Journal of Information Technology Management

Print ISSN: 2008-5893

Online ISSN: 2423-5059

An Archive-based Steady-State Fuzzy Differential Evolutionary Algorithm for Data Clustering (ASFDEaDC)

Nushrat Praveen

Corresponding Author, Department of Computer Science, National Institute of Technology, Mizoram, Aizawl, India. E-mail: nus123nit@gmail.com

Vally D

Department of Computer Science, Vignan's Institution of Information, Vishakhapatnam, India. E-mail: vally2605pg@gmail.com

Abstract

In the current paper, we have assimilated fuzzy techniques and optimization techniques, namely differential evolution, to put forward a modern archive-based fuzzy evolutionary algorithm for multi-objective optimization using clustering. The current work account for the application of a cluster associated approach. Specific quantitative cluster validity measures, i.e., J-measure and Xie-Beni, have been referenced to carry out the appropriate partitioning. The proposed algorithm introduces a new form of strategy which attempts to benefit the feasible search domain of the algorithm by minimizing the analysis and exploration of less beneficial search scope. This clustering method yields a group of trade-off solutions on the ultimate optimal pare to front. Eventually, these solutions are united and maintained in an archive for further evaluation. The current work summarizes and organizes an archive concerned with excellent and diversified solutions in an effort to outline comprehensive non-dominated solutions. The degree of efficiency is revealed with respect to partitioning on gene expression and real-life datasets. The proposed algorithm seeks to reduce the function assessment analysis and maintains a very small working population size. The effectiveness of the proposed method is presented in comparison with some state-of-art methods.

Keywords: Multi-objective optimization, Clustering, Differential evolution, Evolutionary algorithm, Euclidean based distance, Gene expression data.

DOI: 10.22059/jitm.2021.80617

Introduction

Clustering is the classifying together of similar or homogeneous data items into clusters. To handle the vast collection of data, a clustering process was proposed to detect groups in a collection of records (Alizadeh et al., 2000 and Jain, et al., 1999). A clustering algorithm split-up a dataset into several classes such that the similarity within a class is greater than between the class. In simple words, the aim is to set apart species with similar traits and assign them into clusters. The need for organizing the expanding data and eliciting important knowledge from the data makes clustering techniques extensively applied in various application areas such as image segmentation, data mining, artificial intelligence, statistics, pattern recognition, and other scientific applications. The maximization of inter-cluster distance and minimization of intra-cluster distance is the paramount purpose of clustering. Ordinarily, a few homogeneity and similarity approaches are, in general, employed to achieve a combination of classified patterns (Horta, et al., 2011 and Mezura-Montes & Coello, C. 2011). Frequently accessible distance measures include Euclidean distance (Noorbehbahani, F., et al., 2015), cosine similarity (Noorbehbahani, F., et al., 2015), Mahalonabisdistance (Noorbehbahani, F., et al., 2015). In a real-life scenario, the majority of the datasets consist of clusters possessing various structures and convexities.

Moreover, in recent literature, several multi-objective optimizations (MOO) was proposed to deal with data clustering problem and simultaneously determine clusters of the different pattern by implementing the particular algorithm and also attempt to detect every possible partitioning alternatively (Alaei, H. K.,et al., 2013, Kriegel, et al., 2011, Saha, I.,et al., 2012, Bezdek, J.1981 and Tou, J. T., & Gonzalez, R. C. 1974).Usually, these MOO based approaches adopt genetic algorithm or simulated annealing in the form of elementary technique for optimization. In recent years, global optimization problems have become one of the paramount and essential tasks in many branches (Tou, & Gonzalez, & 1974). The main intention behind resolving multi-objective optimization problems (MOPs) comes down to detect a set of nondominated solutions. Multi-objective optimization (MOO) is also known as Pareto optimization or Vector optimization, concerned with optimizing more than one objective function simultaneously. Usually, we consider a collection of best solutions rather than looking for a single eminent solution. All Pareto optimal can be regarded as equally recommendable, but we need some act or process to recognize the most desirable among them. In a little while, ago literature, lot of different optimization strategies like particle swarm optimization (PSO) (Deb, & Tiwari (2008), cuckoo search (Deb, et al 2002), ant colony optimization (Tvrdik, J., & Křivý, I. (2015) and differential evolution (Agustı, & et al., 2012) have consistently been developed which provide improved performance in comparison to the genetic algorithm with respect to convergence criteria (Heyer, L. J., et al 1999).

Motivated by these behaviors, in this current paper, we introduce a modern archive-based steady-state fuzzy differential evolutionary algorithm for data clustering. It explores the search capabilities of some optimization techniques and differential evolution. It solves the problems concurrently by evolving a population of solutions, and the population is generated randomly from the dataset. The algorithm utilizes the general idea of pareto ranking derived from a fast nondominated sorting algorithm (NSGA-II) (Alon, U., et al., 1999). The proposed method is very much convenient to execute and try to attain the finest features of different MOEAs. The optimization performed concurrently by means of this proposed fuzzy clustering algorithm adapts two popular cluster validity measures (de Souto, M. C., et al., 2008) as an objective function utilizing the distance measure Euclidean distance-based J-measure index and Xie-beni index. The utilization of these two existing objective functions benefits the proposed technique to explore the search region comprehensively to determine partitionings having diverse structured clusters.

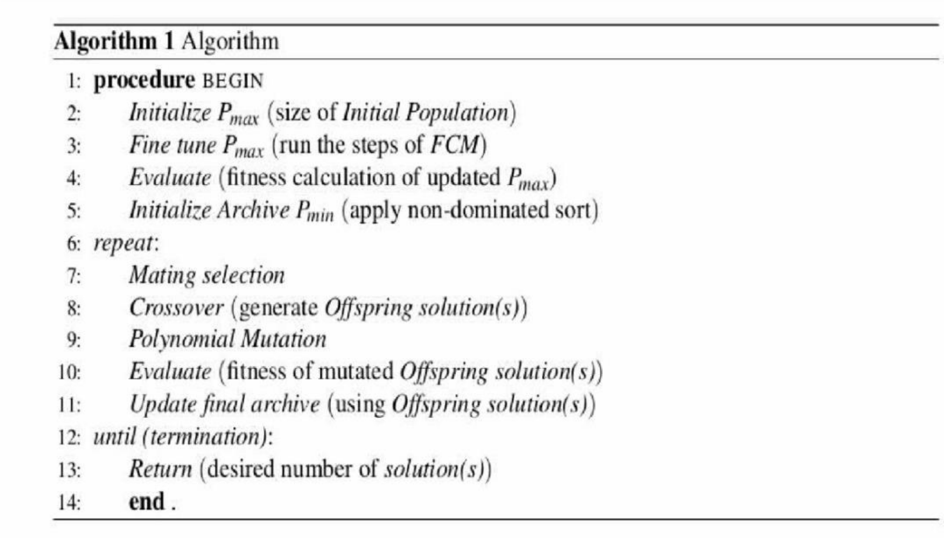
In comparison with the distinctive clustering strategies, the current study targets to explore whether the application of an archive-based fuzzy evolutionary algorithm for the clustering approach produces more desirable and effective result accuracy. In the literature, several archive-based multi-objective evolutionary algorithms have been developed, endeavoring to achieve the better performance of particular clustering algorithms and draw attention to integrate the partitioning solutions achieved from diverse clustering strategies or the similar strategy with varying framework and parameter settings (Rezaee, M. R., et al 1998). The introduced MOO-based clustering approaches introduce a batch of possible solutions situated on the decisive pareto Front (PF), and eventually, with regard to some cluster quality measures, a subset of these solutions is selected (Rezaee, M. R., et al 1998).

Literature Review

In recent literary works, several adaptive and effective clustering approaches based on the genetic algorithm are existing and are competent to figure out the global optimization problem. For the categorization of gene expression data, many clustering-based strategies have been proposed in the bioinformatics discipline. A few examples of clustering algorithms are hybrid niching genetic algorithm (HNGA) (Ester, M., et al 1996) based on Euclidean distance for the distribution of data points to particular clusters, variable string length genetic k-means algorithm (GCUK) (Pal, N. R., &Bezdek, J. C. 1995). In recognition of cancer subspecies, a wide range of clustering techniques is available (Rousseeuw, P. J. 1987). Researchers from the bio-medical discipline favor to opt for standard clustering methods for the purpose of solving issues in the classification and recognition of gene expression datasets and cancer subtypes. In order to detect the clusters of equilateral hyper-spherical structures, a k-means technique based on a genetic algorithm has been demonstrated in (Pal, N. R., &Bezdek, J. C. 1995 and Tamayo, P., et al., 1999). An automatic clustering strategy has been proposed by Liu et al. in (Liu, Y., et al., 2011), which introduces a division absorption and noising particularization-based mutation method with regard to preserve the population discrepancy and selection process. With the micro-array technological advancement, the feasibility to figure out molecular signatures of cancer cells has been enhanced (Ni, et al., 2015 and Deb, & Agrawal, 1995). Some relevant and remarkable acknowledgments of clustering methods have been developed for the sake of the analytic study of micro-array data (cancer specimens) in (Deb, K., & Agrawal, 1995). An evolutionary algorithm has been proposed by Horta, et al., 2011 basing itself on a fuzzy clustering technique attempt to identify the clusters automatically from the corresponding dataset. Authors in (Bandyopadhyay, S. 2005 and Bandyopadhyay, & Maulik, 2002) have developed a splendid scheme for the evolutionary method where the fundamental idea is based on grouping encoding technique and opt explorative and exploitative operator. In an effort to resolve the shortcomings of the k-means clustering algorithm, several other methods have been proposed and put into effect in the clustering of many gene expression datasets (Chu, et al., 1998 and Das, et al., 2005). Apart from the k-means method, there exists a self-organizing map clustering algorithm that achieves eminent performance in the classification of gene expression data (Saha, S., &Bandyopadhyay, S. 2009). Among the applications of non-hierarchical clustering, (Saha, S., et al., 2013) introduced a hybrid strategy through the aggregation of the k-means approach and DE in order to obtain global optimization (Wen, X., et al., 1998 and Yang & Deb, 2014).

An archive-based steady-state fuzzy differential evolutionary algorithm for data clustering

In this paper, the proposed algorithm is a fuzzy evolutionary optimization algorithm based on clustering. A flowchart has been demonstrated in Fig. 1, displaying the different sections of the algorithm. The proposed algorithm is micro-genetic on account of its quite small working population size. It is requisite to keep the size of the working population as small as possible.



Algorithm 1 Algorithm

```
procedure BEGIN
    Initialize $P_{max}$ (size of Initial Population)
    Fine tune $P_{max}$ (run the steps of FCM)
    Evaluate (fitness calculation of updated $P_{max}$)
    Initialize Archive $P_{min}$ (apply non-dominated sort)
repeat:
    Mating selection
    Crossover (generate Offspring solution(s))
    Polynomial Mutation
    Evaluate (fitness of mutated Offspring solution(s))
    Update final archive (using Offspring solution(s))
until (termination):
    Return (desired number of solution(s))
    end .
```

Figure 1. Pseudocode of the Algorithm

A. Population Initialization

This is the most essential part of the proposed technique. The population size significantly influences the performance of the algorithm. It is possible to achieve the initial primitive population (P max) in various ways. Initialization can be performed by either exploring the search region uniformly or any previous information that can be incorporated during the population's initialization. This provides a better inclusive partition of the population and does not require any further interpretation.

We randomly generate the initial population and include the clustering concept during the initialization. Accordingly, (Pmax) is the size of initial solutions which is generated at random. After this step, the initial population is evaluated, which is explained in the next section.

Example: Let N = 4 and C = 3, i.e., the dataset considered is four-dimensional and cluster count taken into account is three. Then the vector

5.5, 2.4, 3.7, 1.0, 5.6, 3.0, 4.5, 1.5, 4.8, 3.4, 1.6, 0.2

illustrates the three centroids of the cluster [[5.5, 2.4, 3.7, 1.0], [5.6, 3.0, 4.5, 1.5], [4.8, 3.4, 1.6, 0.2]] Note that in this manner (P max) vectors are initialized in the population (set of solutions).

B. Fine-tuning of Initial Populations

Since these sets of solutions are randomly generated, so to remove the randomness, we have incorporated Fuzzy c-means (FCM) measure on the initialized population. To estimate the measures, the centers prearranged in the form of the vector are initially derived. The fine-tuning starts with the initially randomized C cluster centroids and obtain the fuzzy membership values of every data objects in each cluster by applying the measure mentioned below:

$$u_{xy} = \frac{\left(\frac{1}{D(v_x, w_y)}\right)^{\frac{1}{m-1}}}{\sum_{c=1}^{C}\left(\frac{1}{D(v_x, w_y)}\right)^{\frac{1}{m-1}}} \tag{1}$$

s.t. $1 \le x \le C$, $1 \le y \le M$, U is the matrix denoting fuzzy partition, M is the count of data objects, C is the cluster count, and m represents the fuzzy component. After the calculation of the membership matrix, we will update the cluster centroids using the following equation:

$$v'_x = argmin \sum_{l=1}^{M} u_{xy}^{m} D\left(v_x, w_y\right) \tag{2}$$

$$s.t\ 1 \le x \le C, \quad 1 \le y \le M \ \ and \quad v_x = w'_{vx}$$

C. Evaluation of Updated Solutions

For evaluation, two cluster validity indices, namely J-measure (Saha, S., et al., 2008) index and XB (Yue, S., et al., 2016) index have been used as objective functions. Moreover, the optimization of the objective functions helps to explore the relevant clustering. A brief idea about these cluster validity indices is given below:

The J-measure index can be written as:

$$J_m = \sum_{x=1}^{C}\sum_{y=1}^{M} \mathcal{U}_{xy}^{m} D^2(v_x, w_y) \tag{3}$$

To calculate XB index, we need to calculate the ratio of cluster compactness. Let's say, Θ is the total variation, and ρ is the minimum separateness among the clusters. Here Θ and ρ can be computed as:

$$\Theta(P,Q;T) = \sum_{c=1}^{C}\sum_{y=1}^{M} \mathcal{U}_{cy}^{2} D^2(v_c, w_y) \tag{4}$$

$$\rho(Q) = min_{y \neq x} D^2(v_c, v_x)$$

$$XB(P,Q;T) = \frac{\Theta(P,Q;T)}{M \times \rho(Q)}$$

Hence, Jm estimates the diversity in clusters globally; on the other note, the XB index calculates the cluster variance globally and locally. Thus the lower value of Jm and XB implies clustering in the best possible way.

D. Archive Initialization using updated population

The ASFDEaDC consistently attempts to keep only non dominated set of solutions in its archive. The size of the archive confines between Pmin and Pmax. In the process of initializing the archive, the fast-nondominated sort (Alon, U., et al., 1999) is implemented on the updated population.

The sorting continues until we obtain Pmin solutions. By expanding the size of Pmin, we can increase the discrepancy between the solutions. Say, Fj be the j th sub front obtained in sorting and |Fj | represents the number of solutions in Fj. Let the method provides a set of sub-fronts F1, F2, ...,Fw such that, (|F1 |, |F2 |, ..., |Fw |) = Pmin is satisfied.

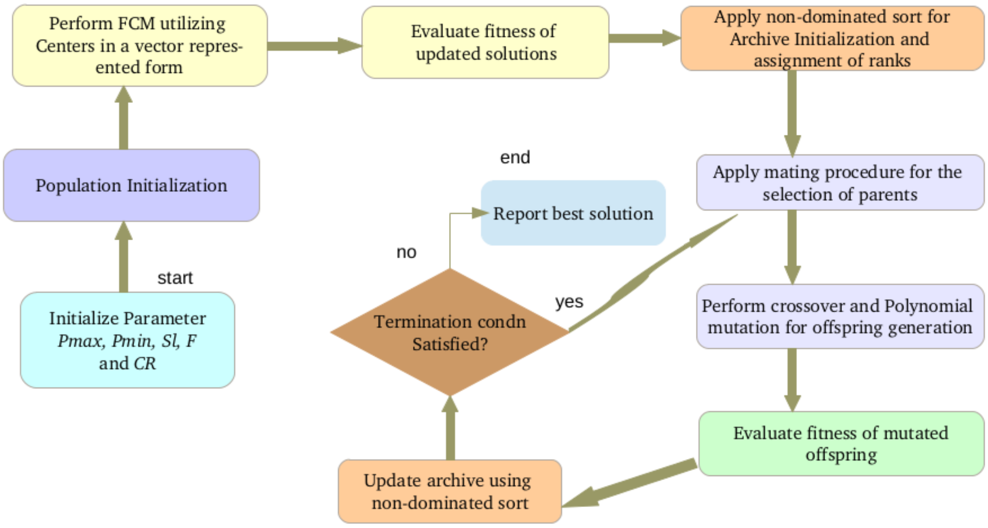


Figure 2. Flowchart of ASFDEaDC

E. Mating selection

In the current work, the working population and mating storage are similar. The main aim is to search the optimal PF region to acquire the divergence. The mating algorithm proceeds by opting for the selection norms such that we choose one primary parent and three auxiliary parents. We randomly assign three solutions from the archive as auxiliary parents. To choose the primary parent, a random number d is induced uniformly in [0, 1]. If d is exceeding the selection ratio Sl, then we randomly choose one solution from the archive and assign it as the primary parent. Else, the normalization of fitness values of entire solutions within the archive is performed. This approach intends to ensure the separateness among these four parents.

F. Crossover and Polynomial Mutation

We compute crossover to produce one offspring for each evaluation. Let ax1, ax2, and ax3 be the three auxiliary parents, and pr is the primary parent. Let d' be another random number generated uniformly in [0, 1]. The crossover operator utilizes two optimizing parameters F and CR. Let, ax_{ij} (pr_j) be the j^{th} component of ax (p_r), O be the offspring, and O_j be the j^{th} component of O, then:

$$O_j = \begin{cases} a_{x3j} + F(a_{x1j} - a_{x2j}), & \text{if } d' \leq CR \\ p_{rj}, & \text{otherwise} \end{cases} \tag{5}$$

Crossover and polynomial mutation is described in more detail in Maulik, U., & Bandyopadhyay, S. 2003 and Iyer, V. R., et al 1999, respectively.

G. Update final archive

We compare the offspring with the complete archive of ASFDEaDC. First, we append the offspring solutions to the archive. In addition to this, the fast-nondominated-sorting is implemented on the archive. We terminate the sorting after acquiring Pmin solutions such that (|F1 |, |F2 |, ..., |Fw |) = Pmin is satisfied. If w = 1, then we opt one special procedure: we append the offspring with the archive. In case that if the offspring is dominated by any of the solutions within the archive, Ar, then the offspring is obsoleted, or else, the offspring will dominate a set of solutions, Si (this could be null). Furthermore, we monitor, if Pmin = (|Ar| − |Si | + 1) = Pmax , append the offspring to Ar , and prune solutions in Si from Ar . Else if, (|Ar| − |Si| + 1) <Pmin, we prune one solution from the Si by normalizing the objective values, then append the leftover solutions to Ar and again perform the non-dominated sort. Else, (|Ar| − |Si|) = Pmax, then the offspring is appended to Ar, and a solution is obsoleted from Ar using the normalization process.

H. Termination criteria

In the ASFDEaDC method, the procedure for tuning, evaluation, archive initialization, mating, crossover, mutation, and archive updation is performed for a fixed number of iterations. The last iteration of the method will provide a set of solutions, and from that particular set, the best and proficient solution for the clustering problem is analyzed in terms of cluster validity measure.

Experimental Results

A. Gene expression datasets

Three gene datasets are considered for the experimental analysis of the proposed method. A brief discussion is given below:

1) Yeast sporulation: This dataset is shared publicly at http://cmgm.stanford.edu/pbrown/sporulation, including 6118 genes. With the 1.6 threshold level, some log-transformed data have been ignored whose transformation in expression levels is not observed significantly, and the resulting set includes 474 genes.

2) Human Fibroblasts Serum: This dataset can be collected from the web address http://www.sciencemag.org/feature/data/984559.shl and contains 8613 human genes. The dataset is 13 dimensional, and a total of 517 genes have been chosen in which the fluctuations of expression levels are somewhat identified.
3) RatCNS: The dataset is taken from this link http://faculty.washington.edu/kayee/cluster. This dataset contains 112 genes obtained during the development of the rat central nervous system.

B. Real-life datasets

A brief introduction to real-life datasets are mentioned below and was collected from the database of UCI machine learning:

1) Iris: This is a four-dimensional dataset 2 (sepal length, sepal width, petal length, and petal width). It consists of 150 data points divided into three classes of Iris flowers, namely, Setosa, Virginia, and Versicolor.
2) Glass: Glass dataset 3 comprises of 214 data objects distributed in six classes.
3) Wine: This dataset 4 contains 178 data objects grouped into thirteen features along with three classes.
4) Cancer: This dataset 5 contains 683 data points, including thirteen features along with two classes.

C. Chosen validity measure

With the objective of obtaining quantitative and qualitative partitioningbyy the suggested algorithm, we have used the Silhouette index validity measures (Heyer, L. J., et al., 1999).

Silhouette index (Heyer, L. J., et al., 1999): Silhouette index (S) is an implicit cluster validity index. It estimates the compactness and separateness among the clusters utilizing Euclidean distance, which varies from −1 to +1. Maximal values of the Silhouette index quantifies eminent partitioning. (S) is evaluated using the equation mentioned below:

$$(S) = \frac{a - b}{\max(a, b)} \tag{6}$$

Where a represents the average mean distance between the data points in the same cluster, and b represents the minimum average distance between data points in the different clusters.

D. Results and Discussion

The proposed ASFDEaDC clustering algorithm utilizes the exploring proficiencies of differential evolution optimization techniques and fuzzy techniques. The parameter settings used for the proposed clustering method are mentioned below: Population size (Pmax) = 50, Fuzzy parameter (F) = 0.5, Crossover parameter (CR) = 0.1, Sl = 0.25, and Pmin = 20. The experimental results of ASFDEaDC are outlined through three gene datasets, namely Yeast sporulation, Serum and RatCNS, and four real-lifedatasets, namely Iris, Glass, Wine, and Cancer. The silhouette index score obtained by the proposed algorithm over three gene datasets and four real-life datasets is shown in Table1 and Table2, respectively.

Table 1. Silhouette Index Score Obtained for Three Gene Expression Datasets by Various Clustering Methods

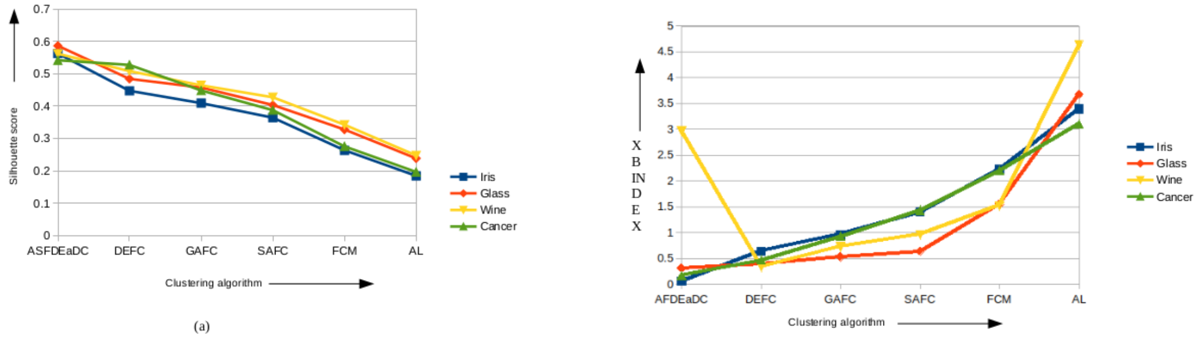
Algorithm	Dataset		
	Yeast Sporulation	Serum	Rat CNS
ASFDEaDC	0.6340	0.4544	0.4943
IDEFC	0.5662	0.3855	0.4681
DEFC	0.5591	0.3628	0.4573
GAFC	0.5421	0.3443	0.4486
SAFC	0.5372	0.3307	0.4401
FCM	0.5163	0.3152	0.4291
AL	0.5007	0.2763	0.4122

Table 2. Silhouette Index Score Obtained for Four Real Life Datasets by Various Clustering Methods

Algorithm	Dataset			
	Iris	Glass	Wine	Cancer
ASFDEaDC	0.5626	0.5864	0.5616	0.5416
DEFC	0.4472	0.4842	0.5082	0.5273
GAFC	0.4092	0.4572	0.4642	0.4482
SAFC	0.3642	0.4033	0.4273	0.3874
FCM	0.2633	0.3274	0.3422	0.2753
AL	0.1842	0.2382	0.2473	0.1962

These tables also describe the silhouette index score obtained by improved differential evolution for microarray analysis (IDEFC) (Ravi, V., et al., 2010), Differential Evolution based Fuzzy Clustering (DEFC) (Sheng, W., et al., 2005), Genetic Algorithm based Fuzzy

Clustering (GAFC) (Xie, X. L., &Beni, G. 1991), Simulated Annealing based Fuzzy Clustering (SAFC) (Bandyopadhyay, S. 2005), Fuzzy C-Means (FCM) (Pal, N. R., &Bezdek, J. C. 1995) and Average Linkage (AL) (Liu, Y., et al 2011) clustering methods. Table1 and Table2 report that the proposed ASFDEaDC clustering method obtains a higher silhouette score in comparison to other clustering techniques in all the runs of the entire approachable datasets.



(a)

Figure 3. Convergence plots of silhouette scores obtained by different clustering algorithms for (a) gene datasets and (b) artificial datasets

Conclusion

In this paper, an archive-based steady-state fuzzy differential evolution algorithm for data clustering (ASFDEaDC) method has been elaborated. The challenge of fuzzy clustering has turned out to be a stochastic optimization tool of a cluster validity measure. Results on different gene datasets and real-life datasets demonstrate that the performance of ASFDEaDC is consistently better than IDEFC, DEFC, GAFC, SAFC, (FCM) and (AL) clustering techniques. The significant performance of the ASFDEaDC method reveals that it may be possibly motivational to use this algorithm in distinct real life application and various data clustering problem of micro-array. As an extension to take forward this research, the proposed algorithm ASFDEaDC can also be made advanced such that the number of clusters can be obtained automatically.

References

Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., ...&Staudt, L. M. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, *403*(6769), 503-511.

Alaei, H. K., Salahshoor, K., &Alaei, H. K. (2013). A new integrated on-line fuzzy clustering and segmentation methodology with adaptive PCA approach for process monitoring and fault detection and diagnosis. *soft computing*, *17*(3), 345-362.

Agustı, L. E., Salcedo-Sanz, S., Jiménez-Fernández, S., Carro-Calvo, L., Del Ser, J., &Portilla-Figueras, J. A. (2012). A new grouping genetic algorithm for clustering problems. *Expert Systems with Applications*, *39*(10), 9695-9703.

Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D., & Levine, A. J. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceedings of the National Academy of Sciences*, *96*(12), 6745-6750.

Bezdek, J. (1981). Pattern Recognition with Fuzzy Objective Algorithms Plenum Press New York Google Scholar.

Bandyopadhyay, S., &Maulik, U. (2002). Genetic clustering for automatic evolution of clusters and application to image classification. *Pattern recognition*, *35*(6), 1197-1208.

Bandyopadhyay, S. (2005). Simulated annealing using a reversible jump Markov chain Monte Carlo algorithm for fuzzy clustering. *IEEE Transactions on Knowledge and data Engineering*, *17*(4), 479-490.

Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P. O., &Herskowitz, I. (1998). The transcriptional program of sporulation in budding yeast. *Science*, *282*(5389), 699-705.

Das, S., Konar, A., &Chakraborty, U. K. (2005, June). Two improved differential evolution schemes for faster global search. In *Proceedings of the 7th annual conference on Genetic and evolutionary computation* (pp. 991-998).

Deb, K., & Agrawal, R. B. (1995). Simulated binary crossover for continuous search space. *Complex systems*, *9*(2), 115-148.

Deb, K., & Tiwari, S. (2008). Omni-optimizer: A generic evolutionary algorithm for single and multi-objective optimization. *European Journal of Operational Research*, *185*(3), 1062-1087.

Deb, K., Pratap, A., Agarwal, S., &Meyarivan, T. A. M. T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE transactions on evolutionary computation*, *6*(2), 182-197.

De Souto, M. C., Costa, I. G., de Araujo, D. S., Ludermir, T. B., &Schliep, A. (2008). Clustering cancer gene expression data: a comparative study. *BMC bioinformatics*, *9*(1), 1-14.

Ester, M., Kriegel, H. P., Sander, J., &Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (Vol. 96, No. 34, pp. 226-231).

Horta, D., De Andrade, I. C., &Campello, R. J. (2011). Evolutionary fuzzy clustering of relational data. *Theoretical Computer Science, 412*(42), 5854-5870.

Iyer, V. R., Eisen, M. B., Ross, D. T., Schuler, G., Moore, T., Lee, J. C., ...& Brown, P. O. (1999). The transcriptional program in the response of human fibroblasts to serum. *science, 283*(5398), 83-87.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR), 31*(3), 264-323.

Kriegel, H. P., Kröger, P., Sander, J., &Zimek, A. (2011). Density-based clustering. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1*(3), 231-240.

Kruglyak, S., &Yooseph, S. (1999). Exploring expression data: identification and analysis of coexpressed genes. *Genome research*, *9*(11), 1106-1115.

Liu, Y., Wu, X., &Shen, Y. (2011). Automatic clustering using genetic algorithms. *Applied mathematics and computation*, *218*(4), 1267-1279.

Mezura-Montes, E., &Coello, C. A. C. (2011). Constraint-handling in nature-inspired numerical optimization: past, present and future. *Swarm and Evolutionary Computation*, *1*(4), 173-194.

Maulik, U., &Bandyopadhyay, S. (2003). Fuzzy partitioning using a real-coded variable-length genetic algorithm for pixel classification. *IEEE Transactions on geoscience and remote sensing*, *41*(5), 1075-1081.

Noorbehbahani, F., Mousavi, S. R., &Mirzaei, A. (2015). An incremental mixed data clustering method using a new distance measure. *Soft Computing*, *19*(3), 731-743.

Ni, Q., Pan, Q., Du, H., Cao, C., &Zhai, Y. (2015). A novel cluster head selection algorithm based on fuzzy clustering and particle swarm optimization. *IEEE/ACM transactions on computational biology and bioinformatics*, *14*(1), 76-84.

Pal, N. R., &Bezdek, J. C. (1995). On cluster validity for the fuzzy c-means model. *IEEE Transactions on Fuzzy systems*, *3*(3), 370-379.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, *20*, 53-65.

Rezaee, M. R., Lelieveldt, B. P., &Reiber, J. H. (1998). A new cluster validity index for the fuzzy c-mean. *Pattern recognition letters*, *19*(3-4), 237-246.

Ravi, V., Aggarwal, N., & Chauhan, N. (2010, December). Differential evolution based fuzzy clustering. In *International Conference on Swarm, Evolutionary, and Memetic Computing* (pp. 38-45). Springer, Berlin, Heidelberg.

Sheng, W., Swift, S., Zhang, L., & Liu, X. (2005). A weighted sum validity function for clustering with a hybrid niching genetic algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 35*(6), 1156-1167.

Saha, I., Plewczynski, D., Maulik, U., &Bandyopadhyay, S. (2012). Improved differential evolution for microarray analysis. *International journal of data mining and bioinformatics*, *6*(1), 86-103.

Saha, S., &Bandyopadhyay, S. (2009). A new point symmetry based fuzzy genetic clustering technique for automatic evolution of clusters. *Information Sciences*, *179*(19), 3230-3246.

Saha, S., Ekbal, A., Gupta, K., &Bandyopadhyay, S. (2013). Gene expression data clustering using a multiobjective symmetry based clustering technique. *Computers in biology and medicine*, *43*(11), 1965-1977.

Saha, S., Das, R., &Pakray, P. (2018). Aggregation of multi-objective fuzzy symmetry-based clustering techniques for improving gene and cancer classification. *Soft Computing*, *22*(18), 5935-5954.

Tou, J. T., & Gonzalez, R. C. (1974). Pattern recognition principles Addison-Wesley Reading.

Tvrdik, J., &Křivý, I. (2015). Hybrid differential evolution algorithm for optimal clustering. *Applied Soft Computing*, *35*, 502-512.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., ...&Golub, T. R. (1999). Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proceedings of the National Academy of Sciences*, *96*(6), 2907-2912.

Xie, X. L., &Beni, G. (1991). A validity measure for fuzzy clustering. *IEEE Transactions on pattern analysis and machine intelligence*, *13*(8), 841-847.

Yang, X. S., & Deb, S. (2014). Cuckoo search: recent advances and applications. *Neural Computing and Applications*, *24*(1), 169-174.

Yue, S., Wang, J., Wang, J., &Bao, X. (2016). A new validity index for evaluating the clustering results by partitional clustering algorithms. *Soft Computing*, *20*(3), 1127-1138.

Wen, X., Fuhrman, S., Michaels, G. S., Carr, D. B., Smith, S., Barker, J. L., & Somogyi, R. (1998). Large-scale temporal gene expression mapping of central nervous system development. *Proceedings of the National Academy of Sciences*, *95*(1), 334-339.

Bibliographic information of this paper for citing:

Praveen, N. & Vally, D. (2021). An Archive-based Steady-State Fuzzy Differential Evolutionary Algorithm for Data Clustering (ASFDEaDC). *Journal of Information Technology Management*, Special Issue, 54-67.