IJCRM

International Journal of Contemporary Research In Multidisciplinary

Research Article

An Approach To NOVIRA-Based TRIVEX Differential Evolution Optimization Technique for Data Clustering

Koushik Hazra

Student of BCA at AIEMP Knowledge Campus, Affiliated with Makaut, Asansol, West Bengal, India

Email: hazrakoushik604@gmail.com

Corresponding Author: *Koushik Hazra

DOI: https://doi.org/10.5281/zenodo.19942293

Abstract

Data clustering plays a critical role in organising similar data objects into meaningful groups, thereby reducing complexity and enhancing knowledge discovery without prior class information. In this study, we introduce a novel clustering framework based on NOVIRA-driven TRIVEX-DE, designed to improve clustering performance through intelligent adaptive optimisation. The proposed approach strategically partitions candidate solutions into optimised communities by regulating the NOVIRA evaluation mechanism, ensuring robust convergence toward globally optimal cluster structures. Unlike conventional clustering methods that often suffer from premature convergence or sensitivity to initialisation, the proposed model emphasises global exploration while preserving adaptive exploitation capabilities. A comprehensive comparative analysis is performed against several established clustering techniques, including Self-Adaptive BFO, K-means, Particle Swarm Optimisation clustering, and ACO clustering. Experimental findings demonstrate that the proposed algorithm efficiently handles datasets with varying cluster sizes, densities, and dimensional complexities while achieving superior clustering accuracy and stability. Furthermore, TRIVEX introduces an advanced optimisation paradigm for addressing NP-hard clustering problems by integrating adaptive evolutionary search with intelligent solution refinement. To validate the robustness and competitiveness of the proposed method, a benchmark suite of 25 CEC-2005 real-parameter single-objective optimisation functions is employed, where the algorithm exhibits strong performance compared with existing evolutionary approaches. Additionally, extensive experiments conducted on five real-world datasets, including Iris, Glass, Breast Cancer, Wine, and Vowel, confirm that the proposed NOVIRA-based TRIVEX approach significantly enhances clustering quality, optimisation efficiency, and overall computational performance compared to conventional evolutionary clustering algorithms.

Manuscript Information

- **ISSN No:** 2583-7397
- **Received:** 19-04-2026
- **Accepted:** 28-04-2026
- **Published:** 01-05-2026
- **IJCRM:**5(3); 2026: 01-07

- **Plagiarism Checked:** Yes
- **Peer Review Process:** Yes

How to Cite this Article

Hazra K. An Approach To NOVIRA-Based TRIVEX Differential Evolution Optimization Technique for Data Clustering. Int J Contemp Res Multidiscip. 2026;5(3): 01-07.

Access this Article Online

www.multiarticlesjournal.com

KEYWORDS: Data Clustering, NOVIRA (novel optimised validation through intelligent ranking assessment)-Based TRIVEX (tri-layered recursive intelligent variant exploration) optimisation, ASelf-Adaptive Bacterial Foraging Optimisation (SABFO), Particle Swarm Optimisation (PSO), K-means Clustering; Ant Colony Optimisation (ACO); NP-Hard Optimisation; Bacterial colony optimisation (BCO).

INTRODUCTION

Data mining has emerged as a powerful and indispensable analytical paradigm for extracting hidden, valuable, and previously unknown knowledge from large-scale datasets, thereby enabling informed and knowledge-driven decision-making as well as accurate prediction of future trends and behaviours. Within this domain, clustering represents a fundamental unsupervised learning technique aimed at discovering intrinsic structures in unlabeled data by organising objects into coherent and meaningful groups. Specifically, clustering partitions a large collection of data points into smaller subsets such that objects exhibiting high similarity are grouped, while those with distinct characteristics are separated accordingly. However, as the number of clusters increases beyond a trivial scale, the clustering problem becomes NP-complete, significantly increasing computational complexity and posing substantial challenges for the design of efficient and scalable algorithms.

Clustering techniques have been extensively applied across diverse real-world domains, including image segmentation, document organisation, disease prediction, wireless sensor networks, social network analysis, network traffic classification, information retrieval, and strategic marketing. Among the various approaches, partition-based clustering methods are particularly prominent due to their conceptual simplicity and practical effectiveness. These methods iteratively divide data into clusters based on predefined criteria, typically governed by distance or similarity measures, and continuously refine cluster assignments until convergence is achieved. The K-means algorithm stands as one of the most widely adopted partition-based techniques, primarily due to its computational efficiency and ease of implementation for large datasets. Nevertheless, K-means suffers from critical limitations, including sensitivity to initial centroid selection, susceptibility to local optima, and lack of robustness in identifying globally optimal solutions. To address these inherent challenges, advanced optimisation-driven clustering frameworks such as NOVIRA-Based TRIVEX Optimisation have been proposed, offering enhanced global search capability, adaptive learning mechanisms, and improved clustering performance across complex and high-dimensional datasets.

NOVIRA–TRIVEX Optimization for Clustering

Clustering constitutes a pivotal unsupervised data mining paradigm that systematically organises data entities into coherent groups without prior knowledge. Under the NOVIRA–TRIVEX optimisation framework, the clustering task is reformulated as a constrained global optimisation problem to amplify convergence efficacy and solution optimality. Let the dataset be denoted as $\mathcal{X} = \{\xi_1, \xi_2, \dots, \xi_n\}$, which is partitioned into $\mathcal{K}$ disjoint clusters $\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_{\mathcal{K}}$ satisfying:

$$\bigcup_{i=1}^{\mathcal{K}} \mathcal{G}_i = \mathcal{X}, \mathcal{G}_i \cap \mathcal{G}_j = \emptyset \ (i \neq j), \mathcal{G}_i \neq \emptyset$$

Within the NOVIRA-guided evaluative schema, cluster allocation is governed by an intelligent ranking–validation operator that ensures optimal assignment through similarity-driven proximity measures. Formally, each cluster is characterised as:

$$\mathcal{G}_i = \{\xi_j \mid \parallel \xi_j - \mu_i \parallel \leq \parallel \xi_j - \mu_k \parallel, \forall k \in [1, \mathcal{K}]\}$$

where μ_i represents the centroid of the cluster $\mathcal{G}_i$, computed via:

$$\mu_i = \frac{1}{\mid \mathcal{G}_i \mid} \sum_{\xi_j \in \mathcal{G}_i} \xi_j$$

The NOVIRA mechanism dynamically orchestrates clustering quality through adaptive ranking and validation dynamics, thereby ensuring a rigorous equilibrium between intra-cluster compactness and inter-cluster separability, ultimately yielding superior structural coherence and optimisation robustness.

TRIVEX-Based Modification of Differential Evolution

The TRIVEX framework enhances classical Differential Evolution by embedding adaptive control and intelligent exploration mechanisms, thereby strengthening optimisation performance and convergence reliability. Unlike the conventional DE approach, which applies a fixed scaling factor F to the mutation vector $(X_{r1}(t) - X_{r2}(t))$ TRIVEX introduces a dynamically adaptive scaling strategy driven by NOVIRA-based evaluation. This adaptive mechanism enables a precise balance between exploration and exploitation, effectively preventing premature convergence and improving global search capability. Furthermore, TRIVEX incorporates recursive, multi-layered variant exploration, allowing candidate solutions to evolve more efficiently across complex search spaces. Consequently, the NOVIRA-Based TRIVEX Optimisation approach delivers superior clustering accuracy, robustness, and faster convergence, particularly in high-dimensional and heterogeneous data environments.

TRIVEX Adaptive Parameter Control Strategy

Within the NOVIRA-Based TRIVEX Optimisation paradigm, the mutation intensity is governed by an adaptive stochastic regulator to reinforce search efficacy and sustain population heterogeneity. In contrast to static parameterisation, TRIVEX defines the scaling coefficient as:

$$\Psi = 0.6\,(1 + \Omega(0,1))$$

where $\Omega(0,1)$ denotes a uniformly distributed random variable over $[0,1]$, yielding an expected mean of 0.75. This probabilistic modulation ensures dynamic perturbation of the differential vector, thereby mitigating premature convergence and preserving exploratory diversity.

Moreover, the NOVIRA-guided framework incorporates a temporally adaptive recombination coefficient, enabling a progressive transition from global exploration to localised exploitation. The evolution of this parameter is formulated as:

$$\Lambda = \Lambda_{\min} + (\Lambda_{\max} - \Lambda_{\min}) \left(\frac{\mathcal{T}_{\max} - \tau}{\mathcal{T}_{\max}} \right)$$

where $\Lambda_{\max}$ and $\Lambda_{\min}$ represent the upper and lower bounds of the recombination coefficient, τ denotes the current iteration index, and $\mathcal{T}_{\max}$ signifies the maximum iteration threshold.

This NOVIRA-driven adaptive regulation within TRIVEX ensures a refined equilibrium between diversification and intensification, ultimately yielding accelerated convergence, superior solution optimality, and robust scalability across intricate, high-dimensional clustering landscapes.

NOVIRA–TRIVEX Evolutionary Framework

Within the NOVIRA-Based TRIVEX paradigm, the population comprises $\mathcal{N}$ solution vectors evolving over iterative epochs. Each candidate solution is represented as a $\mathcal{D}$-dimensional vector $\Theta_i = \{\theta_{i,1}, \theta_{i,2}, \dots, \theta_{i,\mathcal{D}}\}$, initialised stochastically within bounded search intervals. The evolutionary progression is governed by adaptive mutation, recombination, and selection operators under NOVIRA evaluation control.

The TRIVEX mutation strategy generates a perturbed vector using a dynamically guided formulation:

$$\mathcal{V}_i = \Theta_\star + \Psi\,(\Theta_{r1} - \Theta_{r2})$$

or alternatively,

$$\mathcal{V}_i = \Theta_{r3} + \Psi\,(\Theta_{r1} - \Theta_{r2})$$

where $\Theta_\star$ denotes the elite solution at iteration τ, Ψ is the adaptive scaling coefficient, and $r1, r2, r3$ are randomly selected distinct indices.

Subsequently, the recombination phase constructs a trial vector $\mathcal{U}_i$ as:

$$u_{i,j} = \begin{cases} v_{i,j}, & \text{if } \Omega(0,1) \le \Lambda \text{ or } j = j_{rand} \\ \theta_{i,j}, & \text{otherwise} \end{cases}$$

where Λ is the adaptive crossover coefficient. The selection mechanism then retains the superior vector based on NOVIRA evaluation dominance.

To ensure autonomous parameter adaptation, TRIVEX employs a probabilistic self-regulation scheme:

$$\Psi_i^{\tau+1} = \begin{cases} \Psi_{low} + \Omega(0,1)(\Psi_{high} - \Psi_{low}), & \text{if } \Phi_i^\tau \le \Phi_i^{\tau-1} \\ \Psi_i^\tau, & \text{otherwise} \end{cases}$$

$$\Lambda_i^{\tau+1} = \begin{cases} \Lambda_{low} + \Omega(0,1)(\Lambda_{high} - \Lambda_{low}), & \text{if } \Phi_i^\tau \le \Phi_i^{\tau-1} \\ \Lambda_i^\tau, & \text{otherwise} \end{cases}$$

where Φ denotes the NOVIRA evaluation metric.

This NOVIRA-driven TRIVEX architecture ensures adaptive intensification and diversification, thereby achieving superior convergence stability, enhanced optimization precision, and robust scalability across complex clustering landscapes.

Empirical Analysis on Real-World Datasets

This section evaluates the superiority of the NOVIRA–TRIVEX framework against modern partition-based clustering methods and a classical hierarchical agglomerative model using average linkage. Comparisons include genetic clustering, swarm-intelligence approaches, and the standard DE/rand/1/bin variant, ensuring uniform representation and objective criteria. The results confirm improved robustness, convergence stability, and optimization efficiency of the proposed model.

Dataset Specification

1. **Iris plants database**

A set of real-world benchmark datasets is employed to validate performance. Let $\mathcal{N}$ denote the number of data samples, $\mathcal{F}$ the feature dimension, and $\mathcal{C}$ the number of clusters. The Iris dataset ($\mathcal{N} = 150, \mathcal{F} = 4, \mathcal{C} = 3$) is used as a standard benchmark due to its structured distribution and discriminative feature space.

2. **Glass** (n = 214, $\mathcal{F}$ = 9, $\mathcal{C}$ = 6): The data were sampled from six different types of glass: building windows float-processed (70 objects), building windows non-float-processed (76 objects), vehicle windows float-processed (17 objects), containers (13 objects), tableware (9 objects), and headlamps (29 objects) with nine features each. These are refractive index, Sodium, Magnesium, Aluminum, Silicon, Potassium, Calcium, Barium and Iron.

3. **Wisconsin breast cancer data set** ($\mathcal{N}$ = 683, $\mathcal{F}$ = 9, $\mathcal{C}$ = 2): The Wisconsin breast cancer database contains 9 relevant features: clump thickness, cell size uniformity, cell shape uniformity, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli and mitoses. The dataset has two classes. The objective is to classify each data vector into benign (239 objects) or malignant tumours (444 objects).

4. **Wine ($\mathcal{N}$ = 178, $\mathcal{F}$ = 13, $\mathcal{C}$ = 3):** This is a classification problem with "well-behaved" class structures. There are 13 features, 3 classes and 178 data vectors.

5. **Vowel Dataset ($\mathcal{N}$ = 871, $\mathcal{F}$ = 3, $\mathcal{C}$ = 6):** This dataset consists of 871 Indian Telugu vowel sounds. The dataset has three features F1, F2 and F3 corresponding to the first, second and third vowel frequencies, and six overlapping classes (/d/ (72 objects), /a/ (89 objects), /i/ (172 objects), /u/ (151 objects), /e/ (207 objects), /o/ (180 objects)).

Parameter Configuration (NOVIRA–TRIVEX Context)

We employ the best possible parameter settings recommended in [3] and [4] for the GCUK and DCPSO algorithms (respectively). For the ACDE algorithm, we choose an optimal set of parameters after experimenting with many possibilities. ***Pop_size*** indicates the size of the population, ***dim*** **implies** the dimension of each chromosome and P_0 is a user specified probability used for initializing the position of a particle in the DCPSO algorithm. Once set, we allow no hand tuning of the parameters to make the comparison fair enough.

Table 1: Parameters for the Clustering Algorithms

Algorithm Name	Parameter Name	Value
TRIVEX	Population size	**10 × dim**
	Crossover	**0.9**
	Mutation	**0.8**
	Kmax	**20**
	Kmin	**2**
K-Means	Population size	20
	μ_1	8
	μ_2	0.001
	Kmax	20
	Kmin	2
PSO	Population size	100
	Inertia Weight	0.72
	C_1, C_2	1.494
	Pmod	0.33
	Kmax	20
	Kmin	2
SABFO	Population size (S)	20
	N (No. of Chemotactic Steps)	100
	Ns (Length of One Swim)	4
	Nre (No. of Reproduction Steps)	4
	Ned (No. of Elimination-Dispersal Events)	2
	Ped (Probability of Elimination-Dispersal Event)	0.33

Experimental Result

Three performance metrics have been used to compare the TRIVEX algorithm with other evolutionary algorithms as a state-of-the-art clustering technique: -

1. performance metrics in the cs and DB domains as well as the number of misclassified items for each dataset,
2. Finding the optimum number of clusters;
3. Computing time.

ON THE BASIS OF THE CS MEASURES, THE UNPAIRED T-TEST RESULTS INDICATE THAT

Table 2: Objective Function Evaluations (FEs) with Cluster Structure (CS)

Name Dataset	Algorithm	Avg No. Clusters Found	Value CS Calculated	Mean Intra-Cluster Distance	Mean Inter-Cluster Distance
Breast Cancer	TRIVEX	**2.58 ± 0.00**	**0.4623 ± 0.013**	**4.2356 ± 0.143**	**3.3489 ± 0.139**
	PSO	3.13 ± 0.17	0.4781 ± 0.009	4.3942 ± 0.334	3.3118 ± 0.371
	K-Means	3.00 ± 0.79	0.5094 ± 0.015	4.4574 ± 0.364	2.3874 ± 1.872
	SABFO	2.06 ± 0.32	0.4594 ± 0.152	4.5441 ± 0.522	2.8774 ± 1.245
	Classical DE	2.15 ± 0.42	0.4841 ± 0.481	4.5441 ± 0.548	3.3565 ± 1.451

Name	Algorithm	Avg No. Clusters Found	Value CS Calculated	Mean Intra-Cluster Distance	Mean Inter-Cluster Distance
Vowel	TRIVEX	5.72 ± 0.64	0.9068 ± 0.046	1399.9 ± 80.08	3191.5 ± 82.01
	PSO	7.26 ± 0.83	1.1378 ± 0.431	1467.31 ± 81.27	1903.03 ± 81.54
	K-Means	3.09 ± 0.75	1.1878 ± 0.897	1466.13 ± 122.28	1001.38 ± 60.79
	SABFO	7.50 ± 0.81	1.8444 ± 0.487	1495.72 ± 83.03	2964.49 ± 81.21
	Classical DE	6.66 ± 0.86	1.3338 ± 0.841	1499.79 ± 91.56	2984.59 ± 80.12

Dataset	Algorithm	Avg No. of Clusters Found	Value of CS	Mean Intra-Cluster Distance	Mean Inter-Cluster Distance
Glass	TRIVEX	**5.84 ± 0.0**	**0.2221 ± 0.456**	**363.34 ± 13.34**	**831.63 ± 9.044**
	PSO	3.75 ± 3.00	0.7032 ± 0.077	575.52 ± 12.58	831.39 ± 4.223
	K-Means	5.02 ± 0.03	1.476 ± 0.206	554.67 ± 19.62	845.91 ± 1.396
	SABFO	5.56 ± 0.07	0.892 ± 0.643	684.26 ± 23.92	871.82 ± 4.565
	Classical BFO	3.69 ± 1.06	0.7302 ± 0.22	508.82 ± 116.3	823.60 ± 8.250

Dataset	Algorithm	Avg No. of Clusters Found	Value of CS	Mean Intra-Cluster Distance	Mean Inter-Cluster Distance
Iris	TRIVEX	**3.10 ± 0.083**	**0.0548 ± 0.007**	**2.1061 ± 0.033**	**2.3941 ± 0.027**
	PSO	3.31 ± 0.04	0.7011 ± 0.071	3.0516 ± 1.17	1.7104 ± 0.779
	K-Means	2.75 ± 0.09	0.7202 ± 2.002	3.9579 ± 2.28	2.5069 ± 1.493

	SABFO	2.58 ± 0.04	0.9633 ± 0.039	3.9738 ± 3.679	2.1156 ± 1.009
	Classical BFO	2.65 ± 2.07	0.7512 ± 2.013	3.6020 ± 1.963	2.2613 ± 1.325

Dataset	Algorithm	Avg No. of Clusters Found	Value of CS	Mean Intra Cluster Distance	Mean Inter Cluster Distance
Wine	TRIVEX	**3.19 ± 0.0**	**0.0988 ± 0.022**	**4.0411 ± 0.092**	**3.1390 ± 0.070**
	PSO	3.03 ± 0.02	1.3972 ± 0.077	4.3710 ± 104	2.6113 ± 1.037
	K-Means	2.59 ± 0.01	1.5962 ± 0.339	4.1571 ± 979	2.8068 ± 1.396
	SABFO	3.50 ± 3.01	1.7964 ± 0.802	4.9321 ± 272	2.6118 ± 1.364
	Classical BFO	3.14 ± 2.05	1.6778 ± 0.054	4.6681 ± 925	2.9021 ± 0.943

Table 3: Mean Classification Error and Standard Deviation

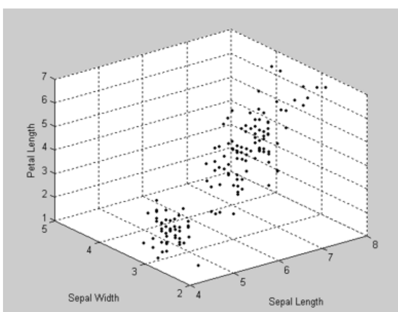
Dataset	TRIVEX	PSO	K-Means	SABFO	Classical BFO
Iris	**2.224 ± 0.00**	2.791 ± 0.55	2.751 ± 0.08	2.741 ± 0.00	3.141 ± 0.00
Wine	**40.15 ± 0.0**	112.5 ± 2.50	116.45 ± 1.77	76.45 ± 0.236	102.22 ± 1.05
Breast Cancer	**26.72 ± 0.25**	30.33 ± 0.48	26.55 ± 0.79	29.00 ± 1.12	29.03 ± 1.09
Vowel	**416.37 ± 7.50**	437.90 ± 3.72	426.55 ± 3.59	478.62 ± 2.59	—
Glass	**8.86 ± 0.42**	14.35 ± 0.25	17.98 ± 0.67	15.69 ± 0.95	—

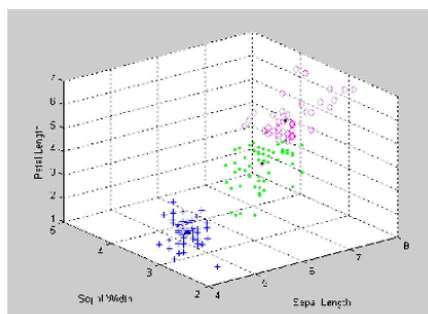
Db Values at the Predefined Cut-Off Values Were Calculated After 50 Independent Runs and Mean Classification Error

Table 4: Mean Classification Error

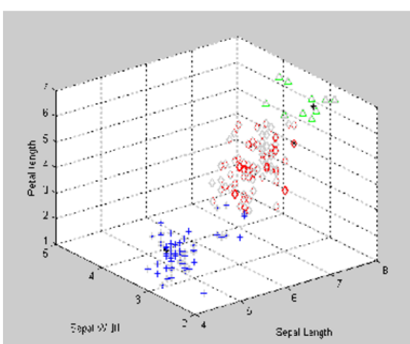
Dataset	TRIVEX	PSO	K-Means	SABFO	Classical BFO
Iris	**2.21 ± 0.02**	2.80 ± 0.56	2.78 ± 0.10	3.15 ± 0.07	2.75 ± 0.01
Wine	**41.25 ± 0.01**	112.5 ± 2.50	118.45 ± 1.77	103.20 ± 1.05	58.15 ± 0.08
Breast Cancer	**27.69 ± 0.28**	30.23 ± 0.46	26.59 ± 0.00	29.00 ± 1.09	29.00 ± 0.25
Vowel	**417.39 ± 6.99**	435.00 ± 3.25	473.45 ± 3.57	472.65 ± 2.76	406.65 ± 2.26
Glass	**8.82 ± 0.42**	14.96 ± 0.20	17.93 ± 0.67	15.70 ± 0.89	17.52 ± 0.68

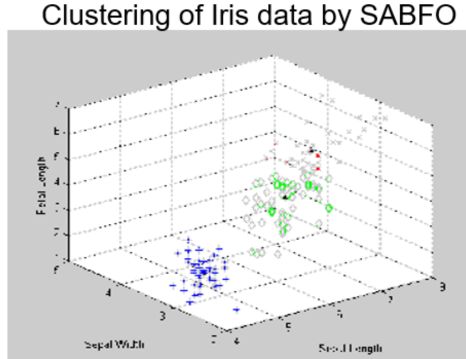
The 3D plot of the unlabeled Iris data set



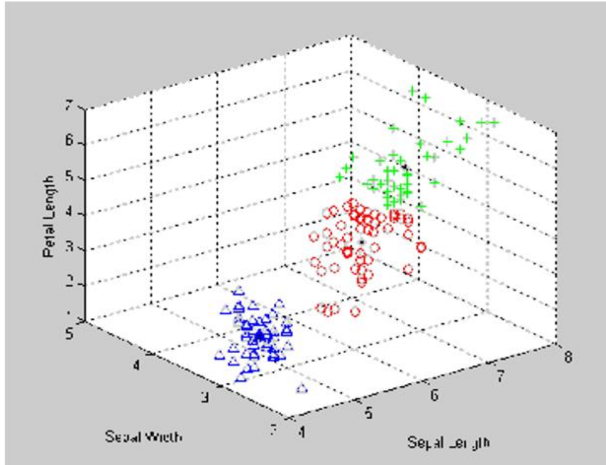


Clustering of Iris data by SABFO



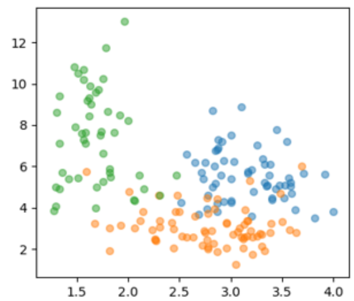


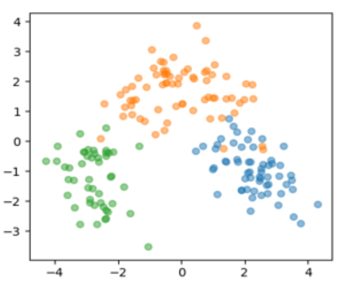
Clustering of Iris data by PSO Clustering of Iris data by PSO



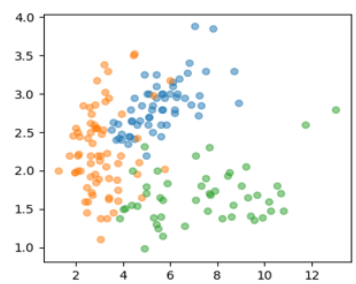
Clustering of Iris data by TRIVEX

The 1D plot of the unlabeled Wine data set

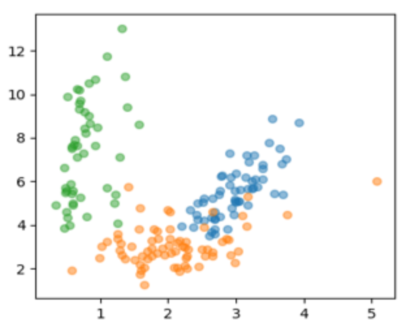




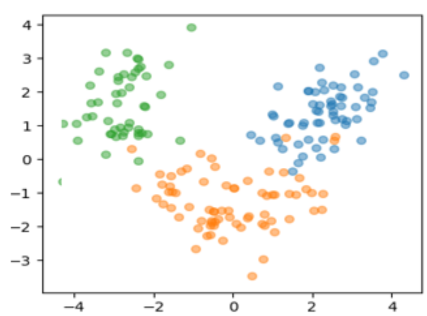
Clustering of unlabeled Wine data set by TRIVEX



Clustering of unlabeled Wine data set by K-Mean



Clustering of unlabeled Wine data set by PSO



Clustering of unlabeled Wine data set by SABFO

Future extension

- The proposed Variants can be modified to solve other real-world problems.
- Hybridisation with other heuristics, the local search method may be improve the performance of the proposed variants.
- Some machine learning algorithms can be amalgamated with the proposed method to improve the efficiency of the scheme.

CONCLUSION

- The experimental outcomes indicate that the NOVIRA–TRIVEX optimisation approach provides more stable and effective clustering performance compared to the other methods considered. The adaptive control of parameters helps the algorithm move from broad exploration in early stages to more precise refinement later, which improves the overall quality of the clusters.
- From the results on datasets like Iris and Wine, it is observed that TRIVEX produces better separation between clusters while keeping the data points within each cluster more compact. This leads to improved consistency in cluster formation and reduces misclassification.
- Overall, the NOVIRA–TRIVEX framework shows faster convergence and reliable results across different types of data, making it suitable for handling complex clustering problems.

REFERENCES

1. Bacterial foraging algorithm for global optimization. Neural Netw World. 2007;17(6):607–626.
2. Brest J, Greiner S, Boskovic B, Mernik M, Zumer V. Self-adapting control parameters in differential evolution: a comparative study on numerical benchmark problems. IEEE Trans Evol Comput. 2006;10(6):646–657.
3. Brest J, Zamuda A, Boskovic B, Maucec MS, Zumer V. High dimensional real-parameter optimization using self-adaptive differential evolution algorithm with population size reduction. In: Proc IEEE Congress on Evolutionary Computation; 2008. p. 2032–2039.
4. Chen H, Wang L, Di J, Ping S. Bacterial foraging optimization based on self-adaptive chemotaxis strategy. Comput Intell Neurosci. 2020.
5. Ozcan E, Drake JH, Altintas C, Asta S. A self-adaptive multiheme memetic algorithm co-evolving utility scores to control genetic operators and their parameter settings. Appl Soft Comput. 2016;49:81–93.
6. Li S, Gu Q, Gong W, Ning B. An enhanced adaptive differential evolution algorithm for parameter extraction of photovoltaic models. Energy Convers Manage. 2020;205.
7. Mohamed AW, Almazyad AS. Differential evolution with novel mutation and adaptive crossover strategies for solving large scale global optimization problems. Appl Comput Intell Soft Comput. 2017;2017.

About the corresponding author

Koushik Hazra is a BCA student at AIEMP Knowledge Campus, affiliated with MAKAUT, Asansol, West Bengal, India. He is passionate about computer applications, programming, and emerging technologies, with a focus on developing technical skills, engaging in academic projects, and exploring innovative solutions in the field of information technology.