International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol. 8, No. 5 (2015), pp. 93-104
http://dx.doi.org/10.14257/ijsip.2015.8.5.10

Eigenvectors Selection in Spectral Clustering by Applying Multi-Objective Genetic Algorithm

Rasoul Taghipour[*], Ehsan Asgarian[**] and Niloofar Chehrenama[***]

[*] *Department of Computer Engineering, Neyshabur Branch, Islamic Azad University, Neyshabur, Iran*
[**] *Department of Computer Engineering, Ferdowsi University of Mashhad, Iran.*
[***] *Department of Computer Engineering, Payam-Noor University of Mashhad, Iran*
asgarian@alum.sharif.edu

Abstract

In recent years, several researches have been conducted on spectral clustering to classify non-linear data in various applications. Considering the effect of selecting the appropriate eigenvectors on spectral clustering performance; various methods have been proposed weighting and ranking features. However, these methods can independently evaluate the impact of each eigenvector. Nevertheless, it is possible that several eigenvectors have duplicate or inadequate information on some clusters. Thus, we have presented a new method for finding the optimal combination of eigenvectors by several different evaluation criteria. In order to detect simultaneously the optimum condition in various criteria, the multi-objective genetic algorithm is applied. Findings of performed experiments on datasets with various features demonstrate a resounding success in the proposed method.

Keywords: *Spectral clustering, Multi-objective genetic algorithm, Feature selection, Clusters evaluation, Pareto Front*

1. Introduction

Grouping data into a set of categories or clusters of similar properties is of crucial importance in controlling and data management methods. In order to learn and understand a new phenomenon, human beings always search for characteristics and features so that they can describe it and compare it with other known phenomena afterwards. This comparison is based on overall similarity criteria and in accordance with specific standards and rules. In general, learning systems are accompanied by supervised or unsupervised method. Clustering is one of the most important methods of unsupervised learning. The aim of clustering is to put unlabeled data into groups, so that data similarities within each cluster are maximized and data similarities within different clusters are minimized. Clustering process leads to reduction of information size; because information about several homogeneous groups is kept instead of keeping information about a large number of objects. Many clustering methods have been proposed in the last few years [1-3]. In addition to the similarity measurement and delimited data methods, the quality of clustering method results is highly dependent on the number of data dimensions and data distribution in space.

One of the major challenges for data clustering problem is the presence of outlier and noisy data in data sets. Moreover, once the data distribution is non-linear and non-convex, classic clustering methods are not able to partition them correctly. In other hand, with the rapid growth of databases and the increase in their dimensions, the accuracy of classic clustering methods is considerably decreased. Thus, subspace learning algorithms have

ISSN: 2005-4254 IJSIP

been proposed that try to map the original high-dimensional data to low dimensional space so that statistical and initial properties of the original data are preserved.

One of the fields of subspace learning method is the spectrum analysis [4, 5], which in recent years has been considered by many researchers. Spectral clustering is an appropriate solution to data clustering in non-linear separable space. Spectral analysis attempts to partition the graph data. In order to model the data by a graph, each data is taken as a given node [6]. Three conditions have been stated to draw the edges [7]: 1) Graph is considered complete2) Edges are simply drawn between data that bear a striking resemblance to each other. 3) Each vertex is connected to several neighbors -The closest ones to the data corresponding to the vertex. Weight of edges will be considered according to the similarity of their corresponding data nodes. Various methods have been proposed for graph clustering [8, 9] which attempt to partition the given graph nodes, so that nodes in each cluster are interconnected strongly and nodes among the clusters have a weak relationship with each other.

Eigenvectors of the adjacency matrix of the graph, which are obtained from its Eigen decomposition, include valuable information about partitioning the vertices of a graph – dataset – [10] and it can also be used for clustering and reducing the dimension of data. Furthermore, Guthrie [10] showed that spectral graph partitioning is equal to its graphical decomposition based on eigenvectors and eigenvalues of the Laplacian matrix of the graph. Calculating the distance matrix of data in spectral clustering and using eigenvectors of its different Laplacian matrixes, it is being attempted to select and apply eigenvectors possessing vast knowledge about data partition by providing various algorithms. For this purpose, eigenvectors of Laplacian matrix are sorted based on their eigenvalues in descending order and a limited number of the sorted list is selected to represent data.

Various methods of spectral clustering have practically been employed in parallel computing [11]. These methods include the VLSI design [12], image, speech, video and texts processing [13-18], bioinformatics [19-22], *etc.*, [23, 24]. One of the major challenges in spectral clustering is the selection method of the appropriate vectors containing adequate information about all the clusters. NJW[1] algorithm [25] is one of the most widely used spectral clustering methods. In this approach, for the clustering problem, data in k clusters choose k eigenvectors corresponding to the largest eigenvalues (adjacent normalized matrix). However, in practice, when data noise is high or clusters are close to each other, these k eigenvectors are not able to recognize data structures well. Furthermore, the selection of inappropriate eigenvectors containing little information results in a reduction in the accuracy of clustering [26]. In 2008, Jiang [26] proposed a method for selecting better eigenvectors for the first time. In recent years, many attempts have been made to weight, rank and select appropriate eigenvectors that contain more information [26-29]. In this paper, a method based on selecting a combination of eigenvectors that lead to the best clustering has been presented. A set of appropriate eigenvectors are first introduced as a sample in the proposed method. These vectors can be selected based on a variety of previous methods. Therefore, Multi-objective Genetic algorithm has been employed so as to investigate various combinations of eigenvectors and evaluate them by different criteria that determine the quality of clustering. Thus, the effect of different combinations of eigenvectors are compared and evaluated with each other in the quality of clustering in order to obtain the best possible eigenvectors in homogeneous clusters to separate data. In the following parts, first, we review the previous work done in the field of spectral clustering and in the second Section the related problems are reviewed. The proposed method is provided in Section 3. Results of the proposed method are shown in Section 4 and

[1] Ng, Jordan and Weiss

they will be compared with other related methods. Finally, the conclusion has been presented in Section 5.

2. Spectral Clustering

In this Section, the general spectral clustering method is firstly presented and NJW method is depicted as one of the most successful (effective) methods of spectral clustering afterwards. Some explanations are provided concerning the importance of selecting the right eigenvectors in the following Section.

Spectral clustering methods have been defined on the basis of spectral graph theory [30]. These methods attempt to partition data graph by extracting the features from eigenvectors of adjacency matrix of graph. Laplacian matrix which is gained from normalization of adjacency matrix of graph has been applied in most spectral clustering methods. Afterwards, the eigenvectors obtained from the Laplacian matrix can be used to discover the existing clusters within the data. Various methods for spectral clustering have been proposed based on the normalization method and graph cuts [6, 9, 14, 25, 31-35], in which NJW spectral clustering algorithm is the most widely used [36].

2.1. Ng-Jordan-Weiss (NJW) Method and its Improvement

Considering the data set X= $\{x_1, x_2, \ldots, x_n\}$ in Υ^d space with K clusters, the similarity matrix $A^{n\times n}$ can be obtained as follows:

$$A_{ij} = \begin{cases} \exp\left(-d(x_i, x_j)/\sigma^2\right) & i \neq j \\ 0 & i = j \end{cases} \quad (1)$$

Where $d(x_i,x_j)$ is the distance between X_i and X_j data and the value of σ parameter determines the radius of adjacency pairs in the same clusters.

A_{ij} is the element of *i-th* row and *j-th* column of the similarity symmetric matrix that is equal to the weight of the edge between *i-th* and *j*-th data in the undirected graph data.

Thus, *NJW* method selects the first *k*(the number of clusters set by the user) eigenvector of the above matrix in a normalized formas the main features of data for the optimal partitioning graph data. *NJW* method is detailed as below [25, 28]:

1- Similarity matrix $A \in \Upsilon^{n\times n}$ calculated using equation (1)

2- Calculating the diagonal matrix $D(D_{ii} = \sum_{j=1}^{n} A_{ij})$ and laplacian matrix $L=D^{-1/2}A D^{-1/2}$, which is the normalized form of the similarity matrix.

3- Computing the large K eigenvalues ($\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$) of matrix L, with $v_1, v_2, \ldots, v_k$ eigenvectors equivalent to them and forming the column matrix $V=[v_1,v_2,\ldots,v_k] \in \Upsilon^{n\times k}$.

4- Forming matrix Y by re- normalization of V as:

$$Y_{ij} = V_{ij} \Big/ \left(\sum_{j=1}^{k} V_{ij}^2\right)^{1/2} \quad (2)$$

5- Each row of matrix *Y* in the space Υ^k is considered as new features of the data corresponding to it in the data set *X*. Then mapped data in the new space are placed in *k* clusters by K-Means algorithm.

By mapping data onto the new space, the spectral clustering method of *NJW*, unlike classical clustering methods such as K-Means, will be able to find non-convex clusters. Due to the importance of the similarity matrix *A* in converting space, it is essential that a proper value is considered for the parameterσby the user in the equation (1); since parameter σ directly controls the degree of similarity data. In various data sets regarding the distribution of data in each cluster, determining a constant value for this parameter can have undesirable effects on spectral clustering results [37, 38]. For this reason, the value of σ_i for each data point X_i is computed locally in [37, 38].The value of σ_i has been

considered equal to the distance between X_i and its *t*-th adjacent neighbor. Thus, the similarity of two different data in a similarity matrix is calculated as follows:

$$A_{ij} = \exp\left(-d(x_i, x_j)/\sigma_i \sigma_j\right), \ i \neq j \quad (3)$$

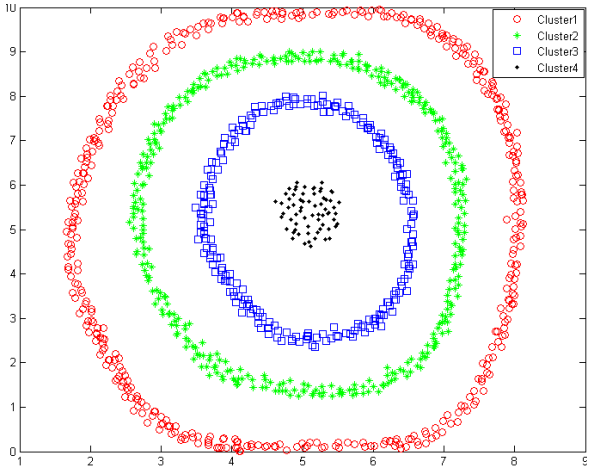
The optimal value for *t* depends on the number of data and clusters. However, in this paper and most similar researches, it is considered as a constant 7.

2.2. Previous Works on Selecting Eigenvectors

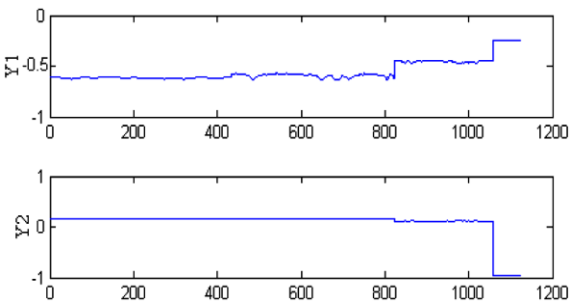
For the first time the importance of selecting appropriate eigenvectors in the quality of clustering results was stated by Xiang [26]. Jiang represented that, contrary to expectations, the accuracy of clustering results decreased by selecting more eigenvectors and the irregular increase in data size. In other words, in spectral clustering, selection of appropriate features not only does not lead to reduction of clustering accuracy, but also the elimination of eigenvectors which contain little information about clusters causes the intra-cluster distances and the convergence within clusters to increase and clusters will be separated easily.

For example, in Figure 1 (a), a simulated non-convex data set with four clusters has been shown in two dimensions. Due to the non-linear structure of the data set, we need to map data into feature space so as to decompose the existing clusters. In Figure 1 (b), the first four eigenvectors of Laplacian matrix of the data set- proportional to the largest eigenvalue- have been shown in feature space. According to the Figure 1 (b), to decompose data in feature space, the third and fourth eigenvectors perform much better than the first and second ones. In the proposed method of Xiang [26], regardless of the data distribution in the input space, the relevance of each eigenvector is determined based on its ability to separate data within clusters. Afterwards, suitable eigenvectors participate in the clustering according to a preference-weighted measure.

(a)



(b)



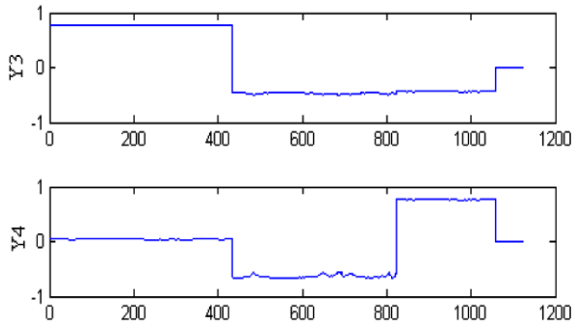


Figure 1. The Importance of Eigenvectors Selection in Separation of Four Clusters in Non-convex Data

Shi [27] in 2009 selected the appropriate vectors, based on data distribution and variation of values for each eigenvector. Shi's fundamental principle [27] was that there are eigenvectors for the data on each cluster that describe them as a large amount and data on other clusters are represented as values approaching zero. In another methodology in 2010, Zhao [28] computed the power of data separation in eigenvectors by the entropy to prove the importance of each feature. Another approach was proposed by Ashkezari in 2011 [38], in which the data are mapped into a nonlinear space by KPCA to extract the nonlinear features. Afterwards, clustering is separately performed based on each obtained eigenvector and the quality of obtained clusters is evaluated. If the quality of clustering is higher than a certain threshold, weight proportional to its fitness clustering result is given to the desired eigenvector. Finally, clustering is performed on specific weighted eigenvectors. In most of the work done for weighting and selection of appropriate eigenvectors, the attempt to determine the significance of each eigenvector is made separately [26, 27, 29, 38].

Zhao [28] in 2010 indicated that the combination of eigenvectors, which efficacy of which for separating data within clusters has been examined distinctly, does not necessarily represent an accurate presentation of the entire data structure. When the clusters are unbalanced or possess different forms and distributions, these eigenvectors contain inadequate or duplicate information about the data or some clusters [28, 38]. It causes these vectors to describe some clusters well; while they have little information about separating other clusters [27].

2.3. SCWES Method

In data-driven method *SCWES*[2] [26], regardless of data distribution in the input space, relevance of each eigenvector is determined based on its ability to separate data from clusters. In this method, it is assumed that the values of the *k-th* large eigenvector (e_k), e_{kn} can be in a unimodal or multimodal distribution based on the relevance of that eigenvector ($0 \leq R_{e_k} \leq 1$). Thus, to calculate the probability density function (PDF) of eigenvector values e_{kn} , combining two components, is as follows:

$$p(e_{kn} \mid \theta_{e_{kn}}) = (1 - R_{e_k}) p(e_{kn} \mid \theta^1_{e_{kn}}) + R_{e_k} p(e_{kn} \mid \theta^2_{e_{kn}}) \qquad (4)$$

Where $\theta_{e_{kn}}$ is the distribution parameter and $p(e_{kn} \mid \theta^1_{e_{kn}})$ is the probability density function (pdf), e_{kn} when e_k is an irrelevant or redundant vector and $p(e_{kn} \mid \theta^2_{e_{kn}})$ is probability density function (pdf), e_{kn} when e_k is a relevant vector. Moreover, R_{e_k} works as weight or probability of the mixture of the second component of the above equation.

[2] Spectral Clustering with Eigenvector Selection

In *SCWES* method, the distribution of the *k-th* eigenvector values (e_{kn}) was considered to be as single Guassian (unimodal) where the vector is irrelevant (data not used for clustering) and if the vector is relevant (*i.e.*, values of the vector can separate data in a cluster from other clusters), it was perceived as two Guassians (multimodal). *SCWES* method employs the EM iterative method (with the initial value for $R_{e_k} = 0.5$) for local optimization of the parameters. Then, eigenvectors which their relevance is more than a certain threshold, will participate in clustering according to their weighted relevance.

In Jiang's proposed method for separating the components of an image, good results were obtained. Although each vector selected in this method is solely informative for separating clusters, the combination of these vectors may lead to inadequate or duplicate information about the overall data structure.

2.4. ESBER Method

Zhao firstly defined the appropriate feature vectors using entropy in *ESBER*[3] approach [28]. Afterwards, all presence permutations of eigenvectors as data features in clustering are searched and the optimum eigenvectors are finally selected as new data features. Due to the lack of access to the probability of points in the real world to calculate the entropy, the similarity between the eigenvectors is applied as follows:

$$E_{\{V\}} = -\sum_{V_i \in V} \sum_{V_j \in V} \left[S_{ij} \log S_{ij} + (1-S_{ij}) \log(1-S_{ij}) \right] \quad (5)$$

Where S_{ij} is the similarity of two eigenvectors V_i and V_j. The similarity of two eigenvectors is calculated as:

$$S_{ij} = \exp\left(-\left[\sum_{k=1}^{n} \left(\frac{V_{ik} - V_{jk}}{\max_k - \min_k} \right)^2 \right]^{1/2} \right) \quad (6)$$

Where $\min_k$ and $\max_k$ are the minimum and maximum values in the eigenvector K (column k of the matrix V). To determine the ranking and importance of the eigenvector V_i, the rate of entropy is calculated after removing that eigenvector ($E_{\{V/V_i\}}$). The higher the entropy value is, the greater the importance of the feature V_i is to distinguish the data clusters. Since the high amount of entropy indicates the increase in similarity between data, clusters are close together or it leads to irregularity of data. Thus, an eigenvector contains additional information about clustering in which removing the feature causes more disruption in data. Finally, the number of K_m eigenvectors with the highest rank of entropy of the V matrix is selected as candidate features of data. Afterwards, using a small training dataset which has actual labels, the suitability for each combination of eigenvectors, which are eliminated from K_m superior features, is calculated by the following equation:

$$f(ec_i) = 10^2 Acc\_train + \left(\frac{len\_remain}{K_m} \right) \quad (7$$

Where ec_i ($1 \le i \le 2^{Km}-1$) is a possible combination of the selected eigenvectors for clustering. *Acc_train* is the clustering accuracy of the training data sets (the ratio of the data whose labels are predicted accurately to the total number of data) and *len_remain* is the number of eigenvectors which were not selected in the ec_i set *(len_remain=*K_m-len(ec_i)). Steps of *ESBER* method to clustering of n data in k cluster is as follows:

1. Construction of similarity matrix using equation (3) and formation of the Laplacian matrix L (similar to *NJW* method)

[3] Eigenvector Selection Based on Entropy Ranking

2. Calculating the eigenvectors of the matrix L, defining the entropy of features and formation of the diagonal matrix $V=[v_1,v_2,...,v_n] \in \Upsilon^{n\times n}$, where $E_{\{V/V_1\}} \geq E_{\{V/V_2\}} \geq ... \geq E_{\{V/V_n\}}$.

3. If the strategy of direct ranking is concerned, with the selection of the first K features of the matrix V as the matrix U, go to step 7, otherwise go to step 4.

4. For the data of a training set, the first K_m features (where K_m=max(k,*10*)) of the matrix $V_{train} \in \Upsilon^{tn\times Km}$ are selected (tn is the number of the training data).

5. For each possible combination of eigenvectors, *eci* (some V_{train} columns), data matrix of $U_{train} \in \Upsilon^{tn\times L}$, where $L=len(ec_i)$ is the number of selected features, is formed.

5.1 With the re-normalization of each row of $U_{train,}$ Y_{train} matrix is formed using equation (2).

5.2 Each row of U_{train} is given as an entry and the clustering of *K-Harmonic Means* is performed on these entries.

5.3 By calculating the accuracy of clustering for data of the training set in the new space, the efficiency of each ($f(ec_i)$) case is determined using equation (7).

6. According to the suitability function in equation (7), the best combination of ec^* eigenvectors is selected (Where $ec^* = Arg \max_{ec_i, 1\leq i\leq m} f(ec_i)$) and therefore, the matrix $U \in \Upsilon^{n\times L^*}$($L^*=len(ec_i^*)$)is formed.

7. Forming matrix Y with re-normalization of each row of U using equation (2).

8. Each row of matrix Y is considered as new data. Therefore, data mapped into the new space are classified in K clusters by the *K-Harmonic Means* algorithm.

Given K_m>10, in step 5, the genetic algorithm is applied to find the optimal combination of features (ec_i^*) instead of examining all possible combinations.

3. The Proposed Method

In the proposed method, a number of proper eigenvectors are selected to map data into a new space by various methods as a set of candidate features. Afterwards, using several different evaluation methods, attempts are made to select the best combination of eigenvectors by a multi-objective genetic algorithm. So that, in the new space, clustering algorithm can separate data into different clusters in the best state. Thee valuation measures and the search procedure to find the optimal combination of eigenvectors are the fundamental differences between the method proposed in this paper and the *ESBER* method. Two measures of data convergence and connectivity in each cluster are considered to estimate the quality of clusters. Convergence of clustering results by the total inner-cluster distance (*IND*) is calculated as follows [39]:

$$IND(C) = \sum_{j=1}^{k} \sum_{x_i \in C_j} d(x_i - m_j) \quad (8$$

Where C_j is the clustering datasets j and m_j is the cluster center j. In order to examine the data connectivity in each cluster, the neighborhood concept of each data is defined as follows:

Definition 1- Existence of x_i in the neighborhood of (close to) data t is given by:

$$nn(x_i,t) = x_j \ : \underset{x_k \in X-\{x_i\}}{Count} \left(d(x_i,x_k) < d(x_i,x_j)\right) = t-1 \quad (9$$

To calculate the benchmark data connectivity of each cluster, instead of each pair of data neighboring each other which are not in a cluster, some penalties in accordance with the following equation is given by [40]:

$$\text{Conn(C)} = \sum_{i=1}^{n}\sum_{t=1}^{L} P\left(x_i, nn(x_i,t)\right) \quad where$$
$$P\left(x_i, x_j\right) = \begin{cases} 1/t & if \;\; s \neq r \wedge x_i \in C_r \wedge x_j \in C_s \\ 0 & otherwise \end{cases} \qquad (10)$$

Where L is the maximum number of given neighbors and the function P is a penalty function for the two adjacent data which are not in a cluster[4].

In the proposed method for ranking and selection of sample eigenvectors, two approaches to eigenvector selection considering the amount of entropy (*EBMOGA*) and the highest eigenvalues (*HEMOGA*) have been employed. Thus, we are searching for the best combination of eigenvectors for the optimal clustering state. In order to search the optimal clustering state, convergence method of clusters (in equation 8) and data connectivity in clusters (in equation 10) must be minimized. Therefore, the multi-objective genetic algorithm has been applied. In order to search through various clustering states, undefeated states are firstly located until the Pareto Front is formed [41].

Definition 2- Answer *C** is called non-dominate among the individuals of population, providing there are the following conditions:

$$\forall C \in Population : IND(C^*) < IND(C) \;\; or \;\; Conn(C^*) < Conn(C) \qquad (11)$$

Thus, in the last generation, from answers found in Pareto Front, the most optimal clustering state is selected by the following equation:

$$IND_{mean}(C) = \sum_{j=1}^{k}\left(\sum_{\forall x_i \in C_j} d\left(x_i, m_j\right) \Big/ \left\|C_j\right\|\right) \Big/ k$$
$$f(C) = 1 \Big/ \left(IND_{means} + \frac{Conn(C)}{k}\right) \qquad (12)$$

Where $\left\|C_j\right\|$ is the number of data in the cluster C_j. The algorithm of the proposed approach is as follows:

Algorithm 1- Spectral Clustering of *HEMOGA* and *EBMOGA*

Input: data set $X \in \Upsilon^{n \times d}$ and the number of clusters K

Output: Vector of the cluster label $C \in \Upsilon^{n}$

1. Constructing the similarity matrix using equation 3 and forming the Laplacian matrix. (Similar to the *NJW* method)
2. Calculating the eigenvectors of Laplacian matrix, ranking eigenvalues and forming the diagonal matrix $V=[v_1,v_2,\ldots,v_{Km}] \in \Upsilon^{n \times Km}$, where v_1 to v_{Km} vectors in *HEMOGA* approach are selected based on eigenvalues and in *EBMOGA* approach based on the value of entropy.
3. Initial population of individuals is created randomly.(Each chromosome is in the form of a binary string of length *l*; where each bit indicates the existence state of a corresponding eigenvector to be present in clustering)
4. In order to optimize chromosomes towards local optimal in the space, each chromosome is considered as an initial clustering condition for the first step in *k-means* clustering algorithm. Then after running the *k-means* algorithm, a new state of data placement within various clusters is coded as new chromosome.
5. The fitness of each chromosome is evaluated by using equations 8 and 10.
6. If over half of the population is non-dominated, it goes to step 9.
7. Each chromosome which has a probability proportional to the inverse number of individuals, who have overcome (dominate) it, can enter a new generation using the rank selection algorithm.

[4] In this paper, the value of L, for various datasets, has been considered 3 according to experimental results.

8. Genetic operators (mutation and crossover) are applied and a new generation of chromosomes is created. Then it goes to step 4.

9. According to the fitness function in equation 12, the best combination of eigenvectors ec^* is selected through the chromosomes of the last generation and then the matrix of $U \in \Upsilon^{n \times L^*}$ (L^*=len(ec_i^*)) is constructed.

10. Constructing the matrix Y by re-normalization of each row of U using equation 2.

Each row of matrix Y is considered as new data features of the data set X. Then by *K-means* algorithm, mapped data in the new space are placed in K clusters. Finally, a vector of the cluster label for each data C is defined as the output.

4. Evaluation of the Proposed Method

Since ESBER method uses 10% of ground-truth data as training data for evaluating the quality of selected properties (in equation 7), it is considered a semi-supervised method. While, it should be noted that the proposed method performs entirely unsupervised in comparison with the *ESBER* method.

The results of the proposed method in comparison with *NJW*, *SCWES* and *ESBER* methods for spectral clustering has been collected from the UCI data sets for the nine data sets with different sizes and features in the following Table.

Table 1. Performance Comparison on UCI Benchmark Datasets

Size	Datasets	Properties			Accuracy of Spectral Clustering Algorithms				
		Data Count	Features Count	Class#	Simple	EigenVectors Selection		Proposed Algorithms	
					NJW	SCWES	ESBER	HEMOGA	EBMOGA
Small	Iris	150	4	3	94.00%	94.66%	93.33%	**96.00%**	93.33%
	wine	178	13	3	97.19%	95.51%	96.63%	96.07%	**98.31%**
	glass	214	9	6	47.20%	47.20%	49.07%	**50.93%**	46.73%
Medium	breast-cancer	286	9	2	71.33%	72.38%	**72.73%**	72.38%	53.15%
	haberman	306	3	2	64.71%	65.36%	66.67%	68.63%	**69.61%**
	Ionosphere	351	34	2	76.07%	75.79%	77.21%	71.23%	**85.75%**
Large	Soybean	683	35	19	60.91%	60.91%	61.06%	62.23%	**63.10%**
	Vowel	990	13	11	46.57%	**46.77%**	**46.77%**	43.13%	42.53%
	CMC	1473	9	3	43.86%	44.20%	43.86%	**44.26%**	43.92%

According to the results in Table 1, in most cases the results of the proposed method for ranking based on the highest eigenvalues (*HEMOGA*) and the entropy value (*EBMOGA*) performs much better than the *NJW*, *SCWES* and *ESBER* methods.

5. Conclusion

In this paper, a method for selecting the best combination of eigenvectors in spectral clustering was presented. Most methods for feature selection in spectral clustering are based on weighting each eigenvector considering the role of that feature in presenting clusters. But in [19], it was shown that some appropriate features may sometimes contain the same information about some clusters. Thus, it is essential that the effect of a set of features is considered together. Then in the proposed method employing the conventional methodologies, a set of sample features is firstly selected and therefore an optimal combination is searched by studying various evaluation methods. In this paper, a combination of the best features which simultaneously optimizes both convergence and connectivity criteria of clusters is discovered by the multi-objective genetic algorithm.

In future research, we seek a heuristic method for determining a relationship to the optimal adjustment of parameters L (number of neighboring to consider) and K_m (number of sample eigenvectors).

References

[1] A. K. Jain, M. N. Murty and P. J. Flynn, ACM computing surveys (CSUR), vol. 31, **(1999)**, pp. 264-323.
[2] R. Xu and D. Wunsch, "Neural Networks", IEEE Transactions on, vol. 16, **(2005)**, pp. 645-678.
[3] P. Berkhin, Grouping Multidimensional Data, **(2006)**, pp. 25-71.
[4] H. Hotelling, Journal of educational psychology, vol. 24, **(1933)**, pp. 417.
[5] T. Zhang, D. Tao, X. Li and T. Yang, "A unifying framework for spectral analysis based dimensionality reduction", IEEE World Congress on Computational Intelligence, IEEE, **(2008)**, pp. 1670-1677.
[6] H. Jia, S. Ding, X. Xu and R. Nie, Neural Computing and Applications, vol. 24, **(2014)**, pp. 1477-1486.
[7] U. Von Luxburg, Statistics and Computing, vol. 17, **(2007)**, pp. 395-416.
[8] S. E. Schaeffer, Computer Science Review, vol. 1, **(2007)**, pp. 27-64.
[9] M. C. V. Nascimento and A. C. de Carvalho, European Journal of Operational Research, vol. 211, **(2011)**, pp. 221-231.
[10] S. Guattery and G. L. Miller, SIAM Journal on Matrix Analysis and Applications, vol. 19, **(1998)**, pp. 701-719.
[11] W.-Y. Chen, Y. Song, H. Bai, C.-J. Lin and E. Y. Chang, "Pattern Analysis and Machine Intelligence", IEEE Transactions on, vol. 33, **(2011)**, pp. 568-586.
[12] C. J. Alpert, A. B. Kahng and S. Z. Yao, Discrete Applied Mathematics, vol. 90, **(1999)**, pp. 3-26.
[13] F. Lauer and C. Schnorr, "Spectral clustering of linear subspaces for motion segmentation", Computer Vision, 2009 IEEE 12th International Conference on, IEEE, **(2009)**, pp. 678-685.
[14] J. Shi and J. Malik, "Pattern Analysis and Machine Intelligence", IEEE Transactions on, vol. 22, **(2000)**, pp. 888-905.
[15] C. Fowlkes, S. Belongie, F. Chung and J. Malik, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, **(2004)**, pp. 214-225.
[16] C. Ding, X. He and H. D. Simon, "On the equivalence of nonnegative matrix factorization and spectral clustering", SIAM Data Mining Conference, **(2005)**, pp. 606-610.
[17] F. R. Bach and M. I. Jordan, "Blind one-microphone speech separation: A spectral learning approach", Advances in Neural Information Processing Systems, The MIT Press, vol. 17, **(2005)**, p. 65.
[18] F. R. Bach and M. I. Jordan, The Journal of Machine Learning Research, vol. 7, **(2006)**, pp. 1963-2001.
[19] D. Tritchler, S. Fallah and J. Beyene, Computational statistics & data analysis, vol. 49, **(2005)**, pp. 63-76.
[20] A. Paccanaro, J. A. Casbon and M. A. S. Saqi, Nucleic acids research, vol. 34, **(2006)**, pp. 1571-1580.
[21] D. J. Higham, G. Kalna and M. Kibble, Journal of computational and applied mathematics, vol. 204, **(2007)**, pp. 25-37.
[22] A. B. Lee, D. Luca, L. Klei, B. Devlin and K. Roeder, Genetic epidemiology, vol. 34, **(2010)**, pp. 51-59.
[23] A. K. Jain, Pattern Recognition Letters, vol. 31, **(2010)**, pp. 651-666.
[24] X. Wang, B. Qian and I. Davidson, Data Mining and Knowledge Discovery, vol. 28, **(2014)**, pp. 1-30.
[25] A. Y. Ng, M. I. Jordan and Y. Weiss, Advances in neural information processing systems, vol. 2, **(2002)**, pp. 849-856.
[26] T. Xiang and S. Gong, Pattern recognition, vol. 41, **(2008)**, pp. 1012-1029.
[27] T. Shi, M. Belkin and B. Yu, The Annals of Statistics, vol. 37, **(2009)**, pp. 3960-3984.
[28] F. Zhao, L. Jiao, H. Liu, X. Gao and M. Gong, Neurocomputing, vol. 73, **(2010)**, pp. 1704-1717.
[29] N. Rebagliati and A. Verri, Neurocomputing, **(2011)**, pp. 1391-1401.
[30] F. R. K. Chung, CBMS Regional Conference Series in Mathematics, vol. 92, **(1997)**.
[31] G. L. Scott and H. C. Longuet-Higgins, "Feature grouping by relocalisation of eigenvectors of the proximity matrix", British Machine Vision Conference, **(1990)**, pp. 103-108.
[32] L. Hagen and A. B. Kahng, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 11, **(1992)**, pp. 1074-1085.
[33] P. Perona and W. Freeman, Computer Vision-ECCV'98, **(1998)**, pp. 655-670.
[34] C. H. Q. Ding, X. He, H. Zha, M. Gu and H. D. Simon, "A min-max cut algorithm for graph partitioning and data clustering", Proceedings IEEE International Conference on Data Mining (ICDM2001), IEEE, **(2001)**, pp. 107-114.
[35] K. Oweiss, R. Jin and Y. Suhail, Neurocomputing, vol. 70, **(2007)**, pp. 1728-1734.
[36] M. Filippone, F. Camastra, F. Masulli and S. Rovetta, Pattern recognition, vol. 41, **(2008)**, pp. 176-190.
[37] H. Zhou, D. Tao, Y. Yuan and X. Li, "Object trajectory clustering via tensor analysis", IEEE International Conference on Image Processing, IEEE, **(2009)**, pp. 1945-1948.
[38] S. Ashkezari and H. S. Yazdi, International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 4, **(2011)**.

[39] H. X.Guo, K. J. Zhu, S. W. Gao and T. L., "An Improved Genetic k-means Algorithm for Optimal Clustering", Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06), **(2006)**.
[40] S. Helbig and D. Pateva, OR Spectrum , vol. 16, **(1994)**, pp. 179-186.
[41] A. Konak, D. W. Coit and A. E. Smith, Reliability Engineering & System Safety, vol. 91, **(2006)**, pp. 992-1007.

Authors

Rasoul Taghipour, He is currently Senior Lecturer at Computer Engineering Department of Islamic Azad University, Neyshabur Branch. He received his M.S. degree in computer engineering from Ferdowsi University of Mashhad in 2007.

Rasoul Taghipour, He received the BS degree (first class Honors) in computer engineerin from Azad University of Mashhad in 2006, and the MS degree in software engineering from the Sharif University of Technology(SUT) in 2008, Tehran. His early research focused on bioinformatics, and haplotype data analysis. His current research interests include integration of machine learning and natural language processing for sentiment analysis and information extraction from the web.

Rasoul Taghipour, She is currently Senior Lecturer at Damavand English Institue and Asma English Institute. He received his B.S. degree in computer engineering from Payam-noor University of Mashhad in 2010.