

Eigenvector Selection in Spectral Clustering using Tabu Search

Soheila Ashkezari Toussi, Hadi Sadoghi Yazdi

Department of Computer Engineering
Ferdowsi University of Mashhad
Mashhad, Iran
Soheila.Ashkezari@stu-mail.um.ac.ir,
h-sadoghi@um.ac.ir

Ensie Hajinezhad, Sohrab Effati

Department of Applied Mathematics
Ferdowsi University of Mashhad
Mashhad, Iran
en_ha496@stu-mail.um.ac.ir, s.effati@um.ac.ir

Abstract— Ng, Jordan Weiss (NJW) is one of the most widely used spectral clustering algorithms. For partitioning data into clusters, this method uses the largest eigenvectors of the normalized affinity matrix derived from the data set. However, this set of features is not always the best selection to represent and reveal the structure of the data. In this paper, we aim to propose a quadratic framework to select the most representative eigenvectors. In this way, we define an objective function which includes two factors. In the first part, the interaction of each pair of eigenvectors is considered. In the second part, the ability of each eigenvector to represent the structure of data is considered separately. Then, we use proposed Tabu Search in [1] to solve this mixed-integer quadratic optimization problem. The experimental results show the success of this method to select relevant eigenvectors.

Keywords: *feature/eigenvector selection, spectral clustering, optimization problem, Tabu search.*

I. INTRODUCTION

Clustering can be considered as one of the most important unsupervised learning problems in pattern recognition and machine learning. This problem deals with finding a structure in a collection of unlabeled data in such a way that objects belonging to a cluster are similar to each other and dissimilar to objects belonging to other clusters. There are many clustering algorithms developed in the literature which are mainly based on hierarchical (such as Single Link, Complete Link, etc.) and partitional (such as K-means, Gaussian Mixture Models, Density Estimation, etc.) clustering. When natural clusters in the input space do not correspond to convex region or when the databases become large, varied, and complex and of unknown shapes, especially in the presence of noise, conventional clustering methods yield poor results. In such cases, subspace learning algorithms are proposed, which aim to discover the important statistical distribution on lower dimensions and obtain a new data representation. Subspace learning contains fields such as spectral analysis [2],[3], kernel machine [4],[5], manifold learning [6],[7], and so on. Spectral clustering techniques [8] originate from the spectral graph theory [2] and make use of the spectrum of the similarity matrix of the data to apply dimensionality reduction for clustering in fewer dimensions. Hence, the

basic idea is to construct a weighted graph from the data set in such a way that the vertices of the graph are data points, and each weighted edge represents the degree of similarity between every correspond pair of vertices. Scott and Longute-Higgines algorithm [9], Perona and Freeman algorithm [10], Normalized cut [11] and NJW [12] are such spectral techniques. This kind of method has been successfully applied to parallel computations [13-14], load balancing [15], VLSI design [16], image segmentation [11], speech separation [17], bioinformatics [18-19], etc.

Spectral clustering methods use the eigenvectors of the normalized affinity matrix obtained from data to carry out data partitioning. In most of these techniques the value of corresponding eigenvalues determines the priority of the eigenvectors. For example, for partitioning data to K clusters, NJW always partitions data using the eigenvectors corresponding to the largest K eigenvalues of the normalized Laplacian matrix L_N of a data set. However, this order does not guarantee to select the best features for representing a dataset [20-22]. In [22], it is shown an example where if the gap size between the $K+1$ -th and K -th eigenvalue of L_N is not large, then there is no guarantee on the information content of these first K eigenvectors. For the first time, Xiang and Gong [20] proposed to use eigenvector selection to improve the performance of the spectral clustering method. They proposed to select eigenvectors based on their relevant. In [23] population analyses have been provide by Tao Shia et al. to gain insights into which eigenvectors should be used in clustering. Another approach for eigenvector selection is proposed by Feng Zhao et al. [21] which is based on entropy ranking.

From the viewpoint of identifying relevant features, the problem of selecting the best eigenvectors of the L_N matrix is a feature selection problem, in which, its goal is to select the most representative subset from the set of all the eigenvectors of L_N . In this way, many of the feature selection criteria and algorithms can be utilized. Generally, the search for selecting a variable subset is a NP-hard problem, and it is not guaranteed to achieve the optimal solution, unless through an exhaustive search, which is clearly expensive. To overcome this problem, metaheuristic methods such as Genetic Algorithm (GA), Simulated

Annealing (SA), Memetic Algorithm (MA), Tabu Search (TS) and etc. may be used [24-27].

In this paper, we propose a mixed-integer constrained optimization framework based on the pairwise and individual eigenvector evaluation for feature selection. In such a way, our proposed objective function is similar to a cardinality constraint portfolio optimization. Three powerful metaheuristic approaches to solve this problem are proposed by Chang et al. [1]. Here, we use Chang's TS algorithm to solve the presented eigenvector selection problem.

Indexes like covariance, correlation [28], pairwise distance and maximum information compression index [29] can be used for a pairwise eigenvector evaluation. Also, for evaluation of eigenvectors individually, some criteria such as Fisher ratio [30-31] can be utilized.

This paper is organized as follows: Sections 2 contains a review of spectral clustering concepts and one of the widely used techniques, i.e. NJW. In addition, we review related works in feature selection methods and eigenvector selection in spectral clustering. In Section 3, we propose the eigenvector selection scheme which is a mixed-integer constrained optimization problem and explain how we use TS to solve it. Section 4 contains the empirical results and related discussion, and Section 5 concludes the article.

II. PRELIMINARIES

In this section, we briefly review the spectral clustering and one of its known algorithms, NJW. Afterward, some pairwise and individual feature evaluation criteria and presented selecting relevant eigenvectors approach in the literature are reviewed.

A. Spectral Clustering

Spectral clustering technique [8] has a strong connection with spectral graph theory [2]. It usually refers to the graph partitioning based on the eigenvalues and eigenvectors of the adjacency (or affinity) matrix of a graph. Given a set of N points in d -dimensional space $X=\{x_1, x_2, \dots, x_N\}$ we can build a complete, weighted undirected graph $G(V, A)$ whose nodes $V=\{v_1, v_2, \dots, v_N\}$ correspond to the N patterns and edges defined through the adjacency matrix A encode the similarity between each pair of sample points. Adjacency between two data points can be defined as follows:

$$a_{ij} = \begin{cases} h(x_i, x_j) & i \neq j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The function h measures the similarity between patterns by a typical Gaussian function

$$h(x_i, x_j) = \exp(-d^2(x_i, x_j)/\sigma^2) \quad (2)$$

where, d measure the dissimilarity or distance between patterns and the scaling parameter σ controls how rapidly the affinity h falls off with the distance between x_i and x_j . The selection of tuning parameter σ greatly affects spectral clustering results, and results under various values of σ are very different from one another. The tuning method proposed by Zelnik and Peona [32] introduces local scaling by selecting a σ_i for each data point x_i instead of the fixed scale parameter σ . The selection is done using the distance between the point x_i and the p^{th} nearest neighbor. In this way, utilizing Euclidian norm, the similarity matrix is defined as, $h(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / \sigma_i \sigma_j)$, $\sigma_i = d(x_i, x_p)$, and x_p is the p^{th} nearest neighbor of point x_i . However, this approach depends on the choice of the nearest neighbor p .

1) Ng-Jordan-Weiss (NJW) method

The idea of NJW algorithm [12] to partition patterns into K clusters, is to find a new representation on the first K eigenvectors of the Laplacian matrix. The algorithm composed of the following steps:

1. Form the affinity matrix A by Equation 1.
2. Compute the degree matrix D and normalized affinity matrix $L_N = D^{-1/2} A D^{-1/2}$. The degree matrix D is a diagonal matrix whose $D_{ii} = \sum_{j=1}^N A_{ij}$ element is the degree of the point x_i .
3. Compute the first K eigenvectors v^1, v^2, \dots, v^K corresponding to the first K largest eigenvalues $\lambda^1, \lambda^2, \dots, \lambda^K$ of L_N and form the column wise matrix $V = [v^1, v^2, \dots, v^K] \in R$.
4. Renormalize the matrix V and form the matrix Y such that all rows have unit length $Y_{ij} = V_{ij} / \sqrt{\sum_j V_{ij}^2}$.
5. Cluster represented data matrix Y into K clusters via K-means.

B. Feature selection methods and spectral clustering

There are many methods for measuring similarity between two features, or two random variables, based on the amount of functional (linear or higher) dependency between them, such as covariance, Correlation coefficient [28], Maximal Information Compression Index [33]. Recently, there has been a focus on selecting appropriate features in spectral clustering whose importance is reflected in the fact that using uninformative features could lead to poor clustering results. In this sub section, we first review the Fisher [30-31], correlation [28], feature pairwise distance and Maximum Information Compression Index [33]. Then, in particular, we discuss about feature/eigenvector selection in the spectral clustering literature.

Fisher ratio [30-31], is a popular feature selection algorithm based on the maximization of the ratio of the between-class scatter to the within-class scatter. Between-class scatter is a measure of the scatter of each class relative to the mean(s) of the other classes. Within-class scatter gives us a measure of how the points in each cluster are distributed about its mean. We use this measure for evaluating eigenvectors' relevance as follows,

$$f = \frac{\sum_{\substack{i,j=1 \\ i \neq j}}^K \|\bar{x}_i - \bar{x}_j\|}{\sum_{i=1}^K \|\text{var}(x_i)\|} \quad (3)$$

where \bar{x}_i is the center of the i^{th} cluster, and $\text{var}(x_i)$ is the variance of the i^{th} cluster. Whatever the value of this index is higher data points are well distributed among clusters.

Correlation coefficient (ρ) [28], is the most well-known measure of similarity between two random variable x and y , which is defined as,

$$\rho(x, y) = \frac{\Sigma_{x,y}}{\sqrt{\text{var}_x \text{var}_y}} \quad (4)$$

where var and Σ denotes the variance of a variable and covariance between two variables respectively. If $\rho(x,y)=1$ means that x and y are completely correlated.

Features pairwise distance (d), for two feature vector x_i and x_j we define l_1 distance, $|x_i - x_j|$, as the pairwise distance between x_i and x_j . Whatever this value is higher it means related features are more distinct from each others. So the sum of pairwise distance can be considered as a dissimilarity criterion, as follows,

$$d(x_i, x_j) = \sum_{i=1}^N |x_i - x_j| \quad (5)$$

Maximum Information Compression Index (λ_2) [33], is a suitable measure to reduce redundancy in a dataset. λ_2 is the smallest eigenvalue of the covariance matrix and is defined as, $2\lambda_2(x, y) = \text{var}(x) + \text{var}(y) -$

$$\sqrt{\text{var}(x) + \text{var}(y) - 4\text{var}(x)\text{var}(y)(1 - \rho(x, y))^2} \quad (6)$$

When the features are linearly dependent, the value of λ_2 is zero. Whatever the amount of dependency decreases, this value increases. Actually, λ_2 is the amount of reconstruction error committed if the two dimension data is projected to the one dimension one.

Feature selection in spectral clustering is an important subject. Selecting irrelevant eigenvectors cause to reduce the accuracy of the clustering. In [22], authors aim to select the best eigenvectors of the Laplacian matrix of NJW algorithm to cluster data into K groups. In this way, after finding the largest $K_m > K$ eigenvectors of the normalized affinity matrix of the data, the relevance of each eigenvector is estimated according to how well it can separate data into different clusters based on the probability density function (*pdf*) of the vectors and then relevance eigenvectors are used to perform data clustering. The result of this method in image segmentation is satisfactory. However, the eigenvector combination, including all these significant eigenvectors cannot always represent the data structure [21]. The proposed method in [23] Concentrate to find the

connection between the population P and the eigenvalues and eigenfunctions of the defined distribution-dependent convolution operator and make a decision to select relevance eigenvectors. Furthermore, another approach for feature selection in eigenspace for spectral clustering based on entropy is proposed in [21]. In this approach, after finding the eigenvectors of the normalized Laplacian matrix, all of these vectors will be ranked according to their importance in clustering and then a suitable subset of these eigenvectors is selected and combined to represent the structure of the original data. To evaluate the importance and obtain the ranking list of eigenvectors, they proposed to remove each eigenvector in turn and compute the corresponding entropy of the set of the rest eigenvectors. If the removal of an eigenvector causes more disorder in the system than another, it shows more importance and higher rank of this eigenvector. This approach is semi-supervised, in which, labels of ten percent of each cluster's data are considered to find out the best combination of eigenvectors.

III. PROPOSED EIGENVECTOR SELECTION APPROACH

The block diagram of the proposed algorithm to select the most relevant eigenvectors in spectral clustering is illustrated in Fig.1. We aim to select the most representative K_m eigenvectors of the normalized Laplacian matrix. In this way, we introduce the zero-one decision variables z_i , $i=1, \dots, N$ to show the presence or absence of the i^{th} eigenvector Y_i in the solution. The cardinality constraint $\sum_{i=1}^N Z_i = K_m$ is considered to select exactly K_m eigenvectors. The proportion of contribution of each eigenvector Y_i is determined with the weight α_i , $i=1, \dots, N$. If the i^{th} eigenvector is selected, then its minimum proportion contribution is ϵ_i which is declared with $\alpha_i \geq z_i \epsilon_i$, for $i=1, \dots, N$. In our proposed method, two factors are considered to select appropriate eigenvectors. The first one is dealing with selecting those eigenvectors that are well-structured, i.e. the clusters on each of them are as separate as possible. Therefore, we define the objective function $\sum_{i=1}^N \alpha_i Y_i$, that Y_i is the result of the evaluation of eigenvector Y_i . A good choice for Y is Fisher ratio (3). The second factor is related to the interaction of each pair of eigenvectors. It is preferred to select those eigenvectors that are independent to each other to reduce the redundancy in a dataset. For this purpose, we define the objective function $\sum_{i,j=1}^N \alpha_i \Psi_{ij} \alpha_j$, where Ψ_{ij} indicates the pairwise evaluation index for two eigenvectors Y_i and Y_j . Feature pairwise distance (5), correlation (4) maximum information compression index (6) and covariance can be considered as some choices for Ψ_{ij} . Now, by introducing a weighting parameter ω ($0 \leq \omega \leq 1$) and taking into account mentioned constraints, we state our multi objective optimization problem in the term of a mixed-integer single objective quadratic optimization to select the most representative eigenvectors of normalized Laplacian matrix for spectral clustering as follows,

$$\text{Min } \omega \sum_{i,j=1}^N \alpha_i \Psi_{ij} \alpha_j - (1 - \omega) \sum_{i=1}^N \alpha_i Y_i$$

subject to

$$\begin{aligned}
 \sum_{i=1}^N z_i &= K_m \\
 \sum_{i=1}^N \alpha_i &= 1 \\
 \alpha_i &\geq \varepsilon_i z_i, i = 1, \dots, N \\
 z_i &\in [0, 1], i = 1, \dots, N
 \end{aligned} \quad (7)$$

This problem is similar to the cardinality constrained portfolio optimization proposed by Chang et. al [1]. Therefore, we use TS heuristic which is presented in [1] to solve optimization problem (7).

TS is a metaheuristic strategy proposed by Glover [34,35]. This procedure explores the solution space beyond the local optimum. In order to escape from local minima, the so-called tabu list is used, which determines the forbidden moves.

In TS heuristic the solution representation has two distinct parts, a set of K_m distinct eigenvectors and K_m real numbers related to the eigenvectors weight. After computing each solution, to ensure the feasibility of the solution is satisfied, *evaluate* algorithm in [1] is called. In our work, we use the first K_m eigenvectors as a starting point.

The move operator takes all K_m eigenvectors present in the solution and multiplying their weights by 0.9 and 1.1. Moreover, the tabu list is a matrix of $2N$ integer values which determines for each of the N eigenvectors whether a particular move (multiplying by 0.9 or 1.1) is currently tabu or not. The search terminates when it reaches a given number of iterations. More details are discussed in [1].

IV. EXPERIMENTAL RESULT

To investigate the performance of the proposed algorithm for selecting relevant eigenvectors, we use a number of data sets from UCI Machine Learning Repository as benchmark comparisons in terms of accuracy of the clustering. Properties of the considered datasets are reported in Table I and parameters of TS algorithm are reported in Table II. To form the affinity matrix, we utilized the proposed method in [32] using the 7th nearest neighbor of each data point. The accuracy of clustering for classic NJW is reported in Table III. To assign equal weights to two factors in our cost function $\omega=0.5$ is considered.

TABLE I. UCI BENCHMARK DATA SETS PROPERTIES

Data sets	Instance	Attributes	Classes
Soybean	47	35	4
wine	178	13	3
Liver	345	6	2
Ionosphere	351	34	2
Breast (wbdc)	569	30	2
Image segmentation	2310	19	7

TABLE II. TABU SEARCH ALGORITHM PROPERTIES

Parameter	Iteration	Population size	Weighting parameter	Lower bound
value	1000	5000	0.5	0.01

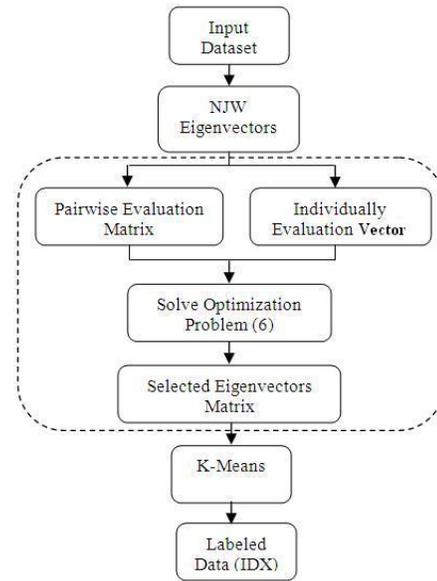


Figure 1 . Block diagram of the proposed algorithm

TABLE III. ACCURACY OF CLUSTERING USING NJW

Wine	Breast	Soybean	Ionosphere	Liver	Image
0.7217	0.8908	0.7868	0.7827	0.5470	0.6169

In optimization problem (7), for pairwise evaluation of eigenvectors we used covariance, correlation (4), feature pairwise distance (5) and λ_2 . In addition, for estimation the ability of each eigenvector individually, we used Fisher ratio (3). The number of considered eigenvectors for TS is $N=40$. For each of six data sets of Table I, the accuracy of clustering based on mentioned indexes vs. different numbers of selected eigenvectors $K_m=1, \dots, 30$, is plotted in Fig.2-3. By comparing the NJW accuracy (Table III) with plotted accuracy of the proposed method, it is obvious that the selected features are able to reveal the structure of data sets well. For example, using just K best selected eigenvectors lead to equal or higher accuracy of the proposed method for all datasets in comparison with classic NJW. This issue shows the fact that first K eigenvectors related to largest eigenvalues are not always the best set of eigenvectors to represent the structure of a dataset.

V. CONCLUSIONS

A framework for selecting relevant eigenvectors in spectral clustering is proposed. This framework is a model

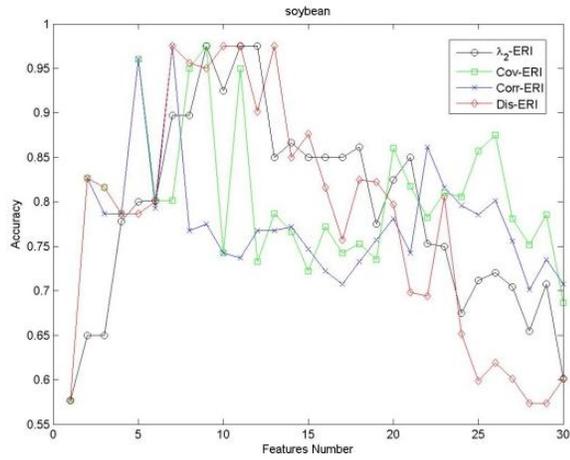
based on two factors pairwise and individually evaluation of normalized Laplacian matrix eigenvectors. Sort of pairwise feature evaluation indexes such as covariance, correlation, pairwise distance are used for pairwise eigenvector evaluation. In addition, for investigating each eigenvector individually, the quality of the formed cluster on each eigenvector is evaluated using the fisher ratio. The TS algorithm proposed by Chang et al. [1] is used to solve the proposed optimization problem. The weighting parameter is considered equal with 0.5, so the contribution proportion of two factors is equal. For the future work, we aim to investigate more indexes for pairwise and individually evaluation of eigenvectors.

ACKNOWLEDGMENT

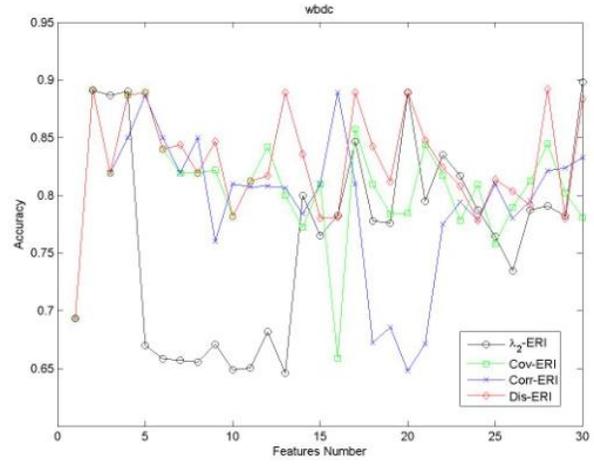
This work has been supported by Iran Research Institute for ICT (ITRC), Tehran, Iran (Contract number: T/500/18496). This support is gratefully acknowledged.

REFERENCES

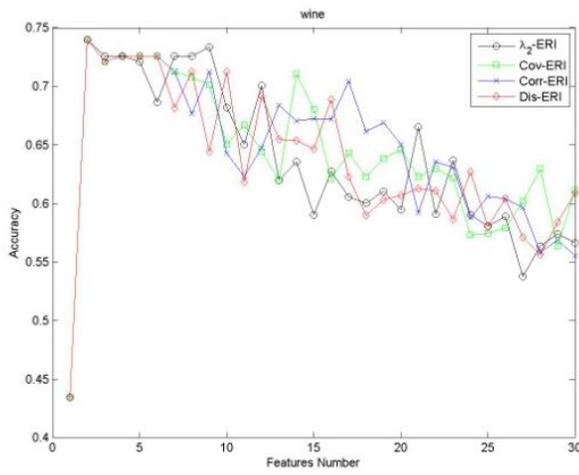
- [1] T.-J. Chang, et al., "Heuristics for cardinality constrained portfolio optimisation," *Comput. Oper. Res.*, vol. 27, pp. 1271-1302, 2000.
- [2] F. R. K. Chung, "Spectral Graph Theory," *CBMS Regional Conference Series in Mathematics*, vol. 92, 1997.
- [3] T. Zhang, et al., "A unifying framework for spectral analysis based dimensionality reduction" in *IEEE World Congress on Computational Intelligence*, 2008, pp. 1670-1677.
- [4] K. R. Müller, et al., "An introduction to kernel-based learning algorithms," *Neural Networks, IEEE Transaction on*, vol. 12, pp. 181-202, 2001.
- [5] B. Schölkopf and A. J. Smola, *Learning with Kernels, Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2002.
- [6] S.T. Roweis and L.K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323-2326, 2000.
- [7] J.B. Tenenbaum, et al., "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319-2323, 2000.
- [8] N. Cristianini, et al., "Spectral kernel methods for clustering," in *Neural Information Processing Systems, NIPS'01*, 2001, pp. 649-655.
- [9] G. L. Scott and H. C. Longuet-Higgins, "Feature grouping by relocalisation of eigenvectors of the proximity matrix," presented at the British Machine Vision Conference, 1990.
- [10] P. Perona and W. T. Freeman, "A factorization approach to grouping," in *ECCV*, 1998, pp. 655-670.
- [11] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, pp. 888-905, Aug. 2000.
- [12] A. Y. Ng, et al., "On spectral clustering: analysis and an algorithm," *Advances in Neural Information Processing Systems*, vol. 14, 2002.
- [13] B. Hendrickson and R. Leland, "An improved spectral graph partitioning algorithm for mapping parallel computations," *SIAM Journal of Scientific Computing*, vol. 16, pp. 452-469, 1995.
- [14] A. Pothen, "Graph partitioning algorithms with applications to scientific computing, *Parallel Numerical Algorithms*," Old Dominion University Norfolk, VA, USA 1997.
- [15] R. V. Driessche and D. Roose, "An improved spectral bisection algorithm and its application to dynamic load balancing," *Parallel Computing*, vol. 21, 1995.
- [16] L. Hagen, "New spectral methods for ratio cut partitioning and clustering," *Computer Aided Design, IEEE Transaction on*, vol. 11, pp. 1074-1085, 1992.
- [17] F.R. Bach and M.I. Jordan, "Learning spectral clustering, with application to speech separation," *Journal of Machine Learning*, vol. 7, pp. 1963-2001, 2006.
- [18] D. Tritchler, et al., "A spectral clustering method for microarray data," *Computational Statistics & Data Analysis*, vol. 49, pp. 63-76, 2005.
- [19] D. J. Higham, et al., "Spectral clustering and its use in bioinformatics," *Journal of Computational and Applied Mathematics*, vol. 204, pp. 25-37, 2007.
- [20] T. Xiang and S. Gong, "Spectral clustering with eigenvector selection," *Pattern Recognition*, vol. 41, pp. 1012-1029, 2008.
- [21] F. Zhao, et al., "Spectral clustering with eigenvector selection based on entropy ranking," *Neurocomputing*, vol. 73, pp. 1704-1717, 2010.
- [22] N. Rebagliati and A. Verri, "Spectral clustering with more than K eigenvectors," *Neurocomputing*, vol. 74, pp. 1391-1401, 2011.
- [23] T. Shi, et al., "Data spectroscopy: Eigenspaces of convolution operators and clustering," *Annals of Statistics*, vol. 37, pp. 3960-3984, 2009.
- [24] Y. Wang, et al., "Feature selection using tabu search with long-term memories and probabilistic neural networks," *Pattern Recognition Letters*, vol. 30, pp. 661-670, 2009.
- [25] S. C. Yusta, "Different metaheuristic strategies to solve the feature selection problem," *Pattern Recognition Letters*, vol. 30, pp. 525-534, 2009.
- [26] P. Bermejo, et al., "A GRASP algorithm for fast hybrid (filter-wrapper) feature subset selection in high-dimensional datasets," *Pattern Recognition Letters*, vol. 32, pp. 701-711, 2011.
- [27] H. Zhang and G. Sun, "Feature selection using tabu search method," *Pattern Recognition*, vol. 35, pp. 701-711, 2002.
- [28] M. A. Hall, "Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning," presented at the Proceedings of the Seventeenth International Conference on Machine Learning, 2000.
- [29] P. Mitra, et al., "Unsupervised feature selection using feature similarity," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, pp. 301-312, 2002.
- [30] R. O. Duda, et al., *Pattern Classification 2ed.*: Wiley-Interscience, 2000.
- [31] F. Nie, et al., "Trace ratio criterion for feature selection," presented at the Proceedings of the 23rd national conference on Artificial intelligence - Volume 2, Chicago, Illinois, 2008.
- [32] L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering," *Advances in Neural Information Processing Systems*, vol. 17, pp. 1601-1608, 2004.
- [33] P. Mitra, et al., "Unsupervised Feature Selection Using Feature Similarity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, pp. 301-312, 2002.
- [34] Glover, F., "Tabu Search: Part I". *ORSA J. Comput.* 1, 190-206, 1989.
- [35] Glover, F., "Tabu Search: Part II". *ORSA J. Comput.* 2, 4-32, 1990.



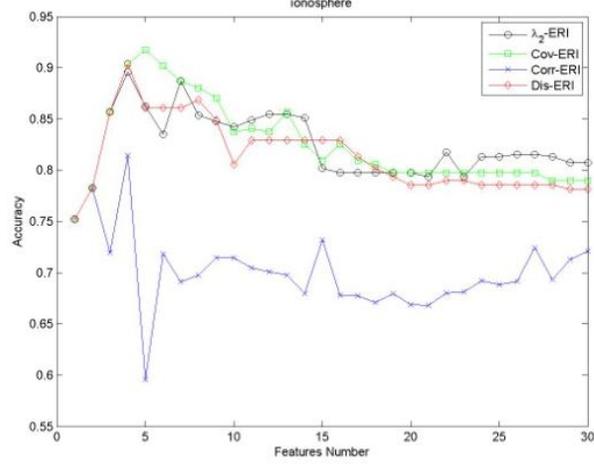
(a)



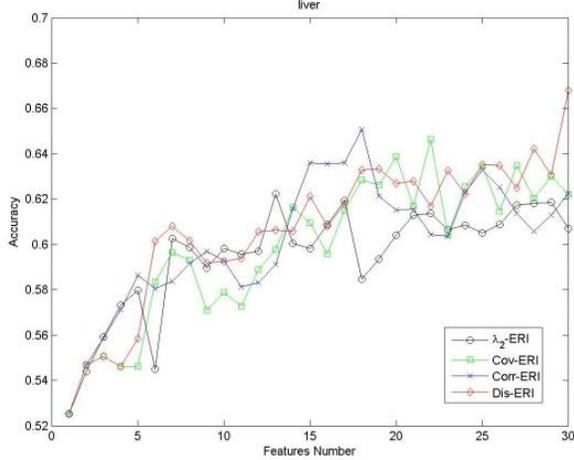
(a)



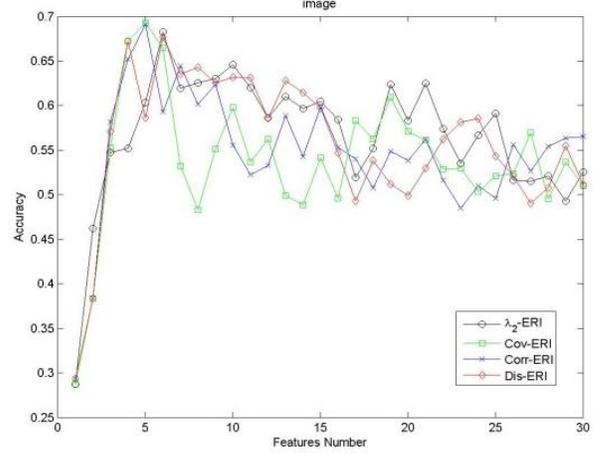
(b)



(b)



(c)



(c)

Figure2 . Clustering accuracy vs. feature number for Soybean(a), Wine (b) and Liver(c)

Figure3 . Clustering accuracy vs. feature number for WDBC(a), Ionosphere(b) and Image(c)