definition of X is the same as related work section. This assumption has been well studied in [17], [18], which holds for most high-dimensional data.

As shown in the related work section, the objective function of spectral clustering is:

$$\begin{aligned} &\min_F Tr(F^T L^* F), \\ &s.t.\ F^T F = I. \end{aligned} \tag{12}$$

In order to make the data have more useful low-dimensional representation, the linear embedded regularization term is added. The basic idea is to minimize the differences between the relaxed assignment matrix F and the label matrix Y. In this way, the spectral embedded representation can be written as:

$$\begin{aligned} &\min_{F,W,b} Tr(F^T L^* F) + \lambda(\|Y - F\|^2 + \mu Tr(W^T W)), \\ &s.t.\ F^T F = I, \end{aligned} \tag{13}$$

in which λ and μ are the parameters used for adjust the weight of different tasks and regularization terms.

In order to optimize Eq. 13, we utilize eigenvalue decomposition to solve this problem. Making the derivatives of W and b in Eq. 13 into zero, the results can be:

$$b = \frac{1}{n}F^T 1, \tag{14}$$

$$W = (XX^T + \mu I)^{-1}X. \tag{15}$$

Substituting W and b into Eq. 13, the objective function can be:

$$\begin{aligned} &\min_F Tr(F^T(L^* + \lambda(H_n - X^T(XX^T + \mu I)^{-1})X)F), \\ &s.t.\ F^T F = I, \end{aligned} \tag{16}$$

in which H_n is the centering matrix.

The optimal solution to Eq. 16 can be regarded as a new kind of representation extracted from the original data. Such representation contains both spectral and linear data mapping information. Different from the classical spectral clustering method, we use an adaptive neighbors method instead of the K-means method to get the cluster assignment matrix.

B. Adaptive Neighbors Clustering

Our goal is to find a method that is able to find the cluster assignment matrix from the above spectral embedded representation F. Based on Ky Fan's theorem [19], a similarity matrix which contains c components can be used for clustering with c classes. In this way, we can convert the cluster assignment problem into the following ways [20], [21]:

$$\begin{aligned} &\min_S \sum_{i,j=1}^{n} \|f_i - f_j\|_2^2 s_{ij}, \\ &s.t.\ s_i^T 1 = 1, s_{ij} > 0, \end{aligned} \tag{17}$$

in which S is the similarity matrix with $s_i \in R^{n\times 1}$ as a vector and f_i is the i-th sample from the embedded spectral output F. However, the above problem has a trivial solution that the nearest data point is the most similar one. In order to tackle this problem, we use a prior knowledge which doesn't consider the relationships between the spectral embedded samples:

$$\begin{aligned} &\min_S \sum_{i,j=1}^{n} s_{ij}^2, \\ &s.t.\ s_i^T 1 = 1, s_{ij} > 0. \end{aligned} \tag{18}$$

By combining Eq. 17 and Eq. 18, we have the adaptive neighbors objective function and its matrix form:

$$\begin{aligned} &\min_S \sum_{i,j=1}^{n} (\|f_i - f_j\|_2^2 s_{ij} + \alpha s_{ij}^2) \\ &\Leftrightarrow \min_S Tr(F^T L_F^* F) + \alpha \|S\|_F^2, \\ &\qquad s.t.\ s_i^T 1 = 1, s_{ij} > 0. \end{aligned} \tag{19}$$

in which L_F^* is the Laplacian matrix corresponding to the spectral embedded representation F and α is the regularization parameter.

According to Ky Fan's theorem, if the clustering results have c clusters, the similarity matrix should have exactly c components. Also, we have the following theorem [22], [23]:

Theorem 1: The multiplicity c of the eigenvalue 0 of the Laplacian matrix L (nonnegative) is equal to the number of connected components in the graph with the similarity matrix S.

From theorem 1 we can see that the number of cluster is subject to the rank of Laplacian matrix. If the rank of Laplacian matrix equals $n - c$, the cluster result would have exactly c clusters. As a result, we add the rank constraint to the original objective function, which is:

$$\begin{aligned} &\min_S Tr(F^T L_F^* F) + \alpha \|S\|_F^2, \\ &s.t.\ s_i^T 1 = 1, s_{ij} > 0, rank(L_F^*) = n - c. \end{aligned} \tag{20}$$

C. Optimization

In the previous sections, we propose a two-stage method. The optimization of the spectral embedded representation in Eq. 16 is relatively simple while the adaptive neighbors clustering with rank constraint is difficult to solve. In this section, we propose the optimization approach to solve Eq. 20.

Without loss of generality, denote $\{e_1, \cdots, e_n\}$ as the eigenvalues of Laplacian matrix L_F^* in ascending order. As shown in [19], we have:

$$\begin{aligned} \sum_{i=1}^{c} e_i = &\min_G Tr(G^T L_F^* G), \\ &s.t.\ G^T G = I, G \in R^{n\times c}. \end{aligned} \tag{21}$$

Meanwhile, the rank constraint in Eq. 20 can be seen as the optimization of the smallest c eigenvalues of the Laplacian matrix. In this way, the problem in Eq. 20 can be rewritten as:

$$\begin{aligned} \min_S Tr(F^T L_F^* F) + \alpha \|S\|_F^2 + \beta \sum_{i=1}^{c} e_i, \\ s.t.\ s_i^T 1 = 1, s_{ij} > 0. \end{aligned} \tag{22}$$

in which β is a large enough regularization parameter. Combing Eq. 21 and Eq. 22, we have:

$$\begin{aligned} &\min_{S,G} Tr(F^T L_F^* F) + \alpha \|S\|_F^2 + \beta Tr(G^T L_F^* G), \\ &s.t.\ s_i^T 1 = 1, s_{ij} > 0, G^T G = I, G \in R^{n\times c}. \end{aligned} \tag{23}$$

When S is fixed, the first two terms of Eq. 23 are constant. So the optimization of G can be:

$$\begin{aligned} &\min_G Tr(G^T L_F^* G), \\ &s.t.\ G^T G = I, G \in R^{n\times c}. \end{aligned} \tag{24}$$

This problem can be solved by eigenvalue decomposition with c eigenvectors corresponding to the smallest c eigenvalues.

When G is fixed, those constraints related with G can be discarded. In this case, the optimization of S can be:

$$\begin{aligned} &\min_S Tr(F^T L_F^* F) + \alpha \|S\|_F^2 + \beta Tr(G^T L_F^* G), \\ &s.t.\ s_i^T 1 = 1, s_{ij} > 0. \end{aligned} \tag{25}$$

In order to have an easier explanation, we use the vector form of Eq. 25. The vector form of $Tr(G^T L_F^* G)$ can be:

$$Tr(G^T L_F^* G) = \sum_{i,j=1}^{n} \|g_i - g_j\|_2^2 s_{ij}. \tag{26}$$

in which $g_i \in R^{1\times c}$ is a row of G. The vector form of Eq. 25 is:

$$\begin{aligned} \min_S \sum_{i,j=1}^{n} (\|f_i - f_j\|_2^2 s_{ij} + \alpha s_{ij}^2 + \beta \|g_i - g_j\|_2^2 s_{ij}), \\ s.t.\ s_i^T 1 = 1, s_{ij} > 0. \end{aligned} \tag{27}$$

We should notice a trick that the solution of Eq. 27 can be divided into every column of S. Then we have:

$$\begin{aligned} \min_{s_i} \sum_{i,j=1}^{n} (\|f_i - f_j\|_2^2 s_{ij} + \alpha s_{ij}^2 + \beta \|g_i - g_j\|_2^2 s_i j), \\ s.t.\ s_i^T 1 = 1, s_i > 0. \end{aligned} \tag{28}$$

Denote $d_{ij}^f = \|f_i - f_j\|_2^2$, $d_{ij}^g = \|g_i - g_j\|_2^2$ and $d_i \in R^{n\times 1}$ as a vector with $d_{ij} = d_{ij}^f + \beta d_{ij}^g$. Then the Eq. 28 can be:

$$\begin{aligned} \min_{s_i} \left\| s_i + \frac{1}{2\alpha} d_i \right\|_2^2, \\ s.t.\ s_i^T 1 = 1, s_i > 0. \end{aligned} \tag{29}$$

The solution of Eq. 29 is:

$$s_{ij} = (-\frac{d_{ij}^f}{2\alpha_i} + \eta_i)_+. \tag{30}$$

in which η_i and α_i are Lagrange multiplier for every i.

The final two-stage objective functions can be:

$$\min_F Tr(F^T(L^* + \lambda(H_n - X^T(XX^T + \mu I)^{-1})X)F),$$

$$s.t.\ F^T F = I. \tag{31}$$

and

$$\begin{aligned} \min_{S,G} Tr(F^T L_F^* F) + \alpha \|S\|_F^2 + \beta Tr(G^T L_F^* G), \\ s.t.\ s_i^T 1 = 1, s_{ij} > 0, G^T G = I, G \in R^{n\times c}. \end{aligned} \tag{32}$$

Here, we propose our SEANC Algorithm.

The SEANC algorithm can be seen as an iterative combination of low-dimensional representation and adaptive neighbors clustering. In this way, the computational complexity should be no higher than any of these two methods. In consideration of the matrix formation of our method, the most computational expensive parts are the eigen-decomposition and inverse calculation. In this way, the computational complexity should be $\mathcal{O}(n^3)$. Frankly speaking, our method achieves the same speed performance as most of the spectral methods while obtains better performance.

Algorithm 1 SEANC Algorithm

Input: Data $X = \{X_1, X_2, \cdots, X_n\} \in R^{d\times n}$, classes number c, regularization parameters $\sigma, k, \lambda, \mu, \alpha, \beta$.
Output: Cluster assignment matrix $C \in R^{n\times 1}$

1: Calculate Laplacian matrix L^* using Eq.3 or Eq. 4 and $L^* + \lambda(H_n - X^T(XX^T + \mu I)^{-1})X$
2: Use eigenvalue decomposition in Eq. 31 to get the representation F
3: Compute the Laplacian matrix L_F^* using the representation F and Eq. 3 or Eq. 4
4: **while** not converge **do**
5: Update G with eigenvectors corresponding to c smallest eigenvalues of L_F^*
6: Update S with Eq 30.
7: **end while**
8: Compute assignment matrix C using S and Ky Fan's theorem

D. Connnection to PCAN

In the SEANC algorithm, the spectral embedded representation is optimized with linear embedded data for high-dimensional problems. In fact, the linear embedded data can be seen as a kind of projection [24]. In view of this projection, we can see the relationships between SEANC and PCAN [12]. The proposed SEANC has a similar solution with PCAN when $\lambda \to \infty$ and $\mu = 0$.

Proof: In [12], the definition of PCAN is:

$$\begin{aligned} &\min_{S,W} \sum_{i,j=1}^{n} (\|x_i^T W - x_j^T W\|_2^2 + \gamma s_{ij}^2), \\ &s.t.\ s_i^T 1 = 1, s_{ij} > 0, W^T S_t W = I, \\ &rank(L_s) = n - c, \end{aligned} \tag{33}$$

in which $S_t = X H_n X^T$ is the scatter matrix. The definition of SEANC is:

$$\begin{aligned} &\min_S Tr(F^T L_F^* F) + \alpha \|S\|_F^2, \\ &s.t.\ s_i^T 1 = 1, s_{ij} > 0, F^T F = I, \\ &rank(L_F^*) = n - c. \end{aligned} \tag{34}$$

When $\lambda \to \infty$ and $\mu = 0$, the optimal solution to Eq. 13 can be:

$$F = X^T W + 1b^T. \tag{35}$$

Combined with Eq. 20, we have:

$$\begin{aligned} &Tr(F^T L_F^* F) + \alpha \|S\|_F^2 \\ \Leftrightarrow &\sum_{i,j=1}^{n} (\|x_i^T W + 1b^T - x_j^T W - 1b^T\|_2^2 + \alpha s_{ij}^2) \\ \Leftrightarrow &\sum_{i,j=1}^{n} (\|x_i^T W - x_j^T W\|_2^2 + \alpha s_{ij}^2) \end{aligned} \tag{36}$$

Without consideration of the bias item, the constraints can be equivalent for SEANC and PCAN. Supposing the data is centralized, we have:

$$\begin{aligned} & F^T F = I \\ \Leftrightarrow & W^T X X^T W = I \\ \Leftrightarrow & W^T X H_n (X H_n)^T W = I \\ \Leftrightarrow & W^T X H_n X^T W = I \\ \Leftrightarrow & W^T S_t W = I \end{aligned} \tag{37}$$

The above proof demonstrates that the proposed method has a similar solution with PCAN when $\lambda \to \infty$ and $\mu = 0$. But there are two different factors between them. The first one is the discarded bias item when proving the equivalence of constraints. The second factor is the different optimization procedure. In our method, the spectral embedded representation is determined by the linear data mapping before the adaptive neighbors clustering stage, while the corresponding item in PCAN, data projection, is optimized in the whole optimization process. These differences ensure our method can make more use of spectral embedded information.

IV. Experiment

In order to verify the effectiveness of the proposed method, we use two kinds of data which are synthetic and real world data to test our method. In order to verify the effectiveness of high-dimensional data, a distinct experiment on high-dimensional and sparse text clustering is conducted in this section.

A. Experiment Setup

Due to the property that the results of K-means method are highly related to initialization, those experiments involving K-means are conducted for 100 times. The average performance and variance are recoded for comparison. Other methods that has closed-form or determined solution are tested for only once. The self-tuning spectral clustering [25] is used for determining the parameters. All of the results of our method run for 100 times and the best parameters are recorded.

As for the four parameters used in SEANC, the principles of setup are explained as follows. The parameter σ is used for adjusting the neighborhood size and k is used for determining the number of neighbors. Thus, these values are tuned according to different types of data. The values of λ and μ are in $[0.1, 0.2, 0.5, 0.8]$ and some random disturbances are added to μ for larger parameter space. The setup of α and β follows the implementation of [12].

B. Experiment on Synthetic Data

In this section, we use synthetic data to demonstrate the effectiveness of the proposed two-stage framework, especially the adaptive neighbors clustering stage.

As shown in Figure 1(a), the synthetic data in two semi-circles shape is utilized. According to Eq. 32, the adaptive neighbors clustering stage optimizes the clustering results and the spectral embedded similarity matrix simultaneously. The learned graph by similarity matrix is shown in Figure 1(b). We can observe that the learned similarity graph is constrained by spectral embedded representation and neighbor distribution instead of single distance information, which shows the effectiveness of the adaptive neighbors clustering stage.

C. Experiment on Real Data

In this section, we conduct our experiment to test our method with real-world datasets. Six datasets with four low-dimensional datasets and two high-dimensional datasets are used for evaluation. The COIL20 dataset [26] has 20 kinds of objects with the image size as 32×32. The USPS dataset is a handwritten recognition dataset [27] in which the image size is 16×16. The YEAST, WINE,ECOLI and GLASS datasets are part of UCI Machine Learning Repository datasets [28]. The detailed information of these datasets can be seen in Table III.

In this experiment, we use two kinds of evaluation metrics to verify the effectiveness of the proposed method, namely the clustering accuracy and the normalized mutual information (NMI) [29]. The clustering accuracy is defined as:

$$ACC = \frac{\#correct\ decisions}{\#total\ decisions}, \tag{38}$$

where $\#$ means "the number of". Denote C as the set of clusters from the ground truth and C' the output. Their mutual information MI is defined as:

$$MI = \sum_{c_i \in C, c_j \in C'} p(c_i, c_j) log \frac{p(c_i, c_j)}{p(c_i) p(c_j)}. \tag{39}$$

And the NMI is defined as:

$$NMI = \frac{MI}{\sqrt{H(C) H(C')}}, \tag{40}$$

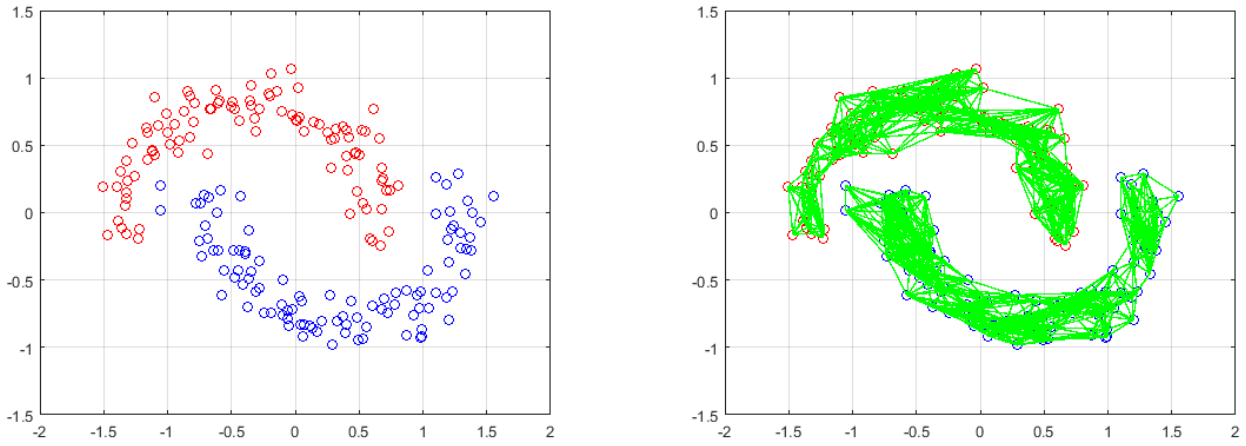
in which $H(\cdot)$ is the entropy of cluster result that is defined as:

$$H(C) = -\sum_{c_i \in C} p(c_i) log p(c_i). \tag{41}$$

Seven algorithms including K-means, Spectral Clustering with ratio cut [30], Spectral Clustering with normalized cut [31], Nonnegative Matrix Factorization [32] [33], PCAN [12], KSEC [16], and ESEC [16] are used for comparison. Due to the limited information, the comparison of KSEC and ESEC only contains the clustering accuracy metric. The missing values are demonstrated as '-'.

The experimental results about clustering accuracy are shown in Table I. The proposed SEANC framework outperforms other methods in five datasets. Compared with ratio cut spectral clustering algorithms, our method shows a better performance by 12.31% on average. The comparison results between normalized cut spectral clustering and the proposed method are similar. In the experiment conducted with ECOLI dataset, our method outperforms normalized cut spectral clustering by 32.32%. This demonstrates that the proposed method takes advantages of spectral embedded representation and shows better performance compared with spectral clustering.

The experimental results about clustering NMI are shown in Table II. Our method outperforms other algorithms on



(a) Original Data

(b) Learned Clustering Graph

Fig. 1. Experiment on Synthetic Data.

TABLE I
CLUSTERING ACCURACY

Dataset	COIL20	USPS	YEAST	WINE	GLASS	ECOLI
K-means	56.54±5.35	64.27±3.08	38.00±2.13	94.65±0.53	45.57±3.51	57.10±6.07
Spectral Clustering(R)	69.42±4.60	67.59±5.11	38.11±2.18	95.44±0.54	38.28±2.27	54.08±5.30
Spectral Clustering(N)	70.30±4.66	68.43±5.10	36.99±2.81	94.99±4.58	38.26±3.00	53.10±4.22
KSEC	72.40±2.70	65.10±1.60	-	61.20±2.30	58.90±2.40	-
ESEC	70.20±3.10	62.30±2.80	-	63.70±1.90	**63.10±2.50**	-
NMF	70.42	67.37	35.65	94.94	37.85	54.17
PCAN	83.33	63.81	50.07	**100.00**	49.53	83.33
SEANC	**85.97**	**71.36**	**51.15**	89.32	53.27	**85.42**

TABLE II
CLUSTERING NMI

Dataset	COIL20	USPS	YEAST	WINE	GLASS	ECOLI
K-means	73.45±2.61	62.07±1.64	25.19±1.06	82.41±1.44	33.13±2.60	53.04±3.02
Spectral Clustering(R)	84.01±1.94	73.95±2.41	24.94±1.37	84.37±1.25	29.10±3.05	48.96±3.11
Spectral Clustering(N)	84.42±1.93	73.97±2.40	23.88±1.35	84.02±3.90	28.58±2.49	49.78±2.32
NMF	81.22	74.15	23.66	83.24	28.70	47.12
PCAN	89.10	68.93	**30.55**	**100.00**	33.82	**72.44**
SEANC	**92.04**	**78.05**	29.21	68.71	**34.82**	70.47

TABLE III
DESCRIPTION OF FIVE DATASETS

Dataset	COIL20	USPS	YEAST	WINE	GLASS	ECOLI
sample	1440	1854	1484	178	219	336
class	20	10	8	3	6	8
dimension	1024	256	8	13	9	7

three datasets. On the YEAST and ECOLI datasets, the proposed method achieves the second highest performance. It is noticeable that our method shows promising results with high-dimensional data, namely COIL20 and USPS datasets, using both kinds of evaluation metrics. This phenomenon demonstrates that the proposed two-stage framework benefits from spectral embedded representation and adaptive neighbors clustering. With the help of these two steps, the SEANC method can process high-dimensional and sparse data.

D. Experiment on High-dimensional and Sparse Data

In this section, we use three text datasets to verify the effectiveness of our method when dealing with high-dimensional data. 3000 random selected articles from Enron Emails, NIPS full papers and KOS blog entries [34] which contain 35284 dimensions are used. Due to the property of writings, an article is hard to cover most part of vocabularies. Thus the text clustering data is an ideal kind of high-dimensional and sparse data to examine our method. All of the setups are the same as the previous section. The experimental results can be seen in Table IV.

As shown in Table IV, our method outperforms other methods when dealing with high-dimensional data. The PCAN method fails in this experiment due to the out of memory error.

TABLE IV
COMPARISON ON TEXT CLUSTERING

Meauserment	Accuracy	NMI
K-means	50.52±2.09	45.29 ± 0.19
Spectral Clustering(R)	47.98±7.70	17.89±9.39
Spectral Clustering(N)	61.20±0.00	22.53±0.00
PCAN	failed	failed
SEANC	**63.00**	**49.67**

This experiment shows the effectiveness of SEANC and the combination of linear embedded representation and adaptive neighbor clustering.

V. CONCLUSION

In this paper, we propose a two-stage framework called SEANC for clustering. The manifold assumption problem is addressed with linear spectral embedded process and adaptive neighbors clustering. The proposed method can not only handle the high-dimensional data with spectral embedded representation, but also learn the similarity graph and clustering results simultaneously. Experiments on both synthetic and real world data shows the promising performance of our method. In the future, applying the SEANC method in the large scale data can be a meaningful work.

REFERENCES

[1] F. Nie, X. Wang, M. I. Jordan, and H. Huang, "The constrained laplacian rank algorithm for graph-based clustering." in *AAAI*, 2016, pp. 1969–1976.

[2] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 11, pp. 2765–2781, 2013.

[3] A. Ben-Hur, D. Horn, H. T. Siegelmann, and V. Vapnik, "Support vector clustering," *Journal of machine learning research*, vol. 2, no. Dec, pp. 125–137, 2001.

[4] L. Xu, J. Neufeld, B. Larson, and D. Schuurmans, "Maximum margin clustering," in *Advances in neural information processing systems*, 2005, pp. 1537–1544.

[5] K. Zhang, I. W. Tsang, and J. T. Kwok, "Maximum margin clustering made practical," *IEEE Transactions on Neural Networks*, vol. 20, no. 4, pp. 583–596, 2009.

[6] G.-T. Zhou, T. Lan, A. Vahdat, and G. Mori, "Latent maximum margin clustering," in *Advances in Neural Information Processing Systems*, 2013, pp. 28–36.

[7] F. Nie, Z. Zeng, I. W. Tsang, D. Xu, and C. Zhang, "Spectral embedded clustering: A framework for in-sample and out-of-sample spectral clustering," *IEEE Transactions on Neural Networks*, vol. 22, no. 11, pp. 1796–1808, 2011.

[8] X. Chang, F. Nie, Z. Ma, Y. Yang, and X. Zhou, "A convex formulation for spectral shrunk clustering." in *AAAI*, 2015, pp. 2532–2538.

[9] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14. Oakland, CA, USA., 1967, pp. 281–297.

[10] I. S. Dhillon, Y. Guan, and B. Kulis, "Kernel k-means: spectral clustering and normalized cuts," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2004, pp. 551–556.

[11] J. Ye, Z. Zhao, and M. Wu, "Discriminative k-means for clustering," in *Advances in neural information processing systems*, 2008, pp. 1649–1656.

[12] F. Nie, X. Wang, and H. Huang, "Clustering and projected clustering with adaptive neighbors," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2014, pp. 977–986.

[13] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.

[14] S. X. Yu and J. Shi, "Multiclass spectral clustering," in *IEEE International Conference on Computer Vision*, 2003, pp. 313–319.

[15] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in neural information processing systems*, 2002, pp. 849–856.

[16] M. Liu, B. Liu, C. Zhang, and W. Sun, "Spectral nonlinearly embedded clustering algorithm," *Mathematical Problems in Engineering*, vol. 2016, 2016.

[17] J. Ye and T. Xiong, "Computational and theoretical analysis of null space and orthogonal linear discriminant analysis," vol. 7, no. 2, pp. 1183–1204, 2006.

[18] J. Ye, "Least squares linear discriminant analysis," in *Machine Learning, Proceedings of the Twenty-Fourth International Conference*, 2007, pp. 1087–1093.

[19] K. Fan, "On a theorem of weyl concerning eigenvalues of linear transformations i," *Proceedings of the National Academy of Sciences*, vol. 35, no. 11, pp. 652–655, 1949.

[20] J. Ye, Z. Zhao, and H. Liu, "Adaptive distance metric learning for clustering," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–7.

[21] T. Li, S. Ma, and M. Ogihara, "Document clustering via adaptive subspace iteration," in *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2004, pp. 218–225.

[22] F. R. Chung, *Spectral graph theory*. American Mathematical Soc., 1997, no. 92.

[23] B. Mohar, Y. Alavi, G. Chartrand, and O. Oellermann, "The laplacian spectrum of graphs," *Graph theory, combinatorics, and applications*, vol. 2, no. 871-898, p. 12, 1991.

[24] J. Ye, "Characterization of a family of algorithms for generalized discriminant analysis on undersampled problems," *Journal of Machine Learning Research*, vol. 6, no. Apr, pp. 483–502, 2005.

[25] L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering," in *Advances in neural information processing systems*, 2005, pp. 1601–1608.

[26] S. A. Nene, S. K. Nayar, H. Murase *et al.*, "Columbia object image library (coil-20)," 1996.

[27] J. J. Hull, "A database for handwritten text recognition research," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 16, no. 5, pp. 550–554, 1994.

[28] P. Franti, "Clustering datasets," 2015. [Online]. Available: http://cs.uef.fi/sipu/datasets/

[29] W. Chen and G. Feng, "Spectral clustering: A semi-supervised approach," *Neurocomputing*, vol. 77, no. 1, pp. 229 – 242, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231211005091

[30] L. Hagen and A. B. Kahng, "New spectral methods for ratio cut partitioning and clustering," *IEEE transactions on computer-aided design of integrated circuits and systems*, vol. 11, no. 9, pp. 1074–1085, 1992.

[31] 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.

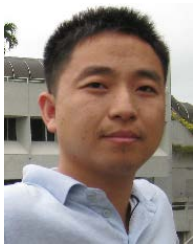
[32] F. Nie, C. Ding, D. Luo, and H. Huang, "Improved minmax cut graph clustering with nonnegative relaxation," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2010, pp. 451–466.

[33] T. Li and C. Ding, "The relationships among various nonnegative matrix factorization methods for clustering," in *Data Mining, 2006. ICDM'06. Sixth International Conference on*. IEEE, 2006, pp. 362–371.

[34] D. Dheeru and E. Karra Taniskidou, "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml

Qi Wang (M'15-SM'15) received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively. He is currently a Professor with the School of Computer Science, with the Unmanned System Research Institute, and with the Center for OPTical IMagery Analysis and Learning, Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision and pattern recognition.

Zequn Qin received the B.E. degree in computer science technology from Northwestern Polytechnical University, Xi'an, China, in 2016. He is currently pursuing the M.E. degree with the Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University, Xi'an, China. His current research intersts include computer vision and machine learning.



Feiping Nie received the Ph.D. degree in Computer Science from Tsinghua University, China in 2009, and currently is full professor in Northwestern Polytechnical University, China. His research interests are machine learning and its applications, such as pattern recognition, data mining, computer vision, image processing and information retrieval. He has published more than 100 papers in the following top journals and conferences: TPAMI, IJCV, TIP, TNNLS/TNN, TKDE, Bioinformatics, ICML, NIPS, KDD, IJCAI, AAAI, ICCV, CVPR, ACM MM. His papers have been cited more than 7000 times. He is now serving as Associate Editor or PC member for several prestigious journals and conferences in the related fields.

Xuelong Li (M'02-SM'07-F'12) is a full professor with the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, P. R. China and with the University of Chinese Academy of Sciences, Beijing 100049, P. R. China.