An Adaptive Masker for the Differential Evolution Algorithm

[1]Nassim Ammour, [2]Naif Alajlan, Member

[1,2] Member, IEEE ALISR Laboratory, College of Computer and Information Sciences, King Saud University
P.O. BOX. 51173, Riyadh 11543, Saudi Arabia

Abstract: *The automatic clustering problem of a large and complex data set into different homogeneous clusters is a challenging task. The choose of a variable length clusters centers using a good method to generate the maskers is an important phase in the evolution of a global search heuristics algorithms. In this paper, a new technique of real-coded modified differential evolution based automatic fuzzy clustering algorithm is proposed which automatically detects the optimal number of clusters and performs the proper partitioning from a data set. The effectiveness of the proposed technique is demonstrated by comparing it with other popular clustering algorithms. In the comparative study, a remote sensing data described in terms of feature vectors and a remote sensing image are used for the automatic partitioning and classifying tasks.*

Index Terms: *Differential evolution (DE), fuzzy clustering, genetic algorithm, remote sensing imagery, and masker.*

I.   Introduction

Clustering is the act of creating an optimal number of objects groups from a data set. each group called cluster gathers objects with a certain degree of similarity and which are dissimilar to the objects of the other groups. Clustering n complex real-world data into k nonempty clusters is a difficult task, and it is infeasible to exhaustively examine all possible clustering solutions. This difficulty emerges from the wide spectrum of cluster structures, highly unbalanced cluster sizes, varying cluster densities and possible cluster overlaps. Automatic clustering can be described as a system which receives an unlabeled data set as a unique input, with no a priori information about the number of clusters in the data, the system detects automatically the optimal number, and issues a labeled set of data producing the data partitions with a certain degree of homogeneity as output. The function of the system is to generate a random scenarios of the data partitioning with different number of clusters, and then evaluates each clustering structure using an objective function optimization as selection criterion. At the end of the clustering progress, the best solution found during all the evolution steps is selected.

The cluster analysis plays a central role in a variety of fields, engineering, computer sciences, life and medical sciences, earth sciences and economics. The clustering problem has been approached from diverse fields of knowledge, hierarchical algorithms [1], statistics [2], graph theory [3], expectation-maximization algorithms [4], artificial neural networks [5], evolutionary computing [6], and so on. Clustering a data set using evolutionary approach is viewed as an optimization problem and solved by using an evolutionary search heuristic such as a Genetic Algorithm (GA) [7] inspired by Darwinian evolution and genetics, and the Differential Evolution algorithm (DE) [8]. Differential Evolutionary algorithm uses a differential operator in the mutation process. In [13], Das et al proposed automatic crisp clustering using improved DE for the image segmentation application. A Modified Differential Evolution algorithm (MoDE) is proposed in [9], in this article three vectors are used in the mutation process, the global best, the local best and a randomly selected vector. This modified mutation process is governed by an adaptive parameter. The clustering process is based on the FCM algorithm and a randomly selected masker used for the clusters validation. In this paper, a new technique of evolutionary DE algorithm is provided, called *adaptive maskers for differential evolution* (AMDE), based on the fuzzy clustering method, and used for the automatic fuzzy clustering. The AMDE algorithm differs from the classic DE algorithms in the choice of the mask matrix process used for the centers validation. Classic DE algorithms use a randomly selected maskers matrix, this action produces random jumps in the choice of the centers to use in the clustering process, which provokes a deviation from the optimal solution direction and slows the convergence process. In the maskers choice process, the maskers matrix is divided into three zones, the first zone contains the best maskers found, the second zone contains a randomly selected maskers with numbers of centers in the vicinity of the global best solution number of centers, the third zone contains maskers with numbers of centers randomly selected. In order to increase the exploration and exploitation capabilities during the evolution process, the maskers matrix zones widths are governed by adaptive parameters. In order to show the efficiency of the proposed algorithm, a comparative study has been done with other fuzzy clustering algorithms, iterative fuzzy c-means (IFCM) [11], improved differential evolution (AFCIDE) [13], variable length genetic algorithm (VGAFC) [14] and a Modified Differential Evolution algorithm (MoDEAFC) [9]. Clustering results are reported for a numeric remote sensing data. The effectiveness of the proposed algorithm has been verified in

image classification. A satellite image of Kvarneric land features is used for demonstrating the effectiveness of the developed adaptive maskers for the differential evolution automatic based fuzzy clustering algorithm for image classification into unknown number of regions.

II.   Fuzzy C-Means Algorithm (FCM)

The fuzzy clustering technique (FCM) [10] generates and evolves the partition matrix $U \in R^{k \times n}$, the matrix of the fuzzy membership of each data points, using the dataset to cluster $X=\{x_1, x_2, ..., x_n\}$ and an evolutionary center set $C =\{c_1, c_2,..., c_k\}$ representing the clusters centers in order to divide the dataset and to identify the k dissimilar subsets regrouping similar patterns in some sense. The clustering FCM algorithm is a widely used technique that uses the fuzzy sets to evolve the partition matrix $U = [u_{i,j}]$. $1 \le i \le k; 1 \le j \le n$, while minimizing the measure,

$$J_m = \sum_{j=1}^{n} \sum_{i=1}^{k} u_{i,j}^m \, D^2(c_i, x_j) \;\; 1 \le m \le \infty \tag{1}$$

m denotes the fuzzy exponent. $D(c_i, x_j)$ represents the distance of point x_j from the kth cluster center c_i. Starting with random initial set C, the FCM algorithm finds, at each iteration, the fuzzy membership of each data points using

$$u_{i,j} = \frac{\left(\frac{1}{D(c_i, x_j)}\right)^{\frac{1}{m-1}}}{\sum_{i=1}^{k} \left(\frac{1}{D(c_i, x_j)}\right)^{\frac{1}{m-1}}}, \;\; for \; 1 \le i \le k; 1 \le j \le n. \tag{2}$$

At each iteration the centers set C is recomputed as follow:

$$c_i = \frac{\sum_{j=1}^{n} u_{i,j}^m x_j}{\sum_{j=1}^{n} u_{i,j}^m}, \qquad 1 \le i \le k. \tag{3}$$

The iteration process is stopped when there no further change in the cluster centers. At the final step, each data point is assigned to the cluster to which it has maximum membership. However, FCM has two major limitations, a priori number of clusters is required, and it often gets stuck at suboptimal solutions based on the initial conditions [11].

III.   Validity Measure

Cluster validation is very important issue in clustering analysis because the result of clustering needs to be validated in most applications. The validity of the clustering solution uses a fitness function, or an ensemble of fitness functions, in order to evaluate the degree of goodness of the presented clustering solution, generally based on the internal and external clustering validation.

The Xie-Beni (XB) index [12] is determined as a function of the ratio of the total variation σ to the minimum separation *sep* of the clusters. Here σ and *sep* are defined as,

$$\sigma(U, C, X) = \sum_{i=1}^{k} \sum_{j=1}^{n} u_{i,j}^2 \; D^2(c_i, x_j). \tag{4}$$

and,

$$sep(C) = \min_{i \ne j} \| c_i - c_j \|^2. \tag{5}$$

Where $\|.\|$ is the Euclidean norm. The XB index is then written as

$$XB(U, C, X) = \frac{\sigma(U, C, X)}{n. \, sep(C)}. \tag{6}$$

If the partition is compact and good, the value of σ should be low while the value of *sep* should be high, thereby yielding lower values of the XB index. Minimizing the XB index is therefore done to detect the proper clustering solution.

IV.   Proposed IDE Clustering Algorithm

A. *Differential Evolution improved Algorithms*

1. *Classical Differential Evolution Algorithm (DE):* For the optimization problems over continuous domains, Differential Evolution is a fast, robust, and efficient population-based heuristic algorithm used to explore and exploit the problem input data space and converge toward the global best existing solution. The chromosome, intermediate solution, is represented in a vector by real numbers. the initial population is generated as a random

matrix and then evaluated to measure the fitness of each solution in the population set. The ith chromosome (individual vector) of the population at time-step (generation) t has d components (dimensions), the chromosome vector is:

$$X_i(t) = \left[X_{i,1}(t), X_{i,2}(t), \ldots, X_{i,d}(t)\right]. \tag{7}$$

For each solution vector $X_i(t)$ that belongs to the current population, three other individuals are randomly chosen from the same generation and used to compose a new trial vector. The lth component of each trial vector is generated by:

$$Y_{k,l}(t+1) = X_{i,l}(t) + Mf\left(X_{m,l}(t) - X_{n,l}(t)\right). \tag{8}$$

Where $Mf \in [0,1]$ is the mutation factor. A crossover with a factor Cr is used to ensure the diversity of the perturbed parameters vectors. To this end, the trial vector is formed:

$$Z_k(t+1) = \left[Z_{k,1}(t+1), Z_{k,2}(t+1), \ldots, Z_{k,d}(t+1)\right]. \tag{9}$$

where

$$Z_{k,j}(t+1) = \begin{cases} Y_{k,j}(t+1) \; if \; rand_j(0,1) \le Cr \;\; or \; j = rand(k) \\ X_{k,j}(i) \; if \; rand_j(0,1) > Cr \;\;\;\; and \; j \ne rand(k). \end{cases} \tag{10}$$

$rand_j(0,1)$ is the jth evaluation of a uniform random number generator with outcome ϵ [0,1]. $rand$(k) is a randomly chosen index ϵ {1,2,...,d}, which ensures that $Z_k(t+1)$ gets at least one parameter from $Y_k(t+1)$. After an evaluation phase, if the offspring yields a better value of the objective function, it is taken to replace its parent in the next generation; otherwise, the parent is chosen and maintained in the population.

$$X_k(t+1) = \begin{cases} Z_k(t+1), if \; \left(f\big(Z_k(t+1)\big) > f\big(X_k(t)\big)\right) \\ X_k(t), \;\;\;\; if \; \left(f\big(Z_k(t+1)\big) \le f\big(X_k(t)\big)\right). \end{cases} \tag{11}$$

where f is the objective function (evaluation function). The process is executed for a fixed number of iterations. The best vector found up to the last generation provides the solution to the problem.

2. *Modified Differential Evolution Algorithm (MoDE):*

A modification has been introduced in Modified Differential Evolution during mutation to push the trial vector quickly to the global optima. In the mutation process, two of the three vectors are substituted by the global best (Gbest) and the local best (Lbest), respectively. The crossover and selection operations are the same as DE.

$$Y_i(t+1) = X_{Gbest}(t) + \alpha\left(X_{Lbest}(t) - X_j(t)\right). \tag{12}$$

Where α is computed as,

$$\propto = \frac{1}{\left[1 + e^{-1/\text{Generation}}\right]} \tag{13}$$

In each generation, the mutation operation uses DE mutation or MoDE mutation process depending on the value of α which is in the range between [1,0.5]. When t increases (iterations), both α and the probability of using MoDE mutation decreases. Also, when MoDE mutation is used, in the first stage of the iterations, Lbest vector has more contribution for evolving the mutant vector than the later stage, where the contribution of Gbest increases.

3. *The proposed Adaptive Maskers Differential Evolution Algorithm (AMDE):*

In The Improved Differential Evolution Algorithm, a new approach is introduced in the phase of selecting the mask in order to speed up the research step for a solution and to push the trial vector to reach the global optima quickly. The population is represented by $k_m \times n_f \times n_p$ matrix, where k_m is the maximum number of centers, n_f is the number of features and n_p is the size of the population. The masker is a $k_m \times n_p$ matrix, each row of the masker is used to activate a random number of centers in a random positions of each gene. The mask matrix is divided into three sub-matrices. The best genes masks zone is the matrix mask area where the masks of the best genes found are located (genes are stored in the ascending order of the fitness function values). The

global best influence zone where the number of clusters and the number of mask vectors are function of the actual global best fitness g_v and the first global best fitness g_m generated by:

$$v_0 = 1 + \frac{g_v.(k_m - 2)}{2.g_m}. \tag{14}$$

and,

$$n_{p0} = 0.5.\left[n_p + \frac{g_v.(2 - n_p)}{g_m}\right]. \tag{15}$$

Where $v_0 \in [1, k_m/2]$ determines the number of clusters variation range, and $n_{p0} \in [1, n_p/4]$ is the number of mask rows in the global best influence zone. $v_{msk} = v_{msk}(t)$ is the number of clusters of the global best solution at iteration t. The random zone mask rows are chosen randomly, so that the exploration of other existing best solution is maintained.
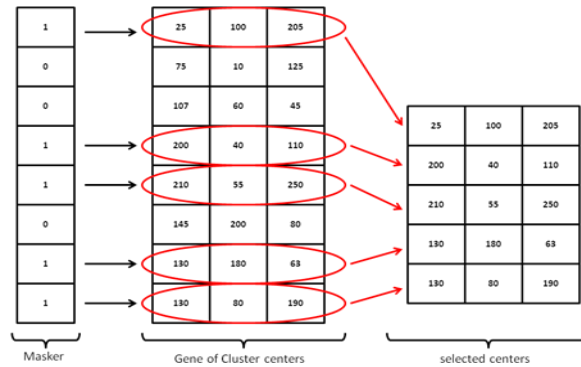
Fig. 1 Mask encoding scheme used to choose a variable number of centers.
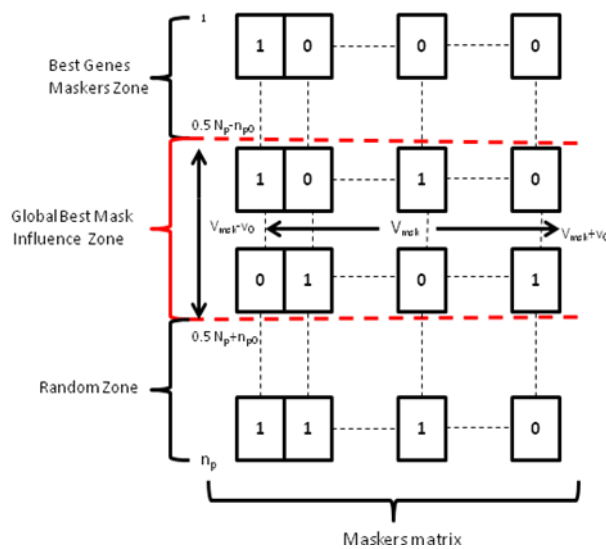
Fig.2 Masker generating scheme.

V.   Experimental Results

In order to establish the effectiveness of the proposed algorithm, a performances comparison study is done with the automatic fuzzy clustering using improved differential evolution, variable length genetic algorithm based fuzzy clustering, and an iterated version of the FCM algorithm. The iterated FCM (IFCM) algorithm is done with running the FCM algorithm for different values of the number of clusters K from K=2 to $n^{1/2}$, with n being the number of data points to cluster. The FCM algorithm is executed 50 times for each K, and the run corresponding to the best J_m value is taken. Each run result of the best solutions is evaluated using the XB index, and the solution producing the minimum XB index value is chosen as the best data partitioning solution. The AFCIDE and VGAFC algorithms are implemented in the same way as in [13] and [14]. The objective of each method is to find the optimal number of clusters automatically.

A. *Remote sensing data set*

The LANDSAT set is a 3-D remote image data set of 5414 samples taken by the L7 ETM+ sensor. partitioned into four distinct classes of sea water (SW) high vegetation (HV), low vegetation (LV) and land (L).

The Fig.V.1 shows the scatter plot of the data set from where it can be seen that the clusters are complex and overlapping.
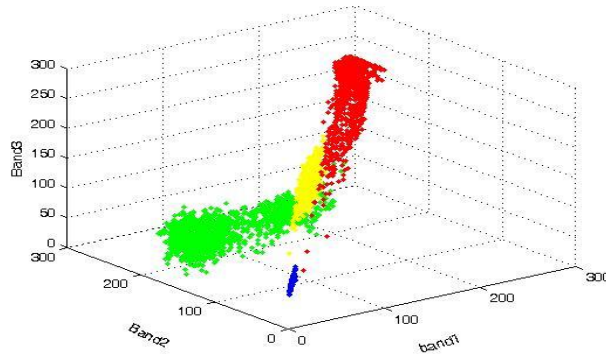

Fig.3 Scatter plot of LANDSAT data set.

B. *Input parameters*

The VMDEAFC, AFCIDE an VGAFC algorithms are executed for 40 generations with a population size of 20. The crossover probability (CR) used for VMDEAFC is 0.3. The mutation factor F and α for VMDEAFC is set to 0.8 or 1/[1+exp(-(1/generation))] randomly. The parameters used in the AFCIDE algorithm are taken by the same procedure as described in [13]. The VGAFC algorithm is executed for 40 generations with a population size of 20, with the parameters crossover probability Cf = 0.8 and mutation probability Mf = 0.3.

C. *Results*

1. Data clustering

A LANDSAT set of data, composed of four clusters, is used as input to the different automatic clustering algorithms. The figures below show the algorithms performances evolution with the generation succession (Local best fitness value(Lbest), Local best number of clusters detected(LNc), Global best fitness value(Gbest), Global best number of clusters(GNc), Real number of clusters(RNc). Clearly, the AMDE is able to detect the actual number of clusters rapidly, in the first stage of the algorithm evolution. And if another zone of optimal number of clusters is detected, the algorithm provides a rapid research in it, and detects the optimal solution. The AFCIDE algorithm detects the actual number of clusters in the 34th generation Fig 3. The VGAFC algorithm provides the solution at the 17th generation Fig. 5. The MoDEAFC has detected 5 clusters Fig. 6. The AMDE provides the solution at the 3rd generation Fig. 7. Fig. 8 illustrates the solution found at each generation for the different algorithms used for the comparative study.
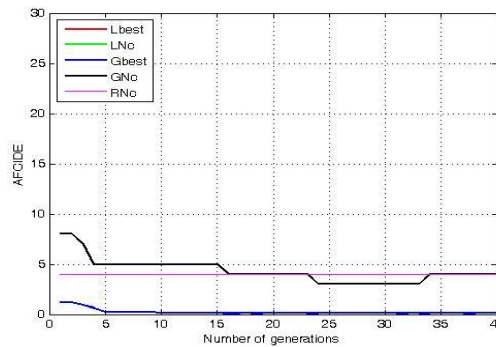

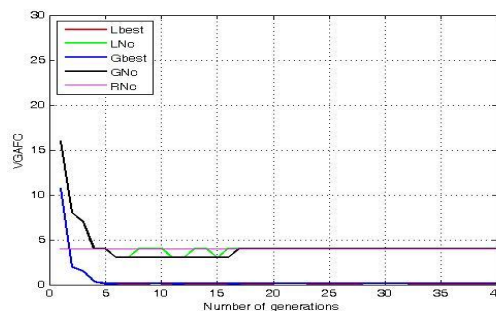Fig.4 AFCIDE performances for LANDSAT data clustering.

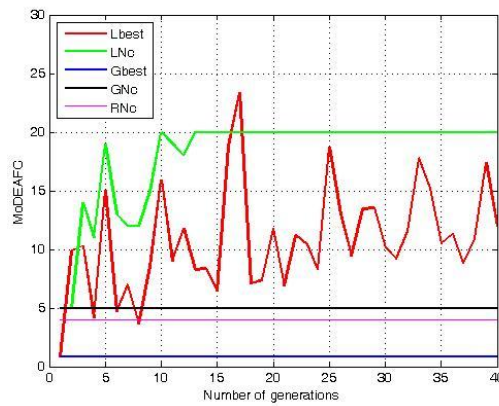Fig.5 VGAFC performances for LANDSAT data clustering

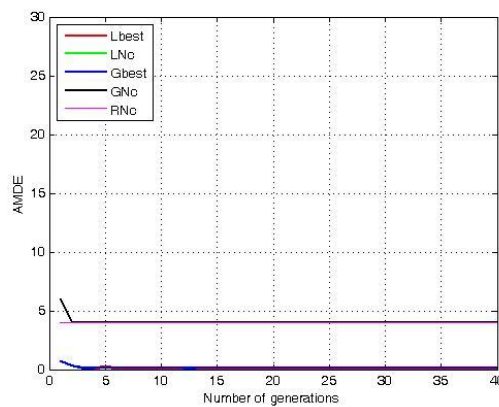Fig.6 MODEAFC performances for LANDSAT data clustering
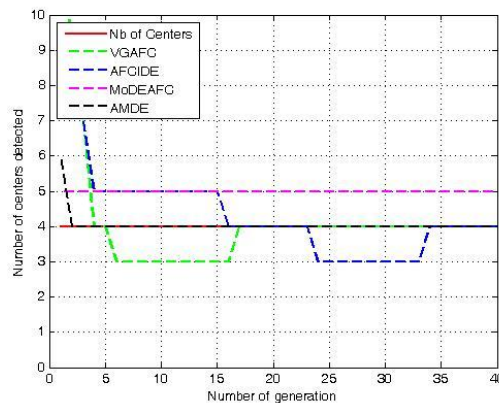

Fig.7 AMDE performances for LANDSAT data clustering

Fig.8 Number of cluster found by the different algorithms for the LANDSAT data set clustering.

2. Image segmentation

A satellite image of Kvarneric land features is used for demonstrating image segmentation. The JPEG image has a size of 500×500 and it shows the Kvarneric Narrows, a group of islands in the southwestern part of Croatia. The image was acquired using the sensor L7 ETM+. The clustering can be verified visually from the clustered images. Fig 9 shows the original satellite Kvarneric image. Figs 10-14 illustrate the corresponding clustered images using AMDE, IFCM, AFCIDE, VGAFC, and MODEAFC algorithms, respectively. The image presents four classes, sea water (SW) determined as blue, open space (OS) determined as red, high vegetation (HV) determined as green, low vegetation (LV) determined as yellow.

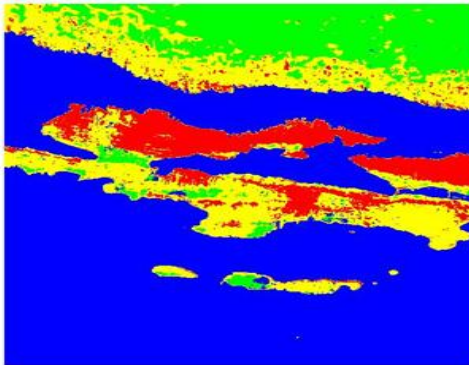Fig.9 Satellite image of Kvarneneric.

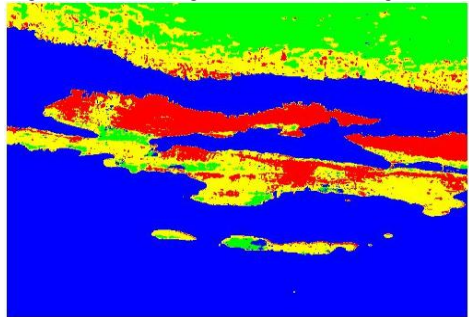Fig 10 Clustered image of Kvarneneric using AMDE

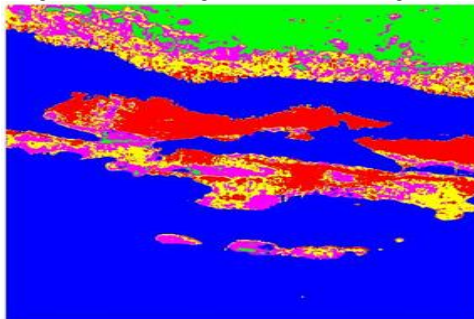Fig 11 Clustered image of Kvarneneric using IFCM
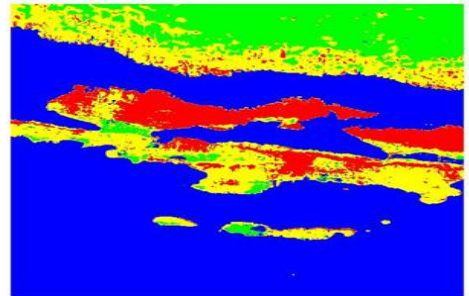

Fig 12 Clustered image of Kvarneneric using AFCIDE
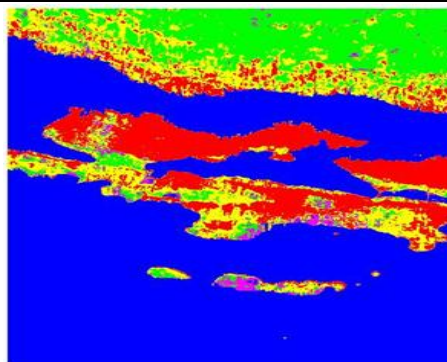

Fig 13 Clustered image of Kvarneneric using VGAFC

Fig 14 Clustered image of Kvarneneric using MoDEAFC

VI. Conclusion

In this paper, a adaptive maskers for differential evolution based fuzzy clustering algorithm has been described for real-world data sets automatic clustering. An important feature of the new technique is that it is able to rapidly detect the number of clusters. The modified version of the differential evolution technique has been used with an adaptive masker. The results of the experiments on real-world data sets indicate that AMDE performances are consistently better than the AFCIDE, VGAFC, and MoDEAFC clustering algorithms.

References

[1] Y. Leung, J. Zhang, and Z. Xu, "Clustering by scale-space filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1396–1410, Dec. 2000.

[2] E. W. Forgy, "Cluster analysis of multivariate data: Efficiency versus interpretability of classification," *Biometrics*, vol. 21, no. 3, pp. 768–769,1965.

[3] C. T. Zahn, "Graph-theoretical methods for detecting and describing gestalt clusters *IEEE Trans. Comput.*, vol. C-20, no. 1, pp. 68–86, Jan. 1971.

[4] T. Mitchell, *Machine Learning*. New York: McGraw-Hill, 1997.

[5] N. R. Pal, J. C. Bezdek, and E. C.-K. Tsao, "Generalized clustering networks and Kohonen's self-organizing scheme," *IEEE Trans. Neural Netw.*, vol. 4, no. 4, pp. 549–557, Jul. 1993.

[6] E. Falkenauer, *Genetic Algorithms and Grouping Problems*. Chichester, U.K.: Wiley, 1998.

[7] J. H. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: Univ. Michigan Press, 1975.

[8] R. Storn and K. Price, "Differential evolution—A simple and efficient heuristic strategy for global optimization over continuous spaces," *J.Glob. Optim.*, vol. 11, pp. 341–359,   1997.

[9] Ujjwal Maulik and Indrajit Saha, "Automatic Fuzzy Clustering Using Modified Differential Evolution for Image Classification, " *IEEE Transactions on Geoscience and Remote Sensing,* 2010.

[10] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. New York: Plenum, 1981.

[11] *L. Groll and J. Jakel, "A new convergence proof of fuzzy c-means," IEEE Trans. Fuzzy   Syst., vol. 13, no. 5, pp. 717–720, Oct. 2005.*

[12] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern   Anal. Mach. Intell.*, vol. 13, no. 8, pp. 841–847, Aug. 1991.

[13] S. Das, A. Abraham, and A. Konar, "Automatic clustering using an improved differential evolution algorithm," IEEE Trans. Syst., Man, Cybern. A, Syst., Humans, vol. 38, no. 1, pp. 218–237, Jan. 2008.

[14] U. Maulik and S. Bandyopadhyay, "Fuzzy partitioning using a real-coded variable-length geneticalgorithmfor pixel classification," IEEE Trans. Geosci. Remote Sens., vol. 41,   no. 5, pp. 1075–1081, May 2003.