Performance Analysis of Single and Multi-Objective Metaheuristic Algorithms for Data Clustering Problems

Arvinder Kaur[1], Pardeep Kumar[2], Prashant Yadav[3] and Geeta Yadav[4]

{er.arvinderdhillon@gmail.com[1], pardeep.kumar3@capitalone.com[2], prashantyadav786@gmail.com[3], yadavgeeta172@gmail.com[4]}

Chitkara University, Mohali, Punjab, India[1],
Capital One Financial Corp, USA[2],
Tata Consultancy Services Ltd., Plano, Texas, USA[3]
School of Pharmaceutical Sciences, CGC university, Mohali, Jhanjeri, India[4]

Abstract. Data clustering is a fundamental approach in unsupervised machine learning that can be utilized to discover patterns with unlabelled data. The variety of algorithms have been adopted to determine the more accurate solutions for data clustering problems. In literature various single and multi-objective metaheuristics algorithms are explored to handle the data clustering problems as well as capable to deal with the high-dimensional and complex data. In this paper, an attempt is made to evaluate and compare the single-objective and multi-objective metaheuristics algorithms for data clustering. Apart from this, this paper also explores the algorithmic design, performance metrics, and empirical outcomes. The simulation results demonstrated that single-objective approaches may be sufficient for simpler clustering tasks, but the multi-objective algorithms often provide superior results in terms of cluster validity, flexibility, and application to real-world, multi-faceted clustering problems.

Keywords: Metaheuristics, Single-objective, Multiobjective, Pareto Front

1 Introduction

The expanding panorama of data driven applications deal with several issues. Finding the patterns from such large amount of data is tedious. Unsupervised learning technique can be applied in identifying the patterns. Clustering is one of the unsupervised learning method that groups the similar data together[1]. The inability to precisely and uniquely define a cluster contributes to the difficulty of data clustering. It is well accepted that when previous knowledge about the underlying data distributions is poorly defined, clustering becomes an ill-posed problem. Figure 1 shows the methodology for clustering the objects using single objective that is based on single criteria. Regardless of the clustering criterion employed, a basic shortcoming of all clustering algorithms is their inability to recognize clusters of various sizes and forms.

IICTDS 2025, December 19-20, Chandigarh, India
Copyright © 2026 EAI
DOI 10.4108/eai.19-12-2025.2363564

Exploratory pattern analysis has long sought to identify most or all of the clusters (of arbitrary forms) in the data [2]. In order to accomplish this, a framework must be established which amalgamates the results of multiple clustering methods employed on the same data, each of which corresponds to a different clustering purpose [3].

An associated issue is that almost all clustering algorithms make the assumption that the clustering criterion is uniform over the whole feature space. Consequently, all of the identified clusters frequently have comparable data densities and shapes. If there are multiple clusters of various shapes in different parts of the feature space, a single clustering method cannot discover them all because the underlying criterion could not match the distribution of data across the feature space. To solve the clustering problems, single objective clustering and multiobjective clustering can be employed.

The process of clustering data by optimizing a single criterion or objective function is known as single objective clustering[4]. It's among the most widely used strategies in both evolutionary and conventional clustering systems. The goal is to minimize or maximize a single objective function. It focuses on intra-cluster compactness, inter-cluster separation. It can also be represented as a single scalar value. Compared to its multiobjective equivalents, single objective optimization is simple to apply, performs calculations quickly, and concentrates on a single area of improvement. The single objective optimization has limited perspective and may overlook trade-offs between conflicting criteria. Also the chosen objective might not capture all aspects of good clustering and also suffers from problem of local optima. In order to produce better-quality and more balanced clusterings, multiobjective clustering entails simultaneously maximizing two or more competing objectives [5].

In complex or real-world datasets, when a single aim might not adequately capture all clustering needs, it is quite useful. As compared to single objective, multiobjective clustering considers multiple goals such as maximizing inter cluster separation, minimizing intra-cluster distance, balancing cluster sizes and minimizing number of clusters if required.

These provide the balanced solutions considering trade-offs between conflicting criteria. The multiobjective clustering results in a set of optimal solutions known as Pareto Optimal (Pareto Front) [6]. Thus these are capable of handling complex and high-dimensional clustering scenarios. Figure 2 shows the process of multiobjective optimization to generate pareto solutions.

1.2 Organization of Paper

The research paper is organized as follows: Section I provides the introduction to the single and multiobjective metaheuristics, section II provides the related works, section III gives the comparison of the single and multiobjective metaheuristics algorithms, whereas results are discussed in section IV. The conclusion and future directions are given in section V.

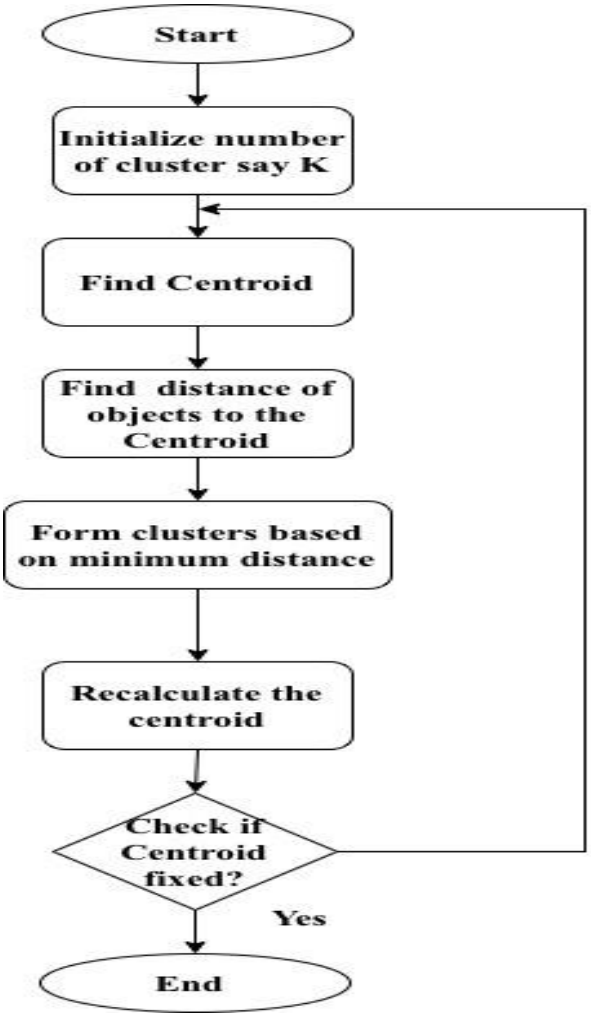


Fig. 1. Illustration of Methodology for Clustering using single objective

2 Related Works

This section provides the summary of single and multiobjective clustering algorithms.

A hybrid algorithm was proposed using Chaos Optimization and Flower Pollination over K-means by Kaur, Pal and Singh [7]. The algorithm is referred to as Chaotic FPA (CFPA). Singh[8] proposed chaotic sequence-guided harris hawks optimization (CHHO). The random number generator of original Harris Hawk Optimization has been replaced by chaotic number to enhance global search in CHHO. Abasi et al.[9] proposed hybrid algorithm H-MVO using

multi-verse optimizer (MVO) and k-means. The global search ability and cluster partition are improved with the proposed H-MVO.

Cho and Nyunt[10] introduced enhanced differential evolution to generate better clustering solutions including faster convergence. The search behavior of DE has been improved by adopting modified mutation strategy. Ahmadi, Ekbatanifard and Bayat [11] proposed modified grey wolf optimizer (MGWO). It has helped in improving the exploration and exploitation in finding the best solution.

Kaur & Kumar [12] presented a new Water Wave Optimization algorithm for data clustering and addresses the absence of global optima and premature convergence. Kuwil et al. [13] proposed a new distance-based algorithm for clustering named as critical distance clustering algorithm. Mansuet and Schoen [14] proposed memetic differential evolution (MDE) for Euclidean minimum sum of squares clustering (MSSC).

For categorical data, Saha and Maulik merged differential equation algorithm and fuzzy concept for multiobjective clustering[15]. To produce better clustering results, amulti objective K-means genetic algorithm has been presented by Peng et al. [16]. Kuo and Zulvia [17] extended original gradient evolution (GE) algorithm for solving multi-objective problem. It considered two objective functions- sum of squared error between cluster (SSB) and sum of distance within cluster (SSW). The Pareto rank assignment is used for sorting vectors based on fitness value. The final clustering outcome is computed using K-means algorithm.

An immune based multi objective fuzzy clustering algorithm was proposed by Liu et al. [18] using two mutation operators for optimal partitioning as well as appropriate number of clusters. Liu et al. [19] developed a multi objective evolutionary algorithm based on different distance measures. Azevedoet al. [22] introduced Multi-objective clustering algorithm for obtaining optimal number of clusters and minimizing the intracluster distance and maximizing inter-clustering for finding partitions among the elements. Abdollahzadeh & Gharehchopogh [23] proposed three different solutions for feature selection that consist of using the Harris Hawks Optimization (HHO) algorithm for first solution, the Fruitfy Optimization Algorithm (FOA) has been multiplied for the second solution, and in the third solution, these two solutions are hybridized and named as MOHHOFOA. Sharma et al. [24] proposed a Butterfly Optimization Algorithm (BOA) for solving multiobjective problems. The proposed algorithm utilizes dominated sorting and crowding distance measure. Kaur and Kumar [31] presented a multiobjective vibrating particle system algorithm for data clustering that utilized the intra-cluster variance and connectedness as objective functions.

3 Comparison of Single and Multiobjective Metaheuristics Algorithm

The single-objective clustering algorithms work on optimizing one criteria such as minimizing the intra-cluster distance whereas multi-objective algorithms balance between the multiple conflicting objectives like compactness and separation simultaneously. Both techniques have wide variety of applications. The single objective algorithms such as k-means, hierarchical clustering can be used to group patients for disease subtyping and resource allocation. Moreover multi-objective clustering

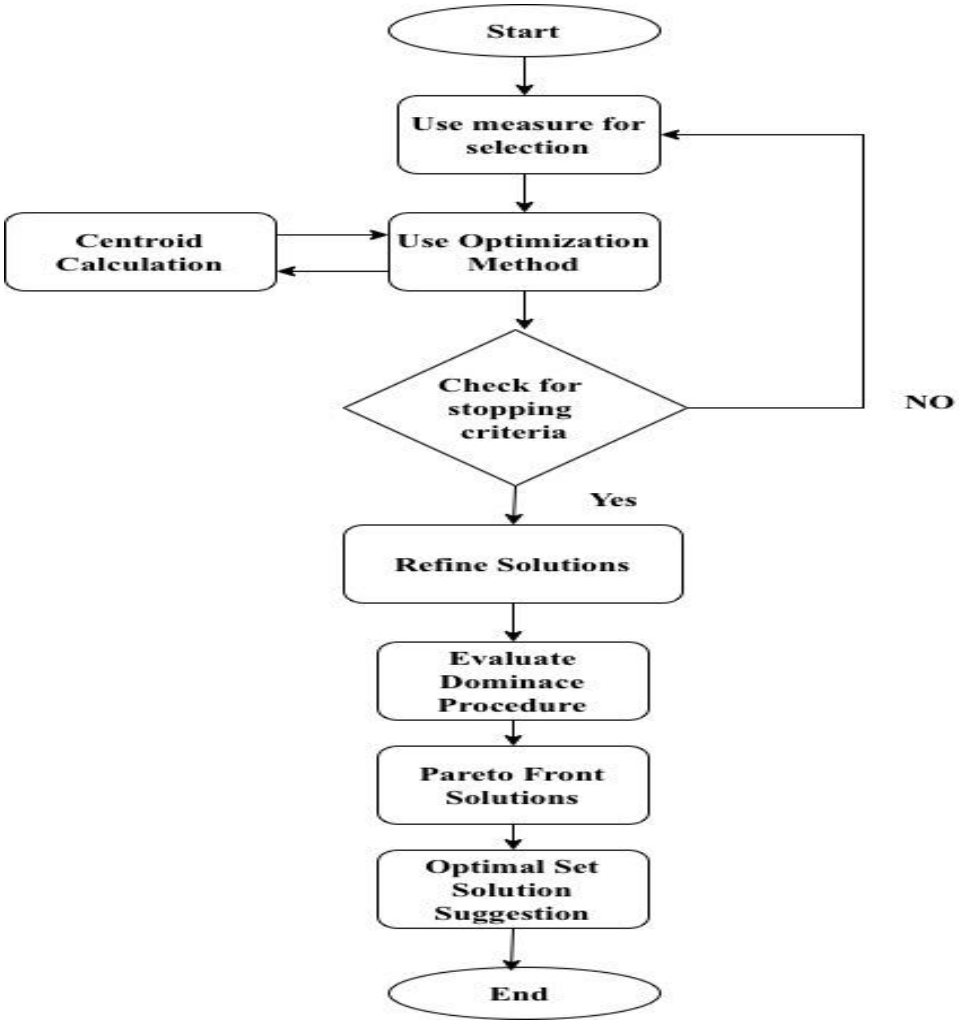


Fig. 2. Illustration of Methodology for Clustering using Multi objective

algorithms can be helpful in handling noisy health records and creating stable patient clusters for genomic and clinical features.

Based upon the different parameters such as number of objective functions optimized, solutions provided and their behaviour the comparison between the Single and Multiobjective Metaheuristics is shown in Table 1.

Table 1. Single and Multiobjective Metaheuristics

Parameter	Single-Objective Metaheuristics	Multiobjective Metaheuristics
Objectives optimized	One (e.g., compactness)	Two or more (e.g., compactness & separation)
Result type	Single solution	Set of Pareto-optimal solutions
Flexibility	Limited	High
Computational cost	Lower	Higher
Interpretability	Less (single solution)	More (trades between objectives shown)
Suitability	Simpler datasets/tasks	Complex, real-world datasets
Example algorithms	PSO-k-means, GA-k-means	NSGA-II, MOEA/D, MOPSO

Empirical results indicate that multiobjective approaches yield higher-quality clusters when multiple criteria are of interest. For example, NSGA-II[20] not only finds compact clusters but ensures well-separated groupings, catering to real-world requirements such as multi-attribute customer segmentation. Single-objective algorithms [21]remain useful for tasks where a clear, primary goal exists, or computational simplicity is a necessity.

4 Results

Popular benchmarking datasets as shown in Table 2 are used to evaluate algorithm performance. The experimental setup consists of 30 independent runs for each algorithm on each dataset. The different metaheuristics algorithms such as Particle Swarm Optimization (PSO)[25], K-means [26], Genetic Algorithm (GA) [27] have been considered for single objective. While for multiobjective metaheuristics, NSGA-II [28], MOEA/D [29], MOPSO[30] have been considered.

Table 2. Description of Benchmark datasets

Datasets	Clusters (K)	Instances	Dimension
Iris	3	150	4
CMC	3	1473	9
Dermatology	6	358	34
Liver	2	345	6
Glass	7	214	9

Single-objective metaheuristics frequently converge faster and provide satisfactory clustering for datasets with clear cluster boundaries. The table 3 gives the results of the Single Objective

metaheuristic for Data Clustering. From the results it is observed that on Iris dataset PSO gives highest accuracy and lowest Std Dev , GA performs well with slight variation whereas K-means lags behind. It is noticed that PSO is stable and effective. On CMC dataset, GA performs best among K-means and PSO.

While on dermatology dataset GA performs best while K-means and PSO gives comparable results. Further on Liver Disorder dataset, the best accuracy is given by PSO than GA and K-means. PSO also shown its best performance on glass dataset, whereas GA give moderates results but better than K-means. It is seen that PSO gives high accuracy for the all the considered datasets except CMC. GA has given consistent results. K-means is weakest among the three and gives lower accuracy compared to GA and PSO

Table 3. Results of Single Objective Metaheuristic for Data Clustering

	K-Means		GA		PSO	
DataSet	**Accuracy**	**Std Dev**	**Accuracy**	**Std Dev**	**Accuracy**	**Std Dev**
Iris	89.5	2.1	91.2	1.9	95.67	1.23
CMC	58.3	2.8	60.75	2.6	53.45	4.12
Dermatology	95.1	1.4	96.2	1.2	95.12	2.05
Liver	64.2	3.7	66.4	3.4	68.25	3.88
Glass	62.7	3	64.15	2.9	72.4	3.2

Multiobjective metaheuristics unfailingly provide results that balance multiple cluster validity indices. Pareto fronts demonstrate non-dominated solutions addressing trade-offs between compactness and separation. Table 4 show the results for multi-objective Metaheuristic for data clustering. The results show that among all the considered datasets MOEA/D yields stable results whereas NSGA-II and MOPSO gives comparable less accurate results.

Table 4. Results of Multi-Objective Metaheuristic for Data Clustering

	NSGA-II		MOEA/D		MOPSO	
DataSet	**Accuracy**	**Std Dev**	**Accuracy**	**Std Dev**	**Accuracy**	**Std Dev**
Iris	92.85	1.7	93.12	1.8	92.45	1.9
CMC	62.1	2.5	62.74	2.4	61.95	2.5
Dermatology	97.05	1.1	97.65	1.1	97.28	1.3
Liver	67.35	3.3	68.54	3.2	67.83	3.5
Glass	65.3	2.8	66.83	2.7	65.92	2.9

Multiobjective metaheuristics consistently deliver solutions that balance multiple cluster validity indices; Pareto fronts demonstrate non-dominated solutions addressing trade-offs between compactness and separation. Figure 3 compares the different single and multiobjective algorithms on the standard datasets based on the accuracy.

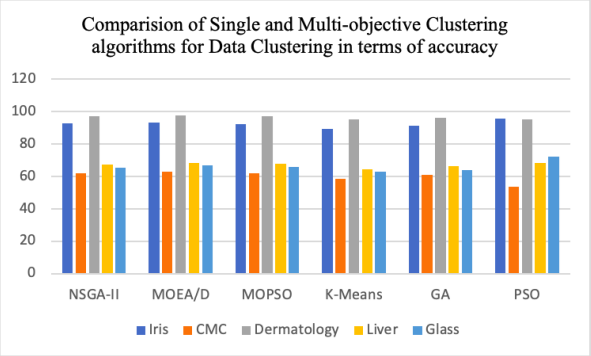


Fig. 3. Comparison of Single and Multiobjective Algorithms for Data Clustering

5 Conclusion

In this work, the comparison of single and multiobjective algorithms is given. It is observed that single-objective algorithms performs well on simple datasets but accuracy drops when complex datasets are considered. Multiobjective show better balance by optimizing compactness and separation. The multiobjective algorithms for data clustering give more stable and accurate results. From results, it is observed that single-objective are biased towards either separation or compactness. While multiobjective metaheuristics generate Pareto-optimal solutions consisting of multiple good solutions. It is concluded that both single and multiobjective metaheuristics have merit for data clustering and can be used according to the suitability of data complexity and problem requirements. The future research includes using the multiobjective methods for reducing computational loads, improving scalability and integrating with deep learning for automate feature extortion generating enhanced clustering results.

References

[1] Celebi, M. E. (Ed.): Partitional clustering algorithms. pp. 42–100. Springer, Berlin (2015)
[2] Kaur, A., Kumar, Y.: A new metaheuristic algorithm based on water wave optimization for data clustering. Evolutionary Intelligence, 15(1), 759–783 (2022)
[3] Gelbard, R., Goldman, O., Spiegler, I.: Investigating diversity of clustering methods: An empirical comparison. Data & Knowledge Engineering, 63(1), 155–166 (2007)
[4] Mirjalili, S.: Dragonfly algorithm: A new meta-heuristic optimization technique. Neural Computing and Applications, 27(4), 1053–1073 (2016)
[5] Benmessahel, I., Xie, K., Chellal, M.: Competitive multiverse optimization technique. Engineering Reports, 2(3), e12124 (2020)
[6] Hua, Y., Liu, Q., Hao, K., Jin, Y.: Survey of evolutionary algorithms for multi-objective optimization. IEEE/CAA Journal of Automatica Sinica, 8(2), 303–318 (2021)

[7] Kaur, A., Pal, S. K., Singh, A. P.: Hybridization of chaos and flower pollination algorithm over K-means. Applied Soft Computing, 97, 105523 (2020)
[8] Singh, T.: Chaotic sequence-guided Harris hawks optimizer for clustering. Neural Computing and Applications, 32(23), 17789–17803 (2020)
[9] Abasi, A. K., Khader, A. T., Al-Betar, M. A., et al.: Hybrid multi-verse optimizer with K-means for text clustering. Neural Computing and Applications, 32(23), 17703–17729 (2020)
[10] Win Cho, P. P., Soe Nyunt, T. T.: Data clustering using modified differential evolution. International Journal of Intelligent Engineering & Systems, 13(6) (2020)
[11] Ahmadi, R., Ekbatanifard, G., Bayat, P.: Modified grey wolf optimizer for clustering. Applied Artificial Intelligence, 35(1), 63–79 (2021)
[12] Kaur, A., Kumar, Y.: Metaheuristic algorithm based on water wave optimization. Evolutionary Intelligence, 15(1), 759–783 (2022)
[13] Kuwil, F. H., Shaar, F., Topcu, A. E., Murtagh, F.: Data clustering using critical distance methodology. Expert Systems with Applications, 129, 296–310 (2019)
[14] Mansueto, P., Schoen, F.: Memetic differential evolution methods for clustering. Pattern Recognition, 114, 107849 (2021)
[15] Saha, I., Maulik, U., Plewczynski, D.: Multi-objective technique for fuzzy clustering. Applied Soft Computing, 11(2), 2765–2776 (2011)
[16] Peng, P., Addam, O., Elzohbi, M., et al.: Multi-objective genetic algorithm for clustering solutions. Knowledge-Based Systems, 56, 108–122 (2014)
[17] Kuo, R. J., Zulvia, F. E.: Multi-objective cluster analysis using gradient evolution. Soft Computing, 24(15) (2020)
[18] Liu, R., Zhang, L., Li, B., et al.: Immune multi-objective fuzzy clustering algorithm. Knowledge and Information Systems, 45(1), 133–157 (2015)
[19] Liu, C., Liu, J., Jiang, Z.: Multiobjective evolutionary algorithm for community detection. IEEE Transactions on Cybernetics, 44(12), 2274–2287 (2014)
[20] Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: NSGA-II: Fast elitist multiobjective genetic algorithm. IEEE Transactions on Evolutionary Computation, 6(2), 182–197 (2002)
[21] Rai, P., Singh, S.: Survey of clustering techniques. International Journal of Computer Applications, 7(12), 1–5 (2010)
[22] Azevedo, B. F., Rocha, A. M. A., Pereira, A. I.: Multi-objective clustering algorithm. In: International Conference on Optimization and Learning, pp. 97–108. Springer (2024)
[23] Abdollahzadeh, B., Gharehchopogh, F. S.: Multi-objective optimization for feature selection. Engineering with Computers, 38(Suppl 3), 1845–1863 (2022)
[24] Sharma, S., Khodadadi, N., Saha, A. K., et al.: Advanced butterfly optimization algorithm. Journal of Bionic Engineering, 20(2), 819–843 (2023)
[25] Kao, Y. T., Zahara, E., Kao, I. W.: Hybridized approach to clustering. Expert Systems with Applications, 34(3), 1754–1762 (2008)
[26] Jain, A. K.: Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8), 651–666 (2010)
[27] Maulik, U., Bandyopadhyay, S.: Genetic algorithm-based clustering. Pattern Recognition, 33(9), 1455–1465 (2000)
[28] Liu, Q., Liu, X., Wu, J., Li, Y.: Improved NSGA-III using K-means. IEEE Access, 7, 185239–185249 (2019)
[29] von Lücken, C., Brizuela, C. A., Barán, B.: Multipopulation approaches in MOEA/D. Computational Optimization and Applications, 81(3), 789–828 (2022)

[30] Armano, G., Farmani, M. R.: Multiobjective clustering using PSO. Expert Systems with Applications, 55, 184–193 (2016)
[31] Kaur, A., Kumar, Y.: Multi-objective vibrating particle system for clustering. Pattern Analysis and Applications, 25(1), 209–239 (2022)