Clustering and its application to GIS

Parthajit Roy
Department of Computer Science,
The University of Burdwan, Burdwan-713104

Abstract
Pattern recognition is an emerging field in the modern computations. The modern world is gradually moving towards the generic solutions rather than specific solution to a problem. For this the recognition of the pattern of the problem is very important. On the other hand, learning is also pattern recognition. Clustering is a field of pattern recognition or pattern classification where the leveled data are not available. i.e. the unsupervised type pattern classification is clustering. This is more challenging because the domain knowledge is not available in clustering. This article is a survey of the clustering techniques that have been explored by the author and the application of them in the field of geographic information systems.

Keywords
Clustering — Pattern Recognition — Geographic Information System

1. Introduction

Pattern Recognition or Pattern Classification an emerging and promising field in the modern day computing and in the field of Pattern Recognition, clustering is one of the interesting and difficult task. Clustering means unsupervised grouping of objects where labeled data for training is not available [1]. This restriction has made the clustering more challenging. Clustering mainly deals with the topology of the data set or proximity of the data set or some kind of density or distribution measures. Learning by these tools only is a major focus area of data clustering.

There are five major steps of data clustering proposed by Jain et al [1]. These are

- Feature extraction
- Introduction of the idea of proximity
- Clustering of grouping
- Data abstraction
- Validation of cluster

Given n objects $\{o_1, o_2, \cdots, o_n\}$ to be clustered, the first thing is to extract the features that represents an object. If feature vectors are m-attribute or m dimensional, then objetcs are represented by $S = p_1, p_2, p_3, \cdots, p_n \forall p_i \in R^m$. After the features are extracted, the notion of proximity is applied. There are different distance measures like Mahalanobis distance, Minkowski distance etc. available but the most frequently used distance measure is Euclidean distance. Having the feature vectors and the distance in hand, the objective of the clustering is to find out a partition matrix $P(S)$ of order $C \times n$ representing the clustering into C number of clusters where each cell of the matrix $u_{ij}, \forall i = 1, \cdots, n \ \ and j = 1, \cdots, C$ has a value representing the degree of membership of i^{th} object to the j^{th} cluster [2].

Duda et al [3] discussed a solid foundation of classification whereas Everitt et al [4] explored the clustering from theoretical aspects. An excellent survey on clustering up to 1999 is done by Jain et al [1]. An outstanding survey on clustering algorithms is done by Xu and Wunsch II [5].

Geographical Information System (GIS), on the other hand, is another emerging field [6] [7]. Clustering is an integrated part of a GIS system. The application of clustering in GIS ranges from soil type grouping to crop clustering and many more. Clustering in GIS is even more challenging because GIS deals with huge data. So, the clustering as well as data structure for storing the knowledge about cluster is one of the major concerns of literally every GIS system. This report is a survey of clustering and its application to geographic information systems.

2. Proposed Methods

The present article covers fuzzy clustering techniques geometric and graph based clustering techniques. First of all a novel variation of traditional fuzzy C-means algorithm has been proposed. This is named as selective scale space based fuzzy C-means technique. The second method is a geometric clustering based on Delaunay triangulation. This method has applied parzen window technique for fine tuning. The third one is a variation of minimum spanning tree based clustering. Following subsections discusses them in brief along with the proosed algorithms.

2.1 Selective Scale Space based FCM

The of this technique is to cluster n number of objects whose representative feature vectors are $S = \{p_1, p_2, p_3, \cdots, p_n\} \forall p_i \in R^m$. First of all we are going to present the standard fuzzy C-means algorithm. Here it is assumed that there are C number of clusters (A *priori* which is essential for FCM). Given such information, the standard Fuzzy C-Means technique is stated using algorithm 1.

Algorithm 1 Standard Fuzzy C-Means Algorithm

Input: $S = \{p_1, p_2, \cdots, p_n\}$ ▷ Points to be clustered.
Output: $C = \{k_1, k_2, \cdots, k_C\}$ ▷ The final cluster centers.

declare $F_{n\times C}S$ **as** *Matrix*
declare $ObjectiveValue$ **as** *Real*
INITIALIZECLUSTERCENTERS(C)
while $ObjectiveValue \leq Benchmark$ **do**
  POPULATEFUZZYPARTITIONMATRIX($F_{n\times C}, C, S$)
  UPDATECLUSTERCENTERS($C, F_{n\times C}$)
  $ObjectiveValue \leftarrow$ OBJECTIVEFUNCTION($C, F_{n\times C}$)
end while

The fuzzy membership should be such that

$$\sum_{k=1}^{n}\sum_{i=1}^{C} \mu_{ik} = n$$

with $0 \leq \mu_{ik} \leq 1 \forall i = 1,2,\cdots,C \ \ and \ \ \forall j = 1,2,\cdots n$, $0 < \sum_{k=1}^{n} \mu_{ik} < 1$ and $\sum_{i=1}^{C} \mu_{ik} = 1$

There is a need of validity index to check the quality of the clusters. In the present model this is measured by Xie-Beni index [8]. This index is a function of Variation v and separation of the clusters centers. Variation v can be stated using equation 1 [9].

$$v(P,C,X) = \sum_{i=1}^{C}\sum_{j=1}^{n} \mu_{ij}^2 E^2(c_i, x_j) \tag{1}$$

where c_i are cluster centers and x_j are feature vectors. The separation measure can be stated using equation 2.

$$d(C) = min_{i\neq j} E^2(c_i - c_j) \tag{2}$$

where the $E(.,.)$ ia the Euclidean norm and finally the Xie-Beni index (XB) can be given using equation 3.

$$XB(P,C)_X = \frac{v_X(P,C)}{nd(C)} = \frac{\sum_{i=1}^{C}\sum_{j=1}^{n} \mu_{ij}^2 E^2(c_i, x_j)}{nmin_{i\neq j} E^2(c_i - c_j)} \tag{3}$$

The Xie-Beni index measures compactness of a single cluster and separability of different clusters. The value of fuzzy membership functions for fuzzy C-menas algorithm is given by the equation 4 [9].

$$\mu_{i,k} = \frac{1}{\sum_{j=1}^{C}\left(\frac{E(c_i - x_k)}{E(c_j - x_k)}\right)^{\frac{2}{m-1}}}, 1 \leq i \leq C, 1 \leq k \leq n \tag{4}$$

The Center update rule [9] is done through the equation 5

$$c_i = \frac{\sum_{k=1}^{n} (\mu_{ik})^m x_k}{\sum_{k=1}^{n} (u_{ik})^m} \tag{5}$$

In the proposed model scale space technique has been applied. This technique works good if the clusters are linearly non-separable. The idea of scale space filter is that some transformation will be applied the parameters so that the clusters will become linearly separable. Some previous work on clustering using scale space filtering can be found in [10].

Scale Space has a tendency to make far points even further. This may destroy the proximity of data sets. For this, a novel Standard Deviation(S.D.) based selective model has been proposed. The scale space is applied to those parameters where the variability is not good. The low standard deviation means not good variability. The scale space has been applied to those parameters only. The parameters having high standard deviation remains untouched. The final model after all kind of preprocessing and internal reorientation has been proposed in the algorithm 2.

Algorithm 2 S.D. based Selective Scale Spaced FCM

Input: $S = \{p_1, p_2, \cdots, p_n\}, sdThreshold$
Output: $C = \{k_1, k_2, \cdots, k_C\}$ ▷ The final cluster centers.

declare $F_{n\times C}S$ **as** *Matrix*
declare $ObjValue$ **as** *Real*
NORMALIZEDATASET(S)
for $All \ \ parameter[i] \ \ of \ \ dataset \ \ S$ **do**
  $sd \leftarrow$ GETSTANDARDDEVIATION($param_i$)
  if $sd < sdThreshold$ **then**
    $p_i \leftarrow$ GAUSSIANSCALESPACE(p_i)
  end if
end for
INITIALIZECLUSTERCENTERS(C)
while $ObjValue \leq Benchmark$ **do**
  POPULATEFUZZYPARTITIONMATRIX($F_{n\times C}, C, S$)
  UPDATECLUSTERCENTERS($C, F_{n\times C}$)
  $ObjValue \leftarrow$ XIEBENIOBJFUNCTION($C, F_{n\times C}$)
end while

2.2 Delaunay Triangulation based clustering

This model is based on a variation of computational geometric model known as Delaunay triangulation [11]. Given a set of representative feature vectors $P = \{p_1, p_2, p_3, \ldots, p_n\}$ in a two dimensional space, The Delaunay Triangulation is a planar decomposition of the point set. The notion of proximity in a Delaunay triangle has been adopted for the first phase of clustering in the present model. The algorithm is given below.

The second phase of this model introduced a fuzzy refinement to the clusters. The Fuzzy based final technique NSFCDT is illustrated in Algorithm 4.

2.3 Spanning Tree based Clustering

This model is a minimum spanning tree based model. The edges of the graph will have weight and this weight is, in our case, Euclidian distance of the two nodes in the two dimensional plane. The edges between two data point gives the notion of proximity in our present model.

Algorithm 3 Procedure Delaunay Classification

```
Input: PointSet = {p_1, p_2, ..., p_n}
    Declare DT As DelaunayTriangleSet
    DT ← DELAUNAYTRIANGLES(PointSet)
    Declare ClusterSet As Set
    Declare NewSet As Set
    Declare TmpSet As Set
    for all dt In DT do
        if VISITED(dt) = False then
            INITIALIZE(NewSet)
            PUSH(dt, Stack)
            while EMPTY(Stack) = False do
                Triangle T ← POP(Stack)
                if SATISFIABLE(T)&NOTVISITED(T) then
                    T.Visited ← True
                    NewSet ← NewSet ∪ {T}
                    TmpSet ← NEIGHBORS(T)
                    for all Element In TmpSet do
                        if NOTVISITED(Element) then
                            PUSH(Element, Stack)
                        end if
                    end for
                end if
            end while
            ClusterSet ← ClusterSet ∪ NewSet
        end if
    end for
```

Algorithm 4 Procedure Fuzzy Based NSFCDT

```
Input: PointSet = {p_1, p_2, ..., p_n}
BUILD( ClusterSet)                          ▷ By Algorithm 3
for all C_i In ClusterSet do
    Area ← GETAREA(C_i)
    Degree ← GETFUZZYDEGREE(Area)
    SETSATISFIABILITYCONDITION(Degree)
    for all Triangle In Cluster do
        N ← Neighbour(Triangle)
        for all E In N do
            if FUZZYSATISFIABLE(E) then
                if E ∈ C_j, ∀j = 1...m & j ≠ i then
                    C_i ← C_i ∪ C_j
                else
                    C_i ← C_i ∪ {E}
                end if
            end if
        end for
    end for
end for
```

As the volume of GIS data is very large, the computation of MST should be very efficient. Delaunay triangulation is a planar decomposition and for planar graph, the upper bound of number of edges is

$$E(G) \leq max\left\{\frac{g}{g-2}(n-2), n-1\right\} \tag{6}$$

where g is the grith of the graph [12]. As the grith of DT is $g = 3$. Putting the value of g in equation 6 we get, $E(V) = 3n-6$. i.e. the edges are of $O(n)$. instead of $O(n^2)$ in case of complete graph K_n. Also, if the distance is Euclidean, the MST of a complete graph is the MST of a Delaunay graph. This suggests that a Delaunay based MST is efficient than bare MST. Also there is efficient algorithm due to Bowyer [13] and Watson [14] for finding DT of a point set with time complexity $O(n\sqrt{n})$. The final time complexity of the algorithm will be $O(n\sqrt{n}) + O(n \lg n) = O(n\sqrt{n})$ instead of $O(n^2 \lg n)$ in case of complete graph.

The algorithm can be stated as follows.

1 Scan The Image and identify the pixels qualified for being a part of the forest and let that set is V and let $|V| = n$.

2 Compute the Delaunay Triangulation of the given point set V.

3 Compute the minimum spanning tree of the complete graph having V as the vertex set. (We have to calculate the MST of the DT because the MST of the Complete Graph is the MST of the DT.)

4 Consider a benchmark distance.

5 Start from a vertex and keep including vertices in a single cluster as long as the edge length is less than benchmark value.

6 If an edge crosses the benchmark distance, then create a separate cluster and start putting the new vertices into the new cluster.

7 continue this process until all the points of V is not processed.

The model further proposed fuzzy post processing known as fuzzy encroachment. The idea is, if a cluster is very large then the nearby cluster is not a separate cluster. instead it is a part of the larger one and the larger cluster will encroach the smaller one. The final fizzy encroachment algorithm can be stated as follows.

- Take the cluster sets C, produced by the SMT-based algorithm.
- For all $c \in C$ do the following:
  - Calculate the size of the cluster

- If the size is large, then
  * Encroach the nearby clusters if any, depending upon the fuzzy strength.
  * Else make no change to the cluster.
- After the completion of the above process, discard all negligible clusters and produce only fair clusters.

3. Results and Analysis

The clustering algorithms are applied to benchmark data, random data and GIS data. The results are discussed below.

3.1 Result of Selective Scale Space based Model

The Result of pure FCM for iris data set [15] is given below.

Table 1. The Result of Pure FCM

Class Name	Actual No.	Pure FCM Correct Pred.	Pure FCM Wrong Pred.	%of Error
Setosa	50	50	0	0%
Versicolor	50	47	13	19%
Verginica	50	37	3	29%

The Result of scale space applied to all parameters for iris data set [15] are shown below.

Table 2. The Result of Scale Space FCM

Class Name	Actual No.	Pure FCM Correct Pred.	Pure FCM Wrong Pred.	%of Error
Setosa	50	50	4	4%
Versicolor	50	49	0	2%
Verginica	50	47	0	6%

The Result of scale space applied to selective parameters for iris data set [15] are shown below.

Table 3. The Result of Standard Deviation based Scale Spaced FCM

Class Name	Actual No.	Pure FCM Correct Pred.	Pure FCM Wrong Pred.	%of Error
Setosa	50	50	0	0%
Versicolor	50	50	3	3%
Verginica	50	47	0	6%

The following table is a comparative study of the variations. This clearly shows the superiority of the selective scale space based model than the other models.

Table 4. The Comparative Study of Models on Iris Data

Model Name	Correct Prediction	wrong Prediction
Pure FCM	89.3%	10.7%
Scale Space FCM	97.3%	2.7
SD Scale Space FCM	**98%**	**2%**

3.2 Result of Delaunay Triangulation based Clustering

The result of proposed NSFCDT is given below. We have compared NSFCDT with the NSCABDT [16]. The Scheme has been tested with a series of data sets consisting of 5000, 6000, 7000, 8000, 9000 and 10000 random points. In each and every case, the proposed NSFCDT shows reasonably good and comparable results. NSFCDT shows better performance when the number of clusters is more (more then 5). The result of a typical data set is shown for NSFCDT in grafical form in Figure 1(a),1(b), 3(a) and 3(b). In Figure 1(a), a data set of 4088 points has been taken. Figure 1(b) shows the Delaunay Triangulation of the data set. In Figure 3(a), the Graphical result of crisp clustering is shown and in the Figure 3(b) the final fuzzy output of the second pass of the model is shown. Figure 3(b) shows, that though crisply there are four clusters, the fuzzy model produces three instead of four clusters as it seems from the Figure 1(a).

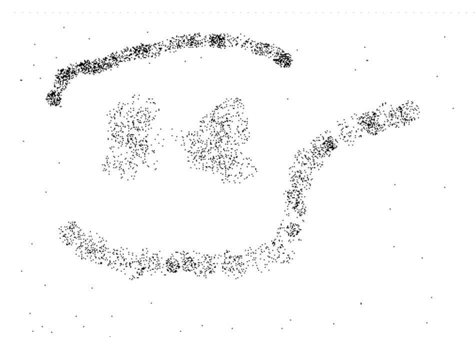

Figure 1. A Set of 4088 data points in a spatial domain.

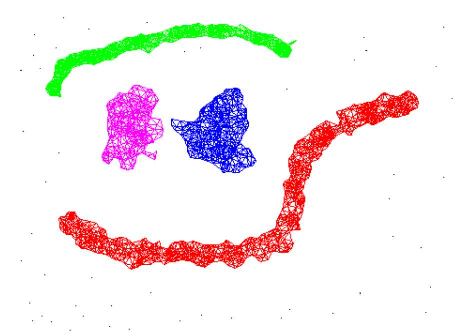

Figure 2. Crisp Clustering produces four different clusters

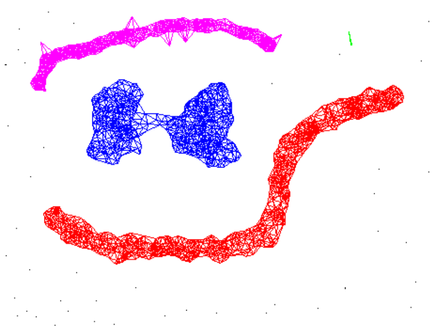

Figure 3. Fuzzy Clustering produces three different clusters.

The comparative study between NSFCDT and NSCABDT for various data sets are presented in the table 5. The table clearly shows that the NSFCDT is superior in crisp clustering. As we have proposed a second pass for fuzzy refinement, there is slight overhead and the time complexity is slightly high. The interesting point is that, the little time overhead of the present model offering a fuzzy rationality which models the real life clustering better.

Table 5. A Comparative Study of NSCABDT and NSFCDT

No of Points	Run Time in Seconds		
	NSCABDT	NSFCDT	
		Crisp Clustering	Fuzzy Clustering
	Run Time	Run Time	Run Time
5000	48.2	40.04	55.18
6000	71.4	54.72	81.99
7000	93.8	63.19	87.45
8000	117.9	97.37	135.82
9000	151.3	129.48	183.88
10000	189.4	167.73	228.31

3.3 Result of Spanning tree based Clustering

The model has also been applied on a set of 5000, 6000, 7000, 8000, 9000 and 10000 points and the results have been compared with Yang et al. [16]. Instead of the post processing proposed in NSCABDT, the present paper proposed a novel fuzzy encroachment technique. The results are neck to neck and as the present model proposes fuzzy refinement, a better rationality can be achieved and also the quality measures like validity, stability of clusters can also be achieved by tuning the fuzzy parameters. The comparison of the proposed model with NSCABDT is given in the figure 4.

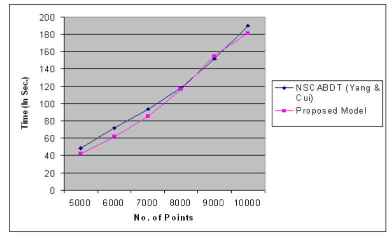

Figure 4. Comparison of the proposed model with NSCABDT [16]

The time-space complexity of the proposed model and the time-space complexity of the model (based on Complete graph) proposed by Foggia et al. [17] is given in the table 6.

Table 6. Complexity measures of Complete Graph based model and the Present model

Attribute	Complete-Graph Model	Present Model
Time Complexity	$O(n^2 \lg n)$	$O(n\sqrt{n})$
Space Complexity	$O(n^2)$	$O(n)$

The proposed NSFCDT model has also been applied on GIS data taken from Remote Sensing image of LANDSAT-7 (*Courtesy of Bidhan Chandra Krishi Viswavidyalaya, West Bengal, India*). A very satisfactory results are found on application of the techniques. A part of Sundarban area of West Bengal having number of rivers and a very dense forest, the model shows a very good result on a gray scale Remote Image. The Crisp classification shows several tree set area (Six large clusters). Even a very low aperture canal separates two clusters distinctly (The Green color and Red color clusters in fig 7(a)). For Blue and Cyan cluster of fig 7(a), a small area having shallow distribution separates the clusters. Both the cases are clearly not realistic (clear from fig 5). The fuzzy inclusion gives a rational classification of the tree set areas and it is established from the algorithm that there are broadly four clusters instead of six, suggested by the crisp clustering. Figure 5, 7(a) and 7(b) shows the results.

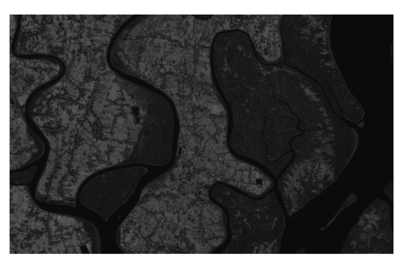

Figure 5. A portion of Sundarban area of West Bengal, India.(*Courtesy of BCKV, West Bengal, India.*)

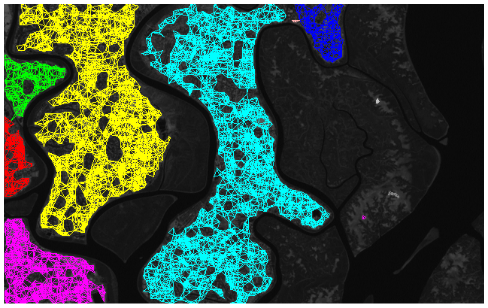

Figure 6. Crisp Clustering produces six main clusters. But clusters in Blue and Cyan are essentially same. The same is true for Green and Red Clusters also.

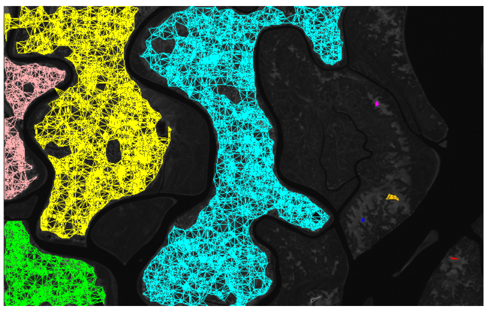

Figure 7. Fuzzy Clustering produces four different clusters. The Cyan and Blue are merged to form a single cluster and so also Red and Green and they have formd a single cluster shown in Pink.

3.4 Application Spanning Tree Model to GIS

The spanning tree based proposed model has been applied on a 640×400 remote gray scale image [Fig 8] of Sundarvan, West Bengal, India. The results are quite satisfactory. In the crisp clustering the result is showing several small clusters [Fig 11], which are not rational because from the figure 8, it is clear that they should form a single cluster. After running the fuzzy encroachment process, the clustering has become even more rational and ends up with five clusters[Fig 12]. The intermediate triangulation and minimum spanning tree generation are also shown in figure 9 and figure 10 respectively.

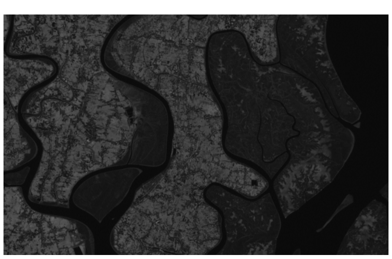

Figure 8. A portion of the Sundarvan Area of West Bengal, India (Gray Scale Image).

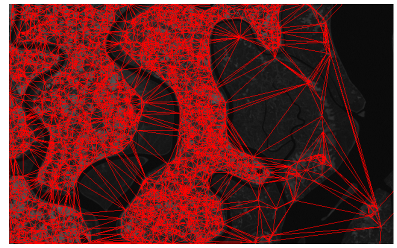

Figure 9. The Delaunay Triangulation of the attribute points.

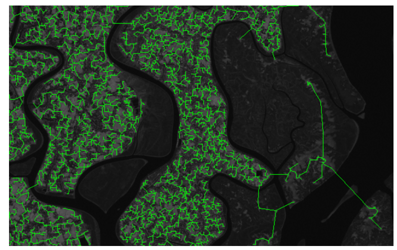

Figure 10. The minimum spanning tree generated from the Delaunay Triangulation of the attribute points.

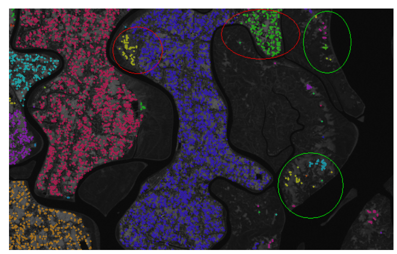

Figure 11. The initial minimum spanning tree based clustering. The model creates clusters correctly but takes benchmark distance for all of the clusters same. That is why it creates some extra clusters which are actually the part of the larger one. The red color circles shows that.

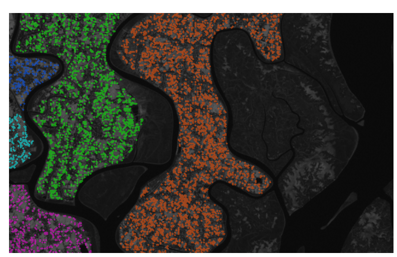

Figure 12. Final output of the proposed model. The model produces only five clusters by the method of fuzzy encroachment, which looks realistic. It also discards negligible clusters shown in green circles in figure 11.

4. Conclusion and Future Scope

The clustering is a very important and effective thing in the field of GIS. The main aim of clustering is to group data sets. The grouping of data is a fundamental part of every GIS systems. Different types of clustering techniques have been proposed. The result of clustering is very promising in the generic data set as well as the standard benchmark data sets. The GIS results are also promising. Nevertheless, there are scope of improvements. The newer area of GIS where the clustering can be applied as well as the newer clustering techniques to GIS can be proposed. The use of more efficient data structures can improve the present models also.

List of Publications

1. Parthajit Roy and J. K. Mandal, "A Novel Fuzzy-GIS Model Based on Delaunay Triangulation to Forecast Facility Locations (FGISFFL)" *International Symposium on Electronic System Design (ISED), 2011, Index at IEEE Xplore, DOI:10.1109/ISED.2011.48*, pp. 341–346, 2011.

2. Parthajit Roy and J. K. Mandal, "A Relaxed Parzen Window Based Multifeatured Fuzzy-GIS Model to Forecast Facility Locations (RPWMFGISFFL)" *Proceedings of the International Conference on Information Systems Design and Intelligent Applications 2012 (INDIA 2012), (Published in Advances in Intelligent and Soft Computing, 2012 of Springer, DOI:10.1007/978-3-642-27443-5)*, vol. 132/2012, pp. 623-630, 2012.

3. Parthajit Roy and J. K. Mandal, "A Novel Spatial Fuzzy Clustering using Delaunay Triangulation for Large Scale GIS Data (NSFCDT)" *PProceedings of the 2nd International Conference on Communication, Computing & Security [ICCCS-2012], (Published in: Procedia Technology of Elsevier, DOI:10.1016/j.protcy.2012.10.054*, vol. 6, pp. 452-459, 2012.

4. Parthajit Roy and J. K. Mandal, "A Novel Spatial Fuzzy Clustering using Delaunay Triangulation for Large Scale GIS Data (NSFCDT)" *Proceedings of the 2nd International Conference on Communication, Computing & Security [ICCCS-2012], (Published in: Procedia Technology of Elsevier, DOI:10.1016/j.protcy.2012.10.054*, vol. 6, pp. 452-459, 2012.

5. Swati Adhikari, J. K. Mandal and Parthajit Roy, "On The Strength of Self Organizing Feature Map for Clustering Non-Categorical Data" *Proceedings of First International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013*, pp. 89-93, September, 2013.

6. Parthajit Roy And J. K. Mandal, "A Novel Selective Scale Space based Fuzzy C-means Model for Spatial Clustering" *Proceedings of Second Internal Conference on Computing and Systems-2013, (Published in: Procedia Technology of Elsevier, DOI:10.1016/j.protcy.2013.12.400)*, vol. 10 pp. 596-603, September, 2013.

References

[1] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: A review. *ACM Computing Surveys*, 31(3):264–323, 1999.

[2] Ujjwal Maulik and Sanghamitra Bandyopadhyay. Performance evaluation of some clustering algorithms and validity indices. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 24(12):1650–1654, December 2002.

[3] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. John Wiley, 2 edition, 2008.

[4] Brian S. Everitt, Daniel Stahl, Morven Leese, and Sabine Landau. *Cluster Analysis*. John Wiley & Sons, 5 edition, 2011.

[5] Rui Xu and Donald Wunsch II. Survey of clustering algorithms. *IEEE Transactions on Neural Network*, 16(3):645–678, May 2005.

[6] C. P. Lo and Albert K.W. Yeung. *Concepts and Techniques of Geographic Information Systems*. Prentice Hall of Iindia, 2 edition, 2007.

[7] T.M. Lillesand, R.W. Kiefer, and J.W. Chipman. *Remote Sensing and Image Interpretation*. Wiley India, 5 edition, 2004.

[8] X. L. Xie and G. Beni. A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analisys and Machine Intelligence*, 13(8):841–847, August 1991.

[9] Ujjwal Maulik and Sanghamitra Bandyopadhyay. Fuzzy partitioning using a real-coded variable-length genetic algorithm for pixel classification. *IEEE Transactions on Geoscience and Remote Sensing*, 41(5):1075–1081, May 2003.

[10] Yee Leung, Jiang-She Zhang, and Zong-Ben Xu. Clustering by scale-space filtering. *IEEE Transactions on Pattern Analisys and Machine Intelligence*, 22(12):1396–1410, December 2000.

[11] Mark de Berg, Otfried Cheong, Marc van Kreveld, and Mark Overmars. *Computational Geometry-Algorithms and Applications*. Springer, 3 edition, 2008.

[12] Béla Bollobás. *Modern Graph Theory*. Springer-Verlag, 2005.

[13] A. Bowyer. Computing dirichlet tessalations. *The Computer Journal*, 24(2):162–166, 1981.

[14] D.F. Watson. Computing the n-dimensional Delaunay Tessalation with application to Voronoi polytops. *The Computer Journal*, 24(2):167–172, 1981.

[15] R. A. Fisher. UCI machine learning repository. 1936.

[16] Xiankun Yang and Weihong Cui. A novel spatial clustering algorithm based on delaunay triangulation. *Journal of Software Engineering and Applications*, 3:141–149, 2010.

[17] Pasquale Foggia, Gennaro Percannella, Carlo Sansone, and Mario Vento. A graph-based clustering method and its applications. *Proceedings of Advances in Brain, Vision, and Artificial Intelligence*, 4729:277–287, 2007.