International Journal of Electrical and Computer Engineering (IJECE)
Vol. 5, No. 2, April 2015, pp. 361 – 370
ISSN: 2088-8708

A Novel Spectral Clustering based on Local Distribution

Parthajit Roy* **and J. K. Mandal****
*Department of Computer Science, The University of Burdwan, Burdwan, West Bengal, India-713104
**Dept. of Comp. Sc. & Engg., The University of Kalyani, Nadia, West Bengal, India-741235

Article Info

Article history:
Received Nov 22, 2014
Revised Jan 20, 2015
Accepted Feb 5, 2015

Keyword:
Spectral Clustering
Affinity
Mahalanobis Distance
Outlier Detection
Random Walk Laplacian
Clustering Indices

Corresponding Author:
J. K. Mandal
Dept. of Comp. Sc. & Engg.
The University of Kalyani
Nadia, West Bengal
India-741235
e-mail: jkm.cse@gmail.com

ABSTRACT

This paper proposed variation of spectral clustering model based on a novel affinity metric that considers the distribution of the neighboring points to learn the underlaying structures in the data set. Proposed affinity metric is calculated using Mahalanobis distance that exploits the concept of outlier detection for identifying the neighborhoods of the data points. Random Walk Laplacian of the representative graph and its spectra has been considered for the clustering purpose and the first k number of eigenvectors have been considered in the second phase of clustering. The model has been tested with benchmark data and the quality of the output of the proposed model has been tested in various cluster validity indexes.

1. INTRODUCTION

Data clustering is the process of grouping the nonuniform data elements by identifying the underlaying structure[1]. Data clustering is a difficult task from computing's point of view. This difficulty is because data clustering is computationally hard. There are different approaches for data clustering and a whole branch of computer science has been devoted for that. General reference on data clustering is due to Everitt et al[2]. Jain et al surveyed clustering, taxonomy of clustering and recent trends[3]. A More recent survey is done by Xu et al[4]. Some recent applications of clustering has been done by Farshchian et al [5] and Sharma et al [6].

There are different approaches towards clustering. Some of the popular approaches are Hierarchical-, Density Based-, Squared-Error based-, Fuzzy-based-, and Graph based clustering[4].

The purpose of the present study is to cluster spatial data sets using graph based model. The cluster shape and its orientation is most important parameter in the present work. The objective of the present research work is to handle the clustering problems where the cluster shapes are non-convex and moreover the present model considers the trend of the data sets and not just the proximities.

The graph based model has gained enormous popularity in recent days. Graph is a very good algebraic structure for defining proximity among the data points and proximity among the points is the key for the success of almost every clustering algorithm. There are different areas of graph theory that can be exploited for data clustering. Some Delaunay graph based model is available due to Yang et al [7] and Roy et al[8]. Some minimum spanning tree based clustering technique has been proposed by Foggia et al [9] and Roy et al[10]. Some early minimum as well as maximum spanning tree based approaches were proposed by Asano et al[11]. An outstanding survey on Graph clustering is done by Shaefer[1] where the recent developments has been outlined explicitly.

Spectral clustering is one of the major branches of the graph based clustering where the spectra i.e. the eigenvalues and eigenvectors of the Laplacian of the graph is considered for clustering. There are several papers on this topic. Some of them are due to Higham et al [12], Toussi et al [13] and Qiu et al [14]. A very good survey on spectral clustering is due to Luxburg[15]. A general survey on graph Laplacian has been done by Chung[16]. Some

variation of the traditional spectral clustering is available due to Spielman et al[17] and Fiedler[18].

Affinity or distance among the data points is the most important input to any clustering model. This paper uses a Mahalanobis distance proposed by Mahalanobis[19] based novel affinity matrix for spectral clustering. A good discussion on Mahalanobis distance can be found in a book by Marsland[20]. Mahalanobis distance has been used effectively for outlier ditection by Hodge et al [21].

The quality of the output produced by any clustering algorithm is measured by several validity indices. There are two major types of cluster validity indices. Internal and external indices. Internal indices measures the quality of the cluster produced, by the distribution of the data and the density/scatterness of the data points. The external indices, on the other hand, considers some labeled data and the quality of the clusters is measured on the basis of that. A good comparative study on the performance of various clustering indices is done by Saha et al[22].

The rest of the article is organized as follows. Section 2. discusses the mathematical backgrounds necessary of the model. Section 3. presents and discusses the proposed model. Section 4. discusses the results and analysis. Conclusion comes in the section 5. and references are drawn at the end.

2. MATHEMATICAL BACKGROUND

This section uncovers the necessary mathematics required to develop and explain the proposed model. The proposed model is a spectral clustering model based on a novel affinity metric that calculates the proximity among the data points in a novel way. The rest of this section discusses spectral graph, similarity measures and other necessary mathematical backgrounds in the following manner. Subsection 2.1. discusses the clustering as an optimization problem. Subsection 2.2. discusses spectral graph theory and subsection 2.4. discusses the distance and similarity measures.

2.1. Data Clustering

Given a set of data points, the consideration of clustering is to identify the grouping by exploring the underlaying structures in a data set, if there exists any. The main assumption in the clustering is that the property of the underlying structure of the data set is not known. Only the number of clusters may be known sometimes.

formally clustering can be defined as follows:

Given $S = \{p_1, p_2, p_3, p_4, \cdots, p_n\}$ is the set of n data points in m-dimensional space R^m, the clustering is the partition of S into k different groups, k being the number of clusters, $C = \{C_1, C_2, C_3, \cdots, C_k\}$ with respect to a distance metric $d(p_i, p_j)$ in such a way that the inter-cluster distance becomes maximum and the intra-cluster distance becomes minimum, over all the partitions[3]. i.e. the objective of clustering is to minimize the ratio of intra cluster distance and inter cluster distance as shown in equation 1.

$$Minimize\ Z = \sum_{\forall C_i \in C} \frac{IntraCluster\ \ Distance}{InterClustter\ \ Distance}\ \ w.r.t.\ \ d(p_i, p_j) \tag{1}$$

2.2. Spectral Graph Theory

A *graph* is an algebraic structure $G =< V, E >$, where V is called the *vertex* set and E is called the *edge* set. The set E can be defined as a relation ρ over the cartesian product $V \times V$, defined as $x\rho y => (x, y) \in E$ which implies that there is an edge between x and y of the vertex set V[23].

Parallel edges implies more than one edge between two vertices and in our case parallel edges are not allowed. A graph consists of self loop if $x\rho x$ for some $x \in V$. A graph $G =< V, E >$ is called an *undirected* graph, if the relation rho is symmetric.

A graph is called a *simple* undirected graph or S-graph, if it is undirected graph without any self loop or parallel edges.

A graph is weighted, if there is a weight function $d(v_i, v_j)$ associated with every edge $e \in E$. A weight function $d(.,.)$ is a *metric* if it follows the following properties:

- **Non-negativity:** $\forall x, y \in R^m, d(x, y) \geq 0$.
- **Symmetricity:** $\forall x, y \in R^m, d(x, y) = d(y, x)$.
- **Triangular Inequality:** $\forall x, y, z \in R^m, d(x, y) + d(y, z) \geq d(x, z)$.

Given a weighted graph G, the Adjacency matrix representation of the graph is a square matrix of order $n = |V|$ where the entry $w_{i,j}$ ($w_{i,j} \geq 0$), is the weight of the edge $e(i, j)$. i.e. the adjacency matrix W can be defined as,

$$W_{n\times n} = (w_{ij})_{i=1,2,\cdots,n;j=1,2,\cdots,n} \tag{2}$$

The degree d_i of a vertex $v_i \in V$ of a graph, represented as per equation 2, can be defined as sum of the weights of the incident edges on a particular vertex as shown in the equation 3.

$$d_i = d(v_i) = \sum_{j=1}^{n} w_{ij} \tag{3}$$

The degree matrix D is a diagonal matrix with the degrees d_i are the leading diagonal of the matrix D.

The tool for spectral clustering is Laplacian matrix of the graph. A Laplacian of a graph is a symmetric matrix that can be formed from the adjacency matrix. The eigenvalues and eigenvectors of the Laplacian are the most important tools for analyzing the structure of a graph. Specially the cut related things can algebraically be analyzed by computing the values of the eigenvalues and eigenvectors of the graph Laplacian. There exists several graph Laplacians and every Laplacian has its own strength and weakness in a particular situation [15]. A detailed literature on spectral graph theory has been given by Chung[16].

Unnormalized Laplacian: The unnormalized Laplacian of a graph, denoted as L, can be defined as per the equation 4.

$$L = D - W. \tag{4}$$

where D is the diagonal degree matrix and W is the adjacency matrix. The important thing to note that the Laplacian defined above, is a real symmetric matrix whose diagonal elements are all non-negative and all other elements are negative. Also, the sum of every row is zero. The spectra is the eigenvectors of the this Laplacian. There are several important properties of this spectra. Some of the properties of this Laplacian [24][25][15] are given as follows.

1. $f'Lf = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(f_i - f_j\right)^2, \ \ \forall f \in R^n$.
2. L is symmetric and positive semi-definite.
3. The smallest eigenvalue of L is 0.
4. All eigenvalues are real with $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$.
5. If the graph is not connected, the the number of 0 eigenvalue, i.e. the multiplicity of eigenvalue 0, is the number of connected components.

Normalized Laplacian: The spectra of unnormalized Laplacian of a graph is independent of the matrix D. Normalized Laplacian overcomes this. Two of the normalized Laplacians are very popular. These are *Symmetric* Laplacian and *Random Walk* Laplacian.

Symmetric Laplacina L_{sym} is a symmetric matrix define as per the equation 5.

$$L_{sym} = D^{-1/2}LD^{-1/2} \tag{5}$$

The Random Walk Laplacian models the Random Walk in a graph. i.e. starting from an arbitrary vertex of a graph, if we randomly go to any of its adjacent vertex and follow the same process repetitively, then in a densely connected subgraph of the original graph, the walk will be ended in the same subgraph in most of the cases. This property is modeled in a Random Walk Laplacian. The random walk graph L_{rw} can be defined as per the equation 6.

$$L_{rw} = D^{-1}L = I - D^{-1}W \tag{6}$$

All of the properties for unnormalized Laplacian holds good for L_{rw} also. In the present paper, the random walk Laplacian and its spectra will be used for clustering.

2.3. Similarity Measurement

The success of the graph based clustering algorithms lie on the choice of similarity matrix. The similarity between two data points is the only means for the creation of an edge in the graph between them. If the edges are more close to the actual relationship, the success possibility of the algorithm is also high. For this, similarity measurement demands adequate research.

There are several types of similarity measures available. Some of them exploits the distance between the data points only. The following part discusses two of them.

$\epsilon-$**neighbor:** In this case, if the distance between two data points is less then a predefined benchmark value ϵ, then the weight is assigned. Otherwise 0 is assigned.

Gaussian Similarity: In this case, the distance between the data points is computed as,

$$g\left(\vec{X},\vec{Y}\right)=e^{-\frac{\|\vec{X}-\vec{Y}\|^2}{2\sigma^2}} \tag{7}$$

The value of σ in the equation 7 known as the *strength of inclusion*, i.e. more the value of σ, the more the weight will be given to the edge between the data points. i.e. far points will also come under consideration with heavy weight.

2.4. Distance Metric

Given two data points, the ultimate tool for designing the Laplacian matrix is similarity among data points and the similarity is based on the distance among data points, which is the weight of the edge between them. There are several distance functions available but the most popular is the Euclidean distance defined as,

$$E(i,j)=\left(\sum_{k=1}^{m}\left|x_i\left(k\right)-x_j\left(k\right)\right|^2\right)^{1/2} \tag{8}$$

The generalization of the Euclidean distance is called the Minkowski distance define as,

$$M(i,j)=\left(\sum_{k=1}^{m}\left|x_i\left(k\right)-x_j\left(k\right)\right|^n\right)^{1/n} \tag{9}$$

There are other distance functions also. But These two are the most popular.

3. PROPOSED MODEL

This section presents proposed model. The Eucledian (Equation 8) and Minkowski (Equation 9) of the previous section gives the proximity but their main disadvantage is they measures the distance from a single point. Even if the distance needs to be measured from a distribution, they first convert it to a single point (like mean, median etc.) and calculate the distance. This mean or median is called representative of the distribution and often they do not show good result. Consider the following situation. Let there are 12 points in two dimensions. Some values are shown in Table 1 and the distrubution are shown graphically in figure 1 (a). Suppose the black-fill spots are some distribution. In our case this will be the neighbor set of some point.

Table 1. Sample points in two dimensions

x	3.0	2.9	2.8	2.7	...	2.0
y	1.0	1.1	1.0	1.0	...	1.1

We would like to find out the distance of any other points, like $P1$ or $P2$ from this distribution. Clearly the point $P2$ is more close to this distribution than $P1$. But how to measure that mathematically? The traditional way will find the mean of the distribution and will try to find the distance of the given point from the mean. But this is not good here, because point $P1$ and $P2$ are equidistance from the mean, so either both will be rejected or both will be accepted. Figure 1 (a) shows this.

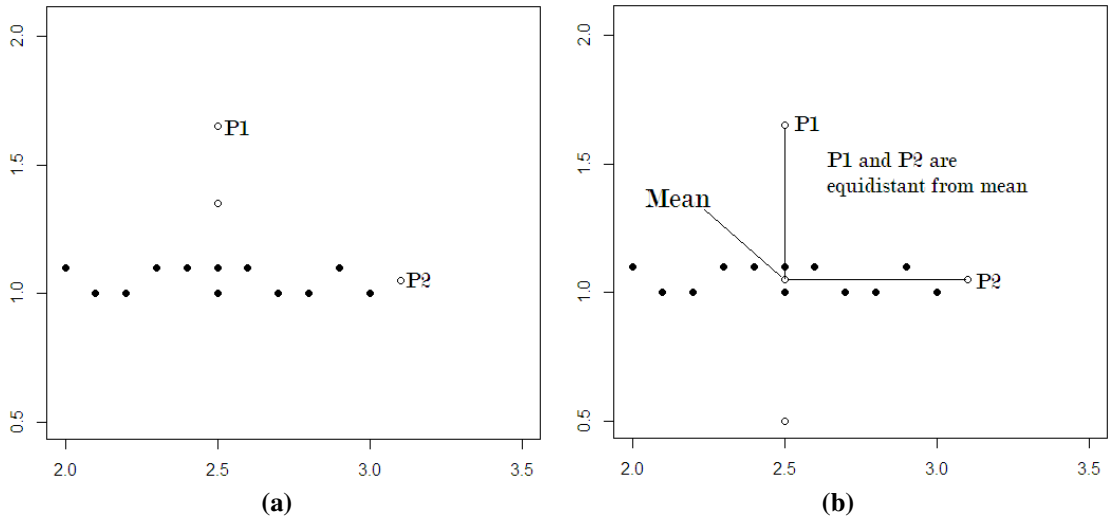

Figure 1. (a) Distribution of the point set of table 1 (b) Point P1 and P2 are equidistant from the mean of the distribution.

The solution to this problem is the Mahalanobis distance[19]. Mahalanobis distance is a distance measure which addresses the above problem in a better way. Following are steps for calculating Mahalanobis distance.

Given a data set $S = \{p_1, p_2, \cdots, p_n\}$ where each $p_i \in R^m$, the mean is defined as per equation 10.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} p_i \tag{10}$$

variance is defined as per equation 11.

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} (p_i - \mu)^2 \tag{11}$$

The Covariance matrix $M_{m\times m}$, where m is the number of attributes of the data set, can be defined as,

$$M = cov(i,j); \forall i = 1.\cdots m; \;\; j = 1, \cdots, m \tag{12}$$

and the Mahalanobis distance between x_1 and x_2 is defined using the equation 12 as,

$$d = \sqrt{x_1 M^{-1} x_2} \tag{13}$$

In the figure 1 (b), the Euclidean distance between $P1$ and the mean is 0.6. The distance of $P2$ from the mean is also 0.6. So, any algorithm that considers Euclidean distance will either include both of them or will reject both of them, which is unrealistic. The Mahalanobis distance (equation 13, on the other hand, for $P1$ from the distribution is 34.28 whereas the distance of $P2$ from the distribution is 24.11. This clearly states that the point $P2$ is more close to the distribution than point $P1$.

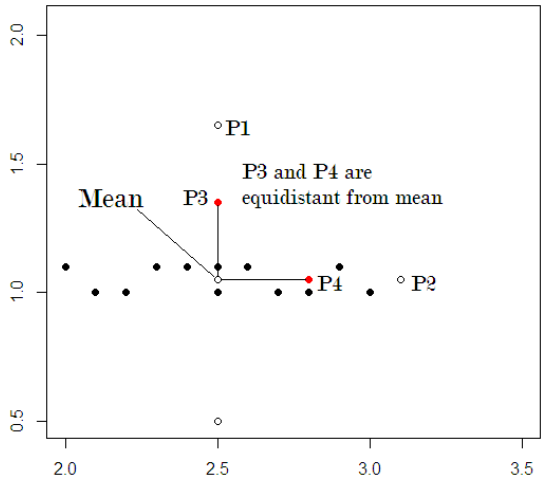

Figure 2. Point P3 falls outside the distribution and the Point P4 falls inside.

Another illustration is shown in figure 2. Here, point $P3$ and $P4$ are considered. Clearly, $P4$ falls inside the distribution and $P3$ falls out side. The Mahalanobis distance states this fact clearly as the distance of $P3$ is 28.64 and that of $P4$ is 23.59. If a suitable cutoff distance can be set, then $P3$ can be excluded and the $P4$ can be included which is not possible in case of traditional similarity measures.

The present model exploits this property of Mahalanobis distance in the neighborhood of each point for calculating the neighbors on the basis of distribution similarity and not on the basis of distance similarity. The model assumes the number of clusters as a prerequisite. This is k in the present case. The model also assumes some knowledge, as in case of other clustering algorithms, as parameter value for the creation of similarity matrix.

Initially, the algorithms has been presented and then the working principle of the proposed model has been discussed explicitely. The algorithm for Mahalanobis distance based similarity matrix computation has been proposed in Algorithm 1. The algorithm for clustering model is shown in the Algorithm 2 .

Algorithm 1 Similarity Matrix Computation

Input: $S = \left\{\vec{P}_1, \vec{P}_2, \ldots, \vec{P}_n\right\}$ ▷ m-dimensional Data Points to be clustered.
Input: ϵ ▷ The Benchmark Distance value for Euclidean distance
Input: η ▷ The Benchmark Distance value for Mahalanobis distance

declare $N_{m\times m}$, $M_{m\times m}$ **as matrix**
$N \leftarrow$ SIMILARITYMATRIX(S, ϵ) ▷ Calculates the ϵ-neighbor similarity matrix in the traditional way.
for all $\vec{P}_i \in S$ **do**
  declare A **as set**
  $A \leftarrow$ NEIGHBORHOODOF$(\vec{P}_i, N)$ ▷ Calculates The neighborhood of $\vec{P}_i$ w.r.t. the already computed N.
  matrix $\Gamma \leftarrow$ COVARIANCEMATRIX(A)
  for all $\vec{P}_j \in S$ **do**
    $d \leftarrow \left(\vec{P}_j{}^T \Gamma^{-1} \vec{P}_j\right)^{1/2}$
    if $d \leq \eta$ **then**
      $M^{(i)(j)} \leftarrow 1/d$
      $M^{(j)(i)} \leftarrow 1/d$
    else
      $M^{(i)(j)} \leftarrow 0$
      $M^{(j)(i)} \leftarrow 0$
    end if
  end for
end for

Algorithm 2 Spectral Clustering using Mahalanobis Distance

Input: $S = \left\{\vec{P}_1, \vec{P}_2, \ldots, \vec{P}_n\right\}$ ▷ m-dimensional Data Points to be clustered.
Input: K ▷ The number of clusters
1: **declare** $M_{m\times m}$ **as matrix**
2: $M \leftarrow$ MAHALANOBISSIMILARITYMATRIX(S, ϵ) ▷ Calculates the Mahalanobis distance based similarity matrix using Algorithm 1.
3: $L_{rw} \leftarrow$ RANDOMWALKLAPLACIAN(M)
4: CALCULATEEIGENS(L_{rw}) ▷ Calculates the eigenvalues and eigenvectors.
5: SORTEIGENVALUES ▷ Sorts the eigenvalues in ascending order.
6: $D \leftarrow$ SMALLESTEIGENVECTORS(k) ▷ D is the set of k eigenvectors for k smallest eigenvalues.
7: K-MEANS(D, k) ▷ Cluster k eigenvectors using K-means algorithm.

From Algorithm 1 and Algorithm 2, it is clear, that the main tricky part is the formation of the similarity matrix or the affinity matrix. Let us discuss the working principle of this part first. What the algorithm 1 is doing is that it is calculating the final similarity matrix or adjacency matrix in two passes. The first pass calculates the similarity matrix or the adjacency matrix using any traditional method. In the present paper, ϵ-neighbor has been adapted but other similarity matrices may also be chosen. Thereafter every point and the neighborhood of them are being computed for the second pass or final adjacency or similarity matrix. This is computation is very tricky. For a particular row, matrix elements with nonzero value is the vertices of the neighbor set. after having vertex set , Lets say A, the present paper computes the covariance matrix of the neighborhood set A. Proposed model considers the other points and the Mahalanobis distance of them. A second new similarity matrix is being created in this way. Here also any suitable similarity measure may be considered. In the present paper, ϵ-neighbor has been chosen. i.e. given the Mahalanobis distance of a point from a set of points, if the distance is less than ϵ, then the point is considered for being a member of the neighborhood of the point otherwise the point is considered as the outlier of the point set and is not considered to be a neighbor of the point set.

From the discussions, it is clear that the Mahalanobis distance based similarity matrix is more realistic. It may include (or reject) a point in neighborhood of point by not measuring the distance only but by considering the distribution also. This method uses the technique of outlier exclusion in micro level for creating affinity or similarity matrix.

Algorithm 2 is the actual clustering algorithm. The present paper assumes the number of clusters, i.e. k is already known. Then the present paper calculates random walk Laplacian of the Mahalanobis distance based similarity matrix. The reason for selecting the random walk Laplacian is that, it identifies dense subgraph in a better way and therefore gives a better result in practical. The proposed method finds the eigenvalue of the Laplacian of the matrix and sorts them in ascending order and the smallest k eigenvalues are identified and the corresponding eigenvectors are considered for k-means clustering and the result of this k-means clustering is the final output.

4. RESULTS AND ANALYSIS

The model has been tasted with two data sets. One is a toy data set proposed by the authors and the second data set is known as spiral data set proposed by Chang and Yeung[26].

The performance of the proposed model has been compared with three other models namely K-means algorithm, hierarchical clustering and one of the standard spectral clustering method. More about K-means and hierarchical clustering can be found in [2, 3]. The standard spectral model is due to Shi-Malik [27].

The toy example consists of 75 data points in a two dimensional plane. The data set has been taken in such a way that the trend of the points are clear. In the figure 3 (a), it is clear that the small portion in the right side of the lower cluster (indicated as C in the figure) is a part of the lower cluster and that the small middle cluster(indicated as B in the figure)is the part of the upper straight line. The trend clearly suggests that. The important thing to note that the small clusters are equidistant from the lower spiral cluster. i.e. portion B and portion C are equidistant from the lower spiral part of the figure 3 (a). Portion B is equidistant from lower spiral part and the upper straight line portion. So, the traditional distance or the k-nearest neighborhood is not at all a good measure, because none of these considers the trend of the data sets. Either they will include both of the small clusters or they will exclude both of them. The proposed Mahalanobis distance based similarity matrix is a very good option in such type of situations because it will include one and exclude the other based on the point distribution. The result of the proposed model is shown in figure 3 (b). The result shows that the proposed model clearly considers the distance as well as the cluster point distribution trends.

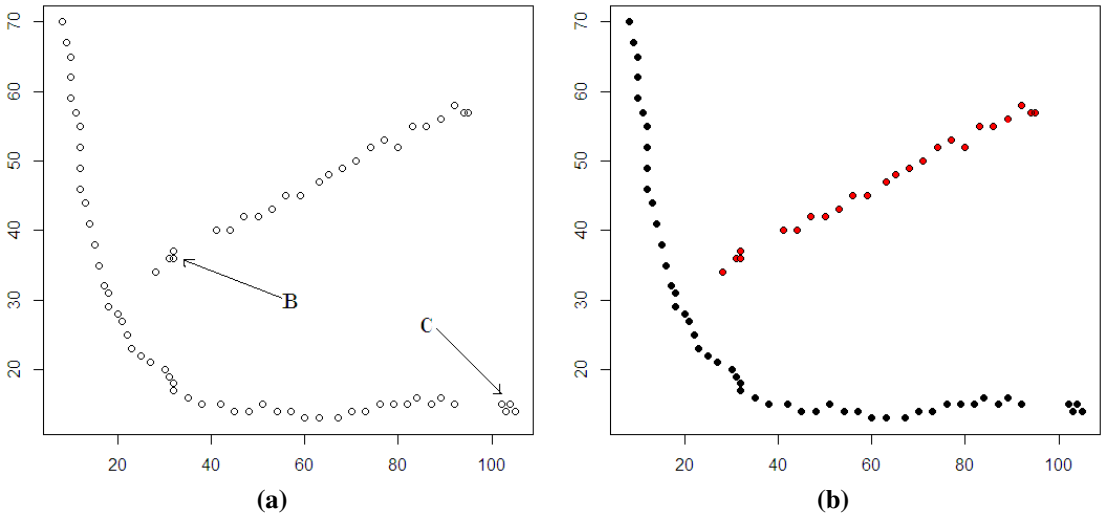

Figure 3. (a) The Sample toy data. (b) The black colors and red colors are the two clusters.

The proposed models performance is compared with the standard models and the results produced are tested by seven internal indices and four external indices. Table 3 and table 2 are the results of the tests. The proposed model shows better result in four out of seven indices. All the four external indices says the strength of the proposed model is better in the proposed situation.

Table 2. Internal Indices for proposed toy data

Internal Indices	Rule	On Synthetic Toy Data				Is the Proposed model best?
		K-Means Model	Hierarchical (Average Link)	Shi-Malik[27] Model	Proposed Model	
Silhouette	Max	0.4835682	0.4326811	0.2461643	0.2506375	No
Scott Symons	Min	800.4114	798.99	861.9192	729.9601	Yes
Wemmert Gancarski	Max	0.5940378	0.5102386	0.1872991	0.2446224	No
Tau	Max	-0.5383263	-0.4735096	-0.2174559	-0.1618588	Yes
Gamma	Max	-0.7610649	-0.6694097	-0.3078273	-0.2304259	Yes
G-Plus	Min	0.4403892	0.4174937	0.3262061	0.3034445	Yes
Ray Turi	Min	0.1788701	0.2293373	0.7925694	1.289401	No

Table 3. External indices for proposed toy data

Internal Indices	Rule	On Synthetic Toy Data				Which One is Best?
		K-Means Model	Hierarchical (Average) Model	Shi-Malik[27] Model	Proposed Model	
Folkes Mallows	Max	0.558085	0.6511637	0.9060712	1	Proposed
Jaccard	Max	0.386073	0.4815563	0.8275653	1	Proposed
Rand	Max	0.532973	0.6302703	0.8976577	1	Proposed
Czekanowski Dice	Max	0.5570745	0.6500682	0.9056478	1	Proposed

The model has been tasted with the another standard data set, known as spiral data, proposed by Chang et al[26]. The data set has 312 data points and 2 dimensions and 3 clusters. The data set is not linearly separable. The proposed model correctly identifies the three clusters of the spiral data with accurate point distribution. i.e. Both path based model[26] and the present model gives 100% accuracy. This shows the strength of the proposed model on

standard situations.

5. CONCLUSION

The proposed Mahalanobis distance based local distribution oriented spectral clustering model is a good model in normal situations as well as it can handle the situations where the distribution of the points needs to be considered. The result shows that the model can handle non-convex data set successfully which is a major strength of the proposed model. Nevertheless, there are scope of improvements also. The model is time demanding because of the eigenvector computations. So, the sparsification of the graph and clustering the sparse graph instead of the original one may be one improvement direction. Secondly, if the inverse of the covariance matrix does not exists, then the Mahalanobis distance cannot be calculated. Handling this will be a major improvement. Finally, model is sensitive to parameters like ϵ in similarity matrix calculation. The automatic calculation of such parameters can make the model even more robust.

REFERENCES

[1] G. Schaefer and H. Zhou, "Fuzzy clustering for colour reduction in images," *Telecommunication Systems*, vol. 40, no. 1-2, pp. 17–25, 2009.

[2] B. S. Everitt, D. Stahl, M. Leese, and S. Landau, *Cluster Analysis*, 5th ed. John Wiley & Sons, 2011.

[3] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999.

[4] R. Xu and D. Wunsch-II, "Survey of clustering algorithms," *IEEE Transactions on Neural Network*, vol. 16, no. 3, pp. 645–678, May 2005.

[5] F. Farshchian, E. Parcham, and S. Tofighi, "Retinal identification and connecting this information to electronic health record," *International Journal of Electrical and Computer Engineering*, vol. 3, no. 3, pp. 359–365, June 2013.

[6] S. Sharma and G. N. Purohit, "Analysis of spectral clustering approach for tracking community formation in social network," *International Journal of Social Networking and Virtual Communities*, vol. 1, no. 1, pp. 31–37, July 2012.

[7] X. Yang and W. Cui, "A novel spatial clustering algorithm based on delaunay triangulation," *Journal of Software Engineering and Applications*, vol. 3, pp. 141–149, 2010.

[8] P. Roy and J. K. Mandal, "A novel spatial fuzzy clustering using delaunay triangulation for large scale gis data (nsfcdt)," *Procedia Technology*, vol. 6, no. 452459, 2012.

[9] P. Foggia, G. Percannella, C. Sansone, and M. Vento, "A graph-based clustering method and its applications," *Proceedings of Advances in Brain, Vision, and Artificial Intelligence*, vol. 4729, pp. 277–287, 2007.

[10] P. Roy and J. K. Mandal, "A delaunay triangulation preprocessing based fuzzy-encroachment graph clustering for large scale gis data," *Proceedings of the International Symposium on Electronic System Design, 2012*, pp. 300–305, 2012.

[11] T. Asano, B. Bhattacharya, M. Keil, and F. Yao, "Clustering algorithms based on minimum and maximum spanning trees," *Proceedings of the 4th Annual Symposium on Computational Geometry*, pp. 252–257, 1988.

[12] D. J. Higham, G. Kalna, and M. Kibble, "Spectral clustering and its use in bioinformatics," *Journal of Computational and Applied Mathematics*, vol. 204, pp. 25–37, 2007.

[13] S. A. Toussi and H. S. Yazdi, "Feature selection in spectral clustering," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 4, no. 3, pp. 179–194, September 2011.

[14] H. Qiu and E. R. Hancock, "Graph matching and clustering using spectral partitions," *Pattern Recognition*, vol. 39, pp. 22–34, 2006.

[15] U. von Luxburg, "A tutorial on spectral clustering," Max Planck Institute for Biological Cybernetics, Tech. Rep. TR-149, August 2006.

[16] F. Chung, "Spectral graph theory," American Mathematical Society,USA, Tech. Rep., 1997.

[17] D. A. Spielman and S.-H. Teng, "Spectral partitioning works: Planar graphs and finite element meshes," *Linear Algebra and its Applications*, vol. 421, p. 284305, 2007.

[18] M. Fiedler, "A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory," *Czechoslovak Mathematical Journal*, vol. 25, no. 4, pp. 619–633, 1975.

[19] P. C. Mahalanobis, "On the generalized distance in statistics," in *Proceedings of the National institute of Sciences of India*, no. 2, 1936, pp. 49–55.

[20] S. Marsland, *Machine Learning, An Algorithmic Perspective*, 1st ed., ser. Machine Learning and Pattern Recognition Series. CRC Press, 2009.

[21] V. J. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artificial Intelligence Review*, vol. 22, pp. 85–126, 2004.
[22] S. Saha and S. Bandyopadhyay, "Performance evaluation of some symmetry-based cluster validity indexes," *IEEE Transactions on Systems, Man, and CyberneticsPart C:Applications AND Reviews*, vol. 39, no. 4, pp. 420–425, July 2009.
[23] C. Godsil and G. Royle, *Algebraic Graph Theory*, ser. Graduate Texts in Mathematics. Springer, 2001, vol. 207.
[24] B. Mohar, *Graph Theory, Combinatorics, and Applications*. Willy, 1991, vol. 2, ch. Laplacian Spectrum of Graphs, pp. 871–898.
[25] ——, *Graph Symmetry: Algebraic Methods and Aplications*, ser. C 497. Kluwer, 1997, vol. NATO ASI, ch. Some Applications of Laplace eigenvalues of graphs, pp. 225–275.
[26] H. Chang and D.-Y. Yeung, "Robust path-based spectral clustering," *Pattern Recognition*, vol. 41, no. 1, pp. 191–203, January 2008.
[27] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 22, pp. 888–905, August 2000.