

A Novel Spectral Clustering Based on Nonlinear Low Dimensional Embedding Feature Selection

Daowen Zhang and Zhiping Zhou

*Engineering Research Center of Internet of Things Technology Applications
Ministry of Education, Wuxi 214122, China
18206181483@163.com, zzp@jiangnan.edu.cn*

Abstract

Spectral clustering is a clustering method based on algebraic graph theory. It has solid theoretical foundation and good performance of clustering. However, during the process of nonlinear low rank approximation, the traditional spectral clustering algorithm can't effectively remove redundant features leading to the phenomenon that the local area can not distinguish. It also suffers from the high computational complexity of eigen-decomposition when dealing with the high dimensional data. In order to resolve the aforementioned problems, in this paper a novel Spectral clustering algorithm called LF-SC is proposed. Firstly, based on the nonlinear low dimensional embedding feature selection, we realize dimension reduction. The multi clustering structure of the data is captured, the potential manifold structure is fully discovered, and the geometry structure of the low dimensional manifold clustering is well maintained. Secondly, utilizing the SVD instead of EVD to obtain the eigenvectors reduces the computational complexity and maintain the local structure of the data as well as low dimensional manifold. Extensive experiments show the effectiveness and efficiency of our approach.

Keywords: *nonlinear low dimensional embedding; feature selection; matrix decomposition; singular value decomposition; spectral clustering*

1. Introduction

Clustering or clustering analysis has been widely used in many fields, such as pattern recognition, data mining, image processing, and other fields of data analysis. During the past few decades, researchers have proposed a large number of clustering algorithms. Among them, spectral clustering algorithms have currently become one of the hotspot in machine learning. Spectral clustering has been utilized in dimension reduction, irregular data clustering, image segmentation, video analysis and so on which has attract much attention[1-2]. Spectral clustering reveals the intrinsic cluster structure of data, and can be used to detect the non convex and non linear structure with eigenvectors of the Laplace matrix. It is considered being one of the most promising clustering techniques due to its better performance compared with the traditional clustering algorithms. The research on spectral clustering mainly includes the following several main aspects: 1) the study of basic theory of spectral clustering, the most remarkable results were from international well-known research group. The theory of spectral clustering includes the role of eigenvectors of Laplace matrix in data classification, the relationship between spectral clustering and kernel PCA, manifold learning. Some work concentrated on the basic problems in the theory, that is the problem of learning feature function of integral operator[3]. 2) The study of spectral clustering algorithms. In this problem, studies pay more attention to the intrinsic quality of dimensionality reduction, the discovery of low rank structure, and the improvement as well as application of the algorithm. The spectral clustering can be divided into two phases: firstly, constructing the graph and getting the

Laplace matrix with adjacency matrix; secondly, calculating the eigenvectors of the Laplace matrix. The drawback of the spectral clustering can be seen from the two phases: the algorithms suffer from slowly running when dealing with large-scale data sets which is caused due to the increase of the computational complexity. The establishment of the graph and eigenvectors need to be time-consuming, and the spectral clustering algorithm becomes inefficient faced with large-scale data sets. In order to resolve the mentioned problems, many researchers have made a lot of improvements on the related spectral clustering algorithms. Cao *et al* [4] proposed an approximate spectral clustering algorithm based on local information of the data points to solve the above problems. A sparse graph is firstly constructed, then the generalization performance of the algorithm is improved with the local information of the data points. Experiments show the superiority of the proposed algorithm. Sparse representation has been demonstrated to be effective in dealing with high-dimensional data, Wu *et al* [5] proposed a spectral algorithm with two weight matrix based on sparse representation vector. Experiments of real high dimensional data show that the performance of the algorithm is higher than the existing spectral clustering algorithms. A kernel spectral clustering model was proposed for large complex networks in paper [6]. In the paper, the model is established with the sub graph of complex networks, and the clusters is automatically recognized through the projection structure of feature space. The algorithm can effectively realize the clustering analysis of complex large data sets, and the performance is greatly improved. The nonlinear low rank approximation of the spectral clustering algorithm is consistent with the current popular manifold embedding. To improve the performance of the algorithms, we need to make corresponding strategies according to the characteristics of the data sets. When different clusters is overlapped, the performance of spectral clustering algorithm will decrease. This is because the performance of spectral clustering algorithms depend on the establishment of undirected graph or adjacency graph, which is easy to be connected to the overlapping area. To solve this problem, a method called SMMC was proposed [7], In the model, it is assumed that the data is close to a smooth low dimensional manifold, where some of the data manifold is separated but has a cross, and then the local geometric information of the sampling data is merged to establish the appropriate adjacency matrix. Spectral clustering can also be applied to the large data network to detect the clustering center to determine the number of clusters. Mall *et al* [8] established a model with the sub graph which keeps the whole cluster structure by the kernel spectral clustering, which take full advantage of the extend sampling characteristics of the invisible nodes of the data. In addition, combined with the current academic research hot spots, many scholars have broadened the train of thought to improve the spectral clustering algorithms. For example, Wang *et al* [9] proposed spectral clustering algorithm based on multi object and multi view that resolved multi-objective optimization problem with Pareto optimal. The paper [10] proposed a new distance measure method based on double root tree. Combined with the spectral clustering, a new clustering algorithm was put forward which can divide the clusters of adjacent non convex shapes and calculate the local or global geometrical features of the data. 3) Application of spectral clustering. Spectral clustering is widely used in the field of image and video segmentation, where fruitful research results are get[11-12].

In summary, spectral clustering suffers from low efficiency in dealing with high dimensional data. The redundant attributes of the data points are not effectively removed leading to the phenomenon that the local area can not be distinguished, which affects the clustering performance. Combined with the research in recent years, A novel spectral clustering based on nonlinear low dimensional embedding feature selection (LF-SC) is proposed in this paper, which realizes the two dimension reduction of data and fully discover potential manifold structures and well keep the clustering geometrical structure. With the method, the performance of the spectral clustering is proved.

The rest of this paper is organized as follows. In section 2, the new algorithm is proposed. we firstly introduce the related theory of spectral clustering and the drawback of the algorithm. Section 3 summarizes our experimental results. We demonstrate the good performance of the algorithm and compare it with the other algorithms. In Section 5 we point out the innovation of the article and give a conclusion to this paper by some remarks.

2. Spectral Clustering based on Nonlinear Low Dimensional Embedding Feature Selection

In this section, we will introduce the improved spectral clustering algorithm in detail after introducing the related theory of spectral clustering . Firstly, we select features of the high dimensional data with the unnormalized graph Laplacian matrix aiming to fully discover potential manifold structures and well keep the clustering geometrical structure. Redundant attributes are removed to avoid the phenomenon that the local area can not be distinguished because of the same similarity measure. Secondly, we select the representative data points which are used to decompose matrix to obtain the representative data matrix of raw data, then utilize singular value decomposition to obtain the matrix made up of eigenvectors. At last, use the traditional K-means to cluster the matrix and obtain the clustering label.

2.1 Spectral Clustering

Spectral clustering is developed from the spectral theory [13] , which is divided into two stages for clustering. Given dataset $X=[x_1, x_2, \dots, x_n] \in \mathbb{R}^{M \times N}$ which has k clusters. In the first stage, the adjacency graph $G(V, E, W)$ that is an undirected graph is constructed, where $V=(v_1, v_2, \dots, v_n)$ is the collection of the vertices in the graph, one vertex represents a data point, E is the collection of the edges between vertices. $W=[w_{ij}]$ is the adjacent matrix calculated by Gauss similarity function: $w_{i,j}=\exp(-\|x_i - x_j\|^2 / (2\sigma^2))$. D is diagonal matrix calculated by $D_{ii} = \sum_j w_{ji}$. Then the unnormalized graph Laplacian matrix is defined as $L=D-W$. In the second stage, decompose Laplacian matrix and get the first k eigenvectors respectively corresponding to the first k minimum eigenvalues. The original M dimensional data set will be changed to a k dimensional data set. At last, use the traditional K-means to analyze the low dimensional representation of the raw data and obtain the clustering label of the data.

2.2 Nonlinear low Dimensional Embedding Feature Selection

The method of feature extraction can be divided into "packaging method" and "filtering method" [14], the "packaging" usually requires a large computational cost, which is not suitable for large scale data set. So, this paper is more inclined to the second method due to its high efficiency. Given dataset $X=[x_1, x_2, \dots, x_n] \in \mathbb{R}^{M \times N}$, obtain the subset of the features through specific method: $X'=[x'_1, x'_2, \dots, x'_n] \in \mathbb{R}^{d \times N}$ includes the most complete features and keep the geometric structure of the low dimensional manifold of original data.

Constructing an undirected graph with N vertices for the dataset X , every vertex represents one data point. For x_i , calculate its p nearest neighbor data points, then the weighted adjacent matrix W is defined as follows.

$$w_{ij} = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}, \quad i = 1, \dots, N; x_j \text{ belongs to the } p \text{ nearest data points of } x_i$$

(1)

W is a symmetric matrix, and every element of the degree matrix D is the sum of the rows or columns of the W . D is defined as $D_{ii} = \sum_j w_{ji}$. Laplacian matrix is defined as :

$$L = D - W \quad (2)$$

And according to formula (3), we can get the eigenvectors matrix $Y=[y_1, y_2, \dots, y_k]$ that is the collection of the eigenvectors corresponding to the k smallest eigenvalue. If the number of the cluster centers of the original data is known, then the k is the number of clusters. Each row of Y is a parallel embedding of each data point, and each column feature vector reflects the distribution of data on the correlation dimension.

$$Ly = \lambda Dy \quad (3)$$

The embedding matrix Y is obtained, which can measure the importance of each feature in the data's intrinsic dimension (each column of Y), correspondingly, determine the role in distinguishing different clusters. In the paper, \mathbf{b}_i is called sparse coefficient vector which is a M dimension vector. \mathbf{b}_i is calculated by minimizing the following expression:

$$\min_{\mathbf{b}_i} \|\mathbf{y}_i - X^T \mathbf{b}_i\|^2 + \beta |\mathbf{b}_i|, i = 1, 2, \dots, k; \quad (4)$$

Where $|\mathbf{b}_i|$ is L1 norm of a vector, due to the penalty nature of L1 norm, the value of some elements in $|\mathbf{b}_i|$ will shrunk to be 0. In this case, select the most relevant feature subsets, corresponding to the nonzero elements in $|\mathbf{b}_k|$. Equation (4) is essentially a regression problem. In order to solve regression problem above, Equation (4) equivalents to the following formula:

$$\begin{aligned} \min_{\mathbf{b}_i} \|\mathbf{y}_i - X^T \mathbf{b}_i\|^2 \\ s.t. \quad |\mathbf{b}_i| \leq \gamma \end{aligned} \quad (5)$$

In the literature [15], the authors proposed a minimum angle regression algorithm which can be used to solve the optimization problem. By setting parameters γ to determine the constraints of the problem, the minimum angle regression algorithm is used to control the sparsity of the vector by determining the number of nonzero elements d in $|\mathbf{b}_i|$ (in this paper, d is the number of features to be selected), d is called the cardinality of each vector. This approach is well suited for feature extraction in the condition of keeping data nonlinear manifold structures.

After the above calculation, if the number of cluster centers of the original data is k , the k sparse vectors with M dimensions are obtained, every nonzero element in the vector corresponds to a feature of the original dataset. The cardinality of each vector is d and each entry in $|\mathbf{b}_i|$ corresponds to a feature. If we select all the features that have at least one non-zero coefficient in k vectors, it is very possible that we will obtain more than d features. In reality, we can use the following simple yet effective method for selecting exactly d features from the k sparse coefficient vectors. For every feature j , we calculate $\mathbf{b}_{k,j}$ and find the maximum value in the k vectors. $\mathbf{b}_{k,j}$ is the j -th element of vector \mathbf{b}_k . We then sort all the features according to the maximum value in descending order and select the top d features. At this point, using the intrinsic dimensions of the data, we fully find the potential manifold structure, and maintain the geometric structure of the data, and complete the feature extraction and dimension reduction, converting data $X = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{M \times N}$ to $X' = [x'_1, x'_2, \dots, x'_n] \in \mathbb{R}^{d \times N}$.

2.3 Matrix Factorization

Through the calculation of the last section, we extend the feature selection based on the nonlinear low dimensional embedding, and expand the clustering region based on the correlation feature of data avoiding the phenomenon that some local areas can not be resolved. While the data features are reduced in the longitudinal dimension, the data is compressed, but the data is not reduced on the horizontal data points. When faced with large-scale data sets, it still needs large computation. In this section, a representative data matrix is formed, which is based on the random selection of data points to realize the matrix factorization. According to section 3.1, we obtain $\mathbf{X}' = [x'_1, x'_2, \dots, x'_n] \in \mathbb{R}^{d \times N}$, Inspired by sparse encoding and semi supervised learning [16], the \mathbf{X}' can be decomposed into U matrix and Z matrix. The approximate decomposition equation is defined as:

$$\mathbf{X}' \approx \mathbf{UZ} \quad (6)$$

Where $\mathbf{U} \in \mathbb{R}^{d \times l}$, $\mathbf{Z} \in \mathbb{R}^{l \times n}$, Z, U is obtained by solving the following minimum optimization problem.

$$\min_{\mathbf{U}, \mathbf{Z}} \|\mathbf{X}' - \mathbf{UZ}\|^2 \quad (7)$$

$\|\cdot\|$ is the Frobenius norm of the matrix which is difficult to get the optimal solution. The existing methods are mostly iterative method, which requires a lot of time consumption. Each column of the U matrix has the same dimension as the data, so we can randomly selected l data points from the original data \mathbf{X}' to constitute U matrix. Assuming that the U matrix has been chose, Z need to be calculated. If we has known U, formula (7) will be constrained linear regression problem which can be solved through the following methods.

For data points x'_i , the approximate values are calculated by the following formula.

$$\hat{x}'_i = \sum_{j=1}^l z_{ji} u_j \quad (8)$$

u_j is the j -th column vector of U, z_{ji} is the element on j -th row and i -th column. Set the parameters r , when u_j does not belong to r nearest data points of x'_i , z_{ji} is set to be 0.

$\mathbf{U}_{\langle i \rangle} \in \mathbb{R}^{d \times r}$ is a matrix made up of r nearest data points of x'_i . z_{ji} is calculated by the following formula

$$z_{ji} = \frac{K_h(x_i, u_j)}{\sum_{j \in U_{\langle i \rangle}} K_h(x_i, u_j)} \quad j \in U_{\langle i \rangle} \quad (9)$$

Where $K_h(x_i, u_j) = \exp(-\|x_i - u_j\|^2 / (2\sigma^2))$.

2.4 Obtaining the Eigenvectors by Singular Value Decomposition

$\mathbf{Z} \in \mathbb{R}^{l \times n}$ is calculated by matrix factorization, so it is easy to get the graph matrix W of data, $W = \hat{\mathbf{Z}}^T \hat{\mathbf{Z}}$, $\hat{\mathbf{Z}} = D_z^{-1/2} \mathbf{Z}$, D_z is a diagonal matrix whose elements are the sum of the row elements in \mathbf{Z} . The purpose is to get eigenvectors the of W , and then construct the new data set with eigenvectors, in the new data set matrix, each column is one of the eigenvectors. At last, cluster every row of data set matrix with k-means. In this paper, we do not use eigen-decomposition, because of its high complexity. Since we have obtained the matrix $\hat{\mathbf{Z}}$, we can consider getting the eigenvectors through some calculation of $\hat{\mathbf{Z}}$. The specific method is the singular value decomposition (SVD).

$$\hat{\mathbf{Z}} = \mathbf{A} \mathbf{\Sigma} \mathbf{B}^T \quad (10)$$

$\Sigma = \text{diag}(\sigma_1, \dots, \sigma_l)$, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_l \geq 0$ is singular value. $A \in \mathbb{R}^{l \times l}$ is left singular vector matrix, $B \in \mathbb{R}^{n \times l}$ is right singular vector matrix. As we can know from the knowledge of the singular value decomposition in matrix theory, each column of the B is the eigenvectors of W. B is calculated by the expression.

$$B^T = \Sigma^{-1} A^T \hat{Z} \quad (11)$$

B is obtained, Each row of the matrix represents a data point, and finally, the k-means method is used to cluster the data of each row of the matrix.

Complexity analysis: for the data set with n data points, the time complexity of the traditional spectral clustering algorithm for clustering analysis on two phases is : $O(n^2)$ and $O(n^3)$.and when the data size is large, the running time of the algorithm is very large. In this paper, the time complexity of spectral clustering algorithm after the feature extraction is: $O(lnd)$ and $O(l^3 + l^2n)$.Because l is far less than n , the algorithm has obviously reduced the complexity of computing.

At last, the flow of the proposed algorithm is described as follow:

Step 1: Input the original dataset and set the parameters including p, r, l , the number of clusters k , the number of features selected d .

Step 2: Construct adjacency matrix W , and then obtain the Laplacian matrix L .

Step 3: Calculate the k eigenvectors corresponding to the k smallest eigenvalue.

Step 4: Optimize the formula (5) to get the sparse coefficient vector .

Step5: Use the sparse coefficient vector to select the d features and get the new data matrix x' .

Step 6: Decompose x' and obtain matrix Z .

Step 7: Decompose Z with Singular value decomposition to get the right singular vector matrix B .

Step 8: Cluster matrix B with K-means.

3. Experiment and Analysis

In this section, we compare the performance of the proposed clustering algorithm and other clustering algorithms to verify the feasibility and effectiveness of the proposed algorithm for the high dimensional large scale data sets. The Algorithm is implemented with matlab which is running on Lenovo computer whose ram, frequency and hard disk are respectively 4G, 2.20GHZ and 500GB. The algorithms compared are K-means, Nystrom [17], LaplacianScore[18], Among them, Nystrom and the K-means algorithm have no process of feature extraction, they uses all the features. LaplacianScore is a clustering algorithm based on feature extraction, which maintains the local manifold structure of the data. This algorithm is mainly used to compare the performance of feature selection. Test data sets are USPS, UMIST[19], COIL20, and Isolet[5], the data sets are described as shown in Table 1.

Table 2. Dataset Description

dataset	size	dimension	class
COIL20	1440	1024	10
USPS	9298	256	20
Isolet	1560	617	26
UMIST	1799	1024	12

In this paper, two evaluation metrics are used to evaluate the clustering performance of the various algorithms, they are the accuracy rate (AC), and the normalized mutual information value (NMI) [19]. Accuracy rate is a comparison between the clustering labels obtained by the algorithms and the original label of the data. It can be considered as a random variable in the amount of information that is contained in another random variable. The value of AC and NMI is a numerical value between 0-1, the greater the value, the better the clustering performance of the algorithm. In the proposed algorithm, there are three parameters: p , r , l , and the values are set to be 5, 6, 500.

In order to compare the clustering performance of the algorithm, the four algorithms are running on the four data sets, and the AC and NMI of different algorithms are obtained, which are shown in Figure 1. The graph shows the changing trend of the different algorithms' performance with the change of the number of features selected. The two straight lines in the graph show the clustering result of the algorithm Nystrom and K-means, which have no feature extraction process but use the whole features. So the clustering performance is not changed with the change of the number of features.

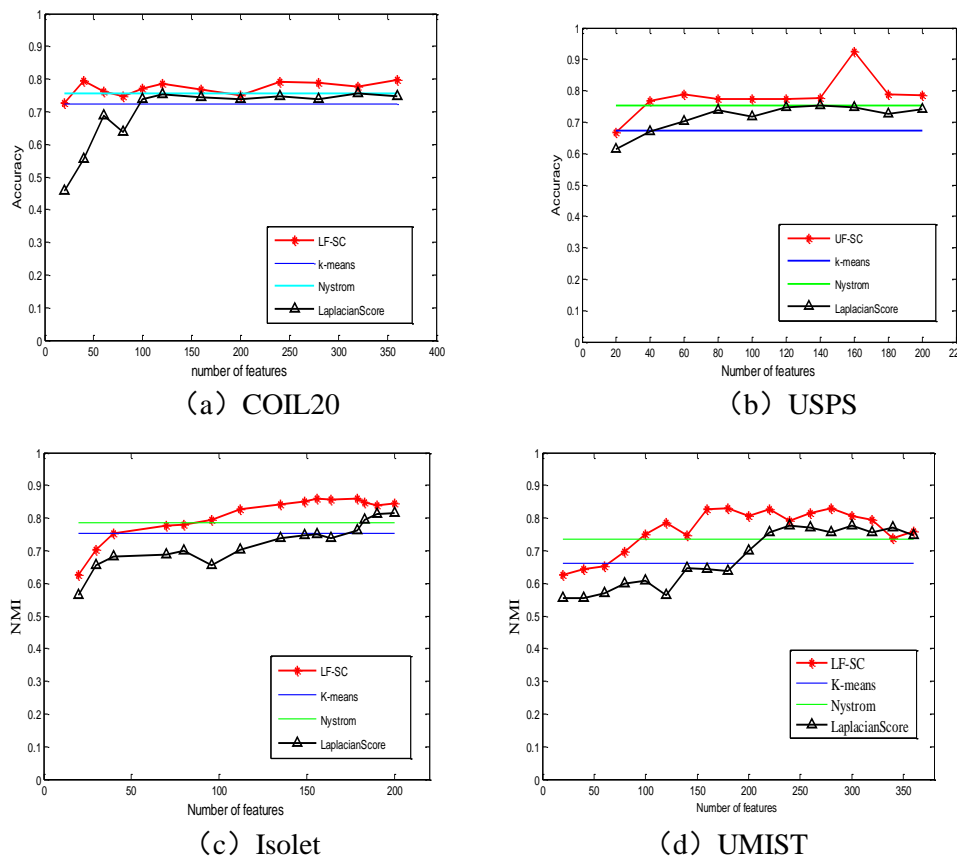


Figure 1. The Comparison of AC and NMI Get by Different Algorithms on four Data Sets

In figure (a) and (d), the number of features selected is ranging from 20 to 360, In figure (b) and (c), the number of features selected is ranging from 20 to 200, and the accuracy and normalized mutual information values of the clustering results are obtained under the different number of features. In (a) and (b) AC is compared and NMI is compared in (c) and (d). From the graph, it can be seen that the overall clustering result of LF-SC algorithm is better than other algorithms can achieve better clustering effect with less features than LaplacianScore algorithm. In other words, the algorithm has better convergence. In this paper, the algorithm is based on the feature extraction of nonlinear low dimensional embedding, and it is easier to capture the multi clusters structure of the

data. The algorithm can capture the data points which is closely related to the Clustering center and fully discover the potential of the popular structure to maintain the geometrical structure. of the data. Nystrom is a classical spectral clustering algorithm, performance has been greatly improved compared to the traditional mean clustering algorithm. However, the algorithm uses the whole features and the redundancy attribute is not removed by combining the intrinsic structure of the data, The same similarity measure caused phenomenon that the local area can not be distinguished. LF-SC algorithm implements the dimension reduction operation based on the nonlinear low dimensional embedding, Therefore, the clustering performance of LF-SC algorithm in most case of the feature extraction is higher than that of the Nystrom algorithm.

In order to verify the clustering performance of the algorithm in different number of clusters(COIL20:5,10,15,20 ; USPS:3,5,7,10;Isolet:10,15,20,26;UMIST:3,6,9,12). Only 100 features are selected in LaplacianScore and LF-SC algorithm and the NMI is obtained as shown in Figure 2. It can be seen that the performance of LF-SC is better than LaplacianScore under different number of clusters on four data sets.

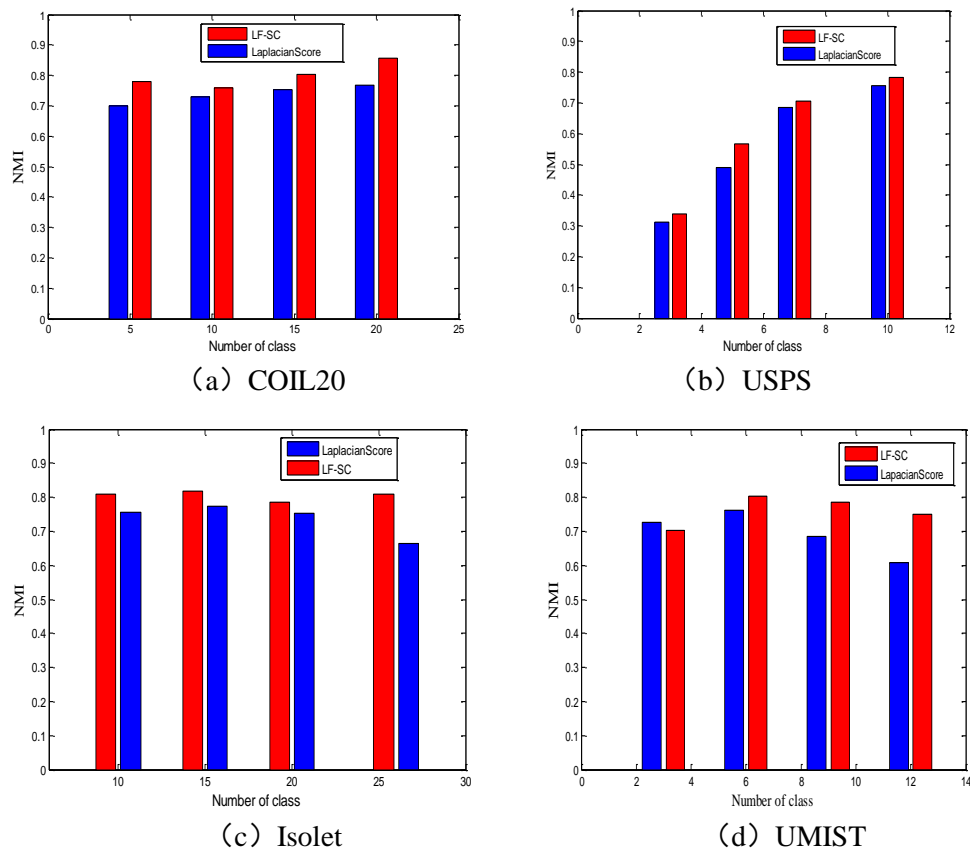


Figure 2. The Comparison of NMI Value under Different Class with 100 Features

The following table shows the total time cost of LF-SC algorithm and Nystrom algorithm in the phases of graph building and eigen-decomposition. In LF-SC algorithm, time cost does not include the time in the process of feature extraction. For the rationality, all features are adopted.

Table 2. Comparison of the Running Time

datasets	algorithm	time(s)
COIL20	LF-SC	7.215
	Nystrom	12.018
USPS	LF-SC	21.657
	Nystrom	39.856
Isolet	LF-SC	9.653
	Nystrom	15.511
UMIST	LF-SC	11.951
	Nystrom	16.256

The data in the table show that LF-SC algorithm effectively reduce the time cost of matrix decomposition and the singular value decomposition of the data points compared with the classical clustering algorithm. LF-SC algorithm has large computational complexity in phase of feature extraction, if we want to combine feature extraction with spectral clustering, it is necessary to reduce the computational complexity in the last phase of the algorithm and reduce the time cost, which is well guaranteed by the LF-SC algorithm.

4. Conclusions

The traditional spectral clustering algorithms can not effectively remove the redundant attribute of data points in the process of low rank approximation which cause the problems that local area can not be distinguished. In order to resolve the mentioned problems, in this paper, a algorithm called LF-SC is proposed which is based on nonlinear low dimensional embedding feature selection. The algorithm can capture the multi clusters structure of the data with Laplacian matrix and fully discover the potential manifold structure and well maintain the geometrical structure. In addition, the data matrix is decomposed with the selected data points and the eigenvectors is obtained by the singular value decomposition which reduce the complexity and maintain the local structure of the data as well as low dimensional manifold. The experimental results verify the feasibility and effectiveness of the proposed algorithm compared with other algorithms.

Acknowledgement

This work is supported by “the Fundamental Research Funds for the Central Universities” (No.JUSRP51510), Natural Science Foundation of Jiangsu Province(No. BK20131107) and National Natural Science Foundation of China(No. 61373126).

References

- [1] Tasdemir K. Vector quantization based approximate spectral clustering of large datasets[J]. Pattern Recognition, 2012, 45(8): 3034-3044.
- [2] Wang L, Dong M. Multi-level Low-rank Approximation-based Spectral Clustering for image segmentation[J]. Pattern Recognition Letters, 2012, 33(16): 2206-2215.
- [3] Rosasco L, Belkin M, Vito E D. On learning with integral operators[J]. The Journal of Machine Learning Research, 2010, 11: 905-934.
- [4] Cao J, Chen P, Dai Q, *et al.* Local information-based fast approximate spectral clustering[J]. Pattern Recognition Letters, 2014, 38: 63-69.
- [5] Wu S, Feng X, Zhou W. Spectral clustering of high-dimensional data exploiting sparse representation vectors[J]. Neurocomputing, 2014, 135: 229-239.
- [6] Mall R, Langone R, Suykens J A K. Self-tuned kernel spectral clustering for large scale networks[C]. Big Data, 2013 IEEE International Conference on. IEEE, 2013: 385-393.
- [7] Wang Y, Jiang Y, Wu Y, *et al.* Spectral clustering on multiple manifolds[J].IEEE Transactions on Neural Networks, 2011, 22(7): 1149-1161.

- [8] Mall R, Langone R, Suykens J A K. Kernel spectral clustering for big data networks[J]. *Entropy*, 2013, 15(5): 1567-1586.
- [9] Wang X, Qian B, Ye J, *et al.* Multi-objective multi-view spectral clustering via pareto optimization[C]. *Proc. SIAM Int. Conf. Data Min.* 2013.
- [10] Galluccio L, Michel O, Comon P, *et al.* Clustering with a new distance measure based on a dual-rooted tree[J]. *Information Sciences*, 2013, 251: 96-113.
- [11] Galasso F, Keuper M, Brox T, *et al.* Spectral graph reduction for efficient image and streaming video segmentation[C]. *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on.* IEEE, 2014: 49-56.
- [12] Wang C, Xu L Z, Wang X, *et al.* A Multi-scale Segmentation Method of Oil Spills in SAR Images Based on JSEG and Spectral Clustering[J]. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 2014, 7(1): 425-432.
- [13] Chung F R K. *Spectral graph theory*[M]. American Mathematical Soc., 1997.
- [14] Liu H, Yu L. Toward integrating feature selection algorithms for classification and clustering[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2005, 17(4): 491-502.
- [15] Efron B, Hastie T, Johnstone I, *et al.* Least angle regression[J]. *The Annals of statistics*, 2004, 32(2): 407-499.
- [16] Liu W, He J, Chang S F. Large graph construction for scalable semi-supervised learning[C]. *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. 2010: 679-686.
- [17] Chen W Y, Song Y, Bai H, *et al.* Parallel spectral clustering in distributed systems[J]. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2011, 33(3): 568-586.
- [18] He X, Cai D, Niyogi P. Laplacian score for feature selection[C]. *Advances in neural information processing systems*. 2005: 507-514.
- [19] Chang X, Nie F, Ma Z, *et al.* A convex formulation for spectral shrunk clustering[J]. preprint arXiv:1411.6308, 2014.

Authors

Daowen Zhang, Master Degree Candidate; Institution: Engineering Research Center of Internet of Things Technology Applications Ministry of Education, jiangnan university; Address: Binhu District Road No.1800, Wuxi Jiangsu, China; Tel: 18206181483; Email: 18206181483@163.com; Title: A Novel Spectral Clustering Based on Nonlinear Low Dimensional Embedding Feature Selection ; Subject: Data Mining, Artificial Intelligence.

Zhiping Zhou, Professor; Institution: Engineering Research Center of Internet of Things Technology Applications Ministry of Education, jiangnan university; Address: Binhu District Road No.1800, Wuxi Jiangsu, China; Tel: 13861867502; Email: zzp@jiangnan.edu.cn; Title: A Novel Spectral Clustering Based on Nonlinear Low Dimensional Embedding Feature Selection ; Subject: data mining, computing intelligence.