http://doi.org/10.35784/iapgos.6921

received: 08.12.2024 | revised: 21.08.2025 | accepted: 04.09.2025 | available online: 30.09.2025

A HYBRID APPROACH COMBINING GENERALIZED NORMAL DISTRIBUTION OPTIMIZATION ALGORITHM AND FUZZY C-MEANS WITH CALINSKI-HARABASZ INDEX FOR CLUSTERING OPTIMIZATION

Moatasem Mahmood Ibrahim, Omar Saber Qasim, Talal Fadhil Hussein
University of Mosul, Department of Mathematics, Mosul, Iraq

Abstract. *In this paper, we propose a new hybrid approach, which combines Generalized Normal Distribution Optimization Algorithm (GNDOA) and fuzzy C-Means clustering (FCM). It is designed for processing unsupervised datasets. This idea target list the development about conventional function option and clustering techniques. The proposed GNDOA-FCM uses normalized normal distribution concept along with FCM for more accurate and efficient clustering outputs leading to accelerated detection in survey region. Calinski-Harabasz index helps finding the number of clusters that has high compactness within each cluster and also apart from other clusters. The performance of the proposed hybrid GNDOA-FCM approach is tested extensively using different benchmark datasets. The results are compared with existing clustering methods using evaluation metrics like silhouette score & feature selection accuracy. Experimental results show that the proposed method can be flexibly set to obtain higher quality of clustering and is more effective than conventional techniques.*

Keywords: feature selection, generalised normal distribution optimisation algorithm, fuzzy C-means clustering, data mining, Calinski-Harabasz index

HYBRYDOWE PODEJŚCIE ŁĄCZĄCE UOGÓLNIONY ALGORYTM OPTYMALIZACJI ROZKŁADU NORMALNEGO I ROZMYTE C-ŚREDNIE ZE WSKAŹNIKIEM CALIŃSKIEGO-HARABASZA DO OPTYMALIZACJI GRUPOWANIA

Streszczenie. *W niniejszym artykule proponujemy nowe podejście hybrydowe, które łączy algorytm uogólnionej optymalizacji rozkładu normalnego (GNDOA) i klasteryzację rozmytych C-średnich (FCM). Zostało ono zaprojektowane do przetwarzania nienadzorowanych zbiorów danych. Pomysł ten ma na celu rozwój konwencjonalnych opcji funkcji i technik klasteryzacji. Proponowany GNDOA-FCM wykorzystuje koncepcję znormalizowanego rozkładu normalnego wraz z FCM w celu uzyskania dokładniejszych i wydajniejszych wyników klasteryzacji, co prowadzi do przyspieszenia wykrywania w badanym regionie. Wskaźnik Calińskiego-Harabasza pomaga znaleźć liczbę klastrów, które charakteryzują się wysoką zwartością w obrębie każdego klastra, a także w odniesieniu do innych klastrów. Wydajność proponowanego hybrydowego podejścia GNDOA-FCM została dokładnie przetestowana przy użyciu różnych zestawów danych benchmarkowych. Wyniki porównano z istniejącymi metodami klastrowania przy użyciu wskaźników oceny, takich jak wynik sylwetki i dokładność wyboru cech. Wyniki eksperymentów pokazują, że proponowana metoda może być elastycznie dostosowana w celu uzyskania wyższej jakości klastrowania i jest bardziej skuteczna niż konwencjonalne techniki.*

Słowa kluczowe: selekcja cech, uogólniony algorytm optymalizacji rozkładu normalnego, klastrowanie rozmytych C-średnich, eksploracja danych, wskaźnik Calińskiego-Harabasza

Introduction

Data science and machine learning that involves high-dimensional data requires efficient processing [1, 8]. This work consists of two main approaches, the Generalized Normal Distribution Optimization Algorithm (GNDOA), and the fuzzy

C-means (FCM) clustering algorithm. One task of feature selection is to keep only most informative features and reduce data dimensionality as such [7]. All three strategies provide predictions that are far more informative than the simple average prediction obtained in very high-dimensional settings and thus significantly enhance model performance [25].

The GNDOA is an advanced optimization approach based on the transformations of normalized normal distributions [26].

It outperforms conventional optimization methods by representing many solutions across its dimensions. By employing a probabilistic framework, GNDOA promotes the balance between exploration and exploitation tactics while establishing a stronger convergence guarantee and improving optimization efficacy [13]. Thanks to these features it can be used in different domains such as industrial design or economic modeling.

FCM is one of the most popular unsupervised learning methods, which allows more freedom in the clustering process [12]. While in most clustering approaches each data point does belong to just one cluster, FCM enables a step of membership in more than one cluster. FCM is capable of providing good clustering when clusters are overlapped and confused over the data such as typical application areas for image analysis, biological data processing etc. [3]. Location optimization in an iterative manner serve as good representation tool reflecting data structure through continuous adaptation cluster accuracy. K-means algorithm a traditional method but FCM functioning mechanism proves more advantageous than k-mean [22].

Feature selection is a key aspect of machine learning preprocessing and consist of the task to find relevant features and Removing irrelevant ones to mitigate learnt system efficacy while minimizing computational complexity [20]. Different techniques such as filtering, bagging and embedding provide unique advantages based on dataset characteristics and study objectives. With larger datasets, careful feature selection plays a vital role in developing reliable machine learning models that perform efficiently [18].

The Calinski-Harabasz index is of key importance in this study to determine the optimal number of clusters Cluster formation can be objectively selected over the number of clusters that best represent structures within a dataset according to cluster compactness

and separation provided by CH Index [10]. Evaluating clustering based on the fact that CH Index well separates final groups. These combined approaches of GNDOA and FCM significantly enhance the performance of unsupervised data analysis. This framework addresses the complexity of modern data analysis and enables data-driven decision-making catering to different domains by providing a novel predictive model that is accurate as well as computationally efficient [4].

This essay is structured as follows: A comprehensive overview of GNDOA is provided in Section I. Section II, we shift to the CH index, a metric to check the quality of classified data generated by the fuzzy c-means clustering algorithm. The fuzzy C-Means method of convergence is covered in Section III. Section IV outlines the suggested approach and concentrates on our framework. Section V concludes by presenting the findings of the thorough analysis. The data analysis's experimental findings are shown in Section VI.

1. Generalized normal distribution optimization algorithm (GNDOA)

The case of generalized normal distribution optimization (GNDOA) is explored in this kind of section, and is broken down into sub-divisions. The first subsection discusses primary motives driving the GNDOA. After that, the next subsection describes the GNDOA. Last but not least, the use of GNDOA within optimization has been explained [9, 27].

1.1. Description

Normal Distribution method is used to sustain the individual in state based on population location data. The effect of the normal distribution on GNDOA or the normal distribution or the Gaussian is a great way to describe constants. In addition, the Normal Classification concept is described [19]. Where x has a fixed distribution that is parameterized with location and scale parameters. A probability density function for x can look like the following:

$$f(x) = \frac{1}{\sqrt{2\pi\delta^2}}\, e^{(\frac{-(x-\mu)^2}{2\delta^2})} \tag{1}$$

Thus δ is scale parameter and μ the location parameter. The location and scale parameters control the mean value and standard deviation of random variables distributed in this manner. For population-based optimization methods, the process typically follows three steps: all individuals in the population initially have a dispersed distribution; next, they start to move towards the best solution, based on exploration and exploitation strategies; lastly, individuals congregate around their best-known solution. We can consider normal fields as random variables following a distribution with mean and an optimal position that lies between the first guesses with large standard deviations among individuals [17]. The second stage closes the gap again, but this only minimizes changes to the entire position and the last stage can also enter a zero distance between mean and best rates in an optimal world where standard deviations are reduced by individual ends.

The information exchange in GNDOA is based on both local and global research, and its procedures are straightforward and clear. The current mean and normal value guide local use and are built into a generalized normal distribution model. In addition, the global study includes three randomly selected topics. A detailed explanation of the two learning methods will be provided.

1.2. Local exploitation

The local operation is to obtain accurate results including the location of all individuals in the search area. GNDOA models can be developed to enhance the relationship between the natural distribution of the population and the actual distribution of the population [6].

$$v_i^t = \mu_i + \delta_i \times \eta\ ,\ i = 1,2,3,\ldots,N \tag{2}$$

$\boldsymbol{\delta_i}$ is the normalized standard variable, v_i^t is the posterior vector of the i^{th} person at time t, μ_i is the normalised mean of the i^{th} person, and $\boldsymbol{\eta}$ is the normalized penalty factor and thus, $\boldsymbol{\mu_i}$, $\boldsymbol{\delta_i}$, and $\boldsymbol{\eta}$ can be described in eq. (3), (4) and (5).

$$\mu_i = \frac{1}{3}(x_i^t + x_{Best}^t + M) \tag{3}$$

$$\eta = \begin{cases} \sqrt{-\log(\lambda_1) \times \cos 2\pi\lambda_2)}\ , if\ a \le b \\ \sqrt{-\log(\lambda_1) \times \cos 2\pi\lambda_2 + \pi\ )}, otherwise \end{cases} \tag{4}$$

a, b, λ_1, and λ_2 are random numbers between 0 and 1. M (10) is the midpoint, x_{Best}^t is the current top position in the existing population of all M, and N is the samples.

$$M = \frac{\sum_{i=1}^{N} x_i^t}{N} \tag{5}$$

1.3. Global exploitation

Global Search: Searches for promising journals related to global place consideration. In GNDOA, the global search is restricted to three randomly selected individuals, which can be given as follows:

$$v_i^t = x_i^t + \beta \times (|\lambda_3| \times v_1) + (1-\beta) \times (|\lambda_4| \times v_2) \tag{6}$$

In Eq.(7), $\beta \times (|\lambda_3| \times v_1)$ is the local shared information and $(1-\beta) \times (|\lambda_4| \times v_2)$ Represents shared global information. p_1, p_2 and p_3 are fixed numbers based on standard normal distribution [14]. (β) is a random number between from 0 to 1, and there are two path vectors. Furthermore, it could be seen that:

$$v_1 = \begin{cases} x_i^t - x_{p1}^t\ , \text{if } f(x_i^t) < f(x_{p1}^t) \\ x_{p1}^t - x_i^t\ , \text{otherwise} \end{cases} \tag{7}$$

$$v_2 = \begin{cases} x_{p2}^t - x_{p3}^t\ , \text{if } f(x_{p2}^t) < f(x_{p3}^t) \\ x_{p3}^t - x_{p2}^t\ , \text{otherwise} \end{cases} \tag{8}$$

Here p_1, p_2, and p_3 are three distinct numbers from 1 to N, occurring before $i \neq p_1 \neq p_2 \neq p_3$. We can say that the next term on the right-hand side of equations (7) and (8), equation (6) is a local knowledge term, which means that the result contains the same information as p_1. Global information and shared decisions. The third term of control of equation (6) describes the individual in their p_2 and p_3 and the information specified by. Moreover, λ_3, λ_4 are random numbers with normal distribution. And that modification criterion (β) makes the two information distribution methods equal. The marking in all equation (6) must remain fixed in the direction of the object in equations (7) and (8).

1.4. GNDOA function for optimization

This section discusses the operation of the GNDOA. The optimization algorithm is based on the developed local exploitation and global exploration strategies. These two strategies are equally important for GNDOA and have the same selection opportunities. In addition, as with population-based optimization strategies, GNDOA starts with the population:

$$x_{i,j}^t = I_j + (u_j - I_j) \times \lambda_5 \tag{9}$$

here $i = 1, 2, 3, 4, \ldots, N$, and $j = 1, 2, 3, 4, \ldots, D$.
λ_5 Is an arbitrary number between 0 and 1, D is the design number of variables, in the design of the j, the upper bounds of the variables are u_j and I_j. Remember that the typical does not find the best results through global search or local exploitation methods. A demonstration method in equation (10) is assumed to carry the improved solution to the next generation; that can be stated as follows:

$$x_i^{t+1} = \begin{cases} v_i^t\ , \text{if } f(v_i^t) < f(x_i^t) \\ x_i^t\ , \text{otherwise} \end{cases} \tag{10}$$

1.5. Computational complexity of the algorithm

The computational complexity is an important estimate of the time required to execute the algorithm. The computational challenges of GNDOA include the time to compare and update positions, depending on the number of participants, repetitions, and the flexibility with which N individuals can update their positions in each cycle, all GNDOA mathematical complexities can be expressed as $O\,(+)$. The flowchart of the proposed GNDOA is shown in Fig. 1.

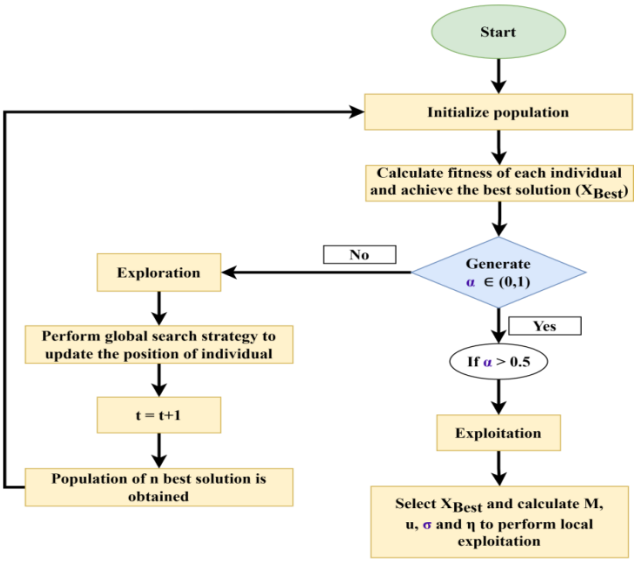

Start
Initialize population
Calculate fitness of each individual and achieve the best solution (X_{Best})
Generate $\alpha \in (0,1)$
No
Yes
Exploration
Perform global search strategy to update the position of individual
t = t+1
Population of n best solution is obtained
If $\alpha > 0.5$
Exploitation
Select X_{Best} and calculate M, u, σ and η to perform local exploitation

Fig. 1. GNDOA flowchart

2. The Calinski-Harabasz index

The Calinski-Harabasz index (CH) is a widely used metric for evaluating clustering schemes with an integrative nature as well as the K-means clustering technique [24]. This index is essential when evaluating the effectiveness of clustering algorithms as it shows how well data are separated into a specified number of clusters by the aforementioned algorithms [11]. The effectiveness is then statistically measured using contrast ratio standards, also called the CH index. It is computed by dividing the total dispersal (defined as the sum of square distance) with the total dispersal [23]. In a nutshell, it measures the distribution present inside each cluster vs. how separated the data points are between clusters:

$$\mathrm{CH} = \frac{\frac{BGSS}{K-1}}{\frac{WGSS}{N-K}} = \frac{BGSS}{WGSS} * \frac{N-K}{K-1} \tag{11}$$

where: N – the total amount of observations, K – the overall cluster count.

The formula for calculating the between-group sum of squares inter-cluster dispersion is as follows:

$$BGSS = \sum_{k=1}^{k} n_k * \|c_k - c\|^2 \tag{12}$$

where: n_k – how many observations there are in cluster k, c_k – the cluster k centroid, c – the barycenter, or centroid, of the dataset, k – how many clusters there are enter an equation here.

The following equation is used to determine WGSS or within-group sum of squares intra-cluster dispersion.

$$WGSS = \sum_{i=1}^{n_k} \|x_{ik} - c_k\|^2 \tag{13}$$

where: n_k – the quantity of data in cluster k, c_k – the cluster k centroid, x_{ik} – the i-th observation of cluster k.

Then add up each individual square sum within a group:

$$WGSS = \sum_{k=1}^{k} WGSS_K \tag{14}$$

where $WGSS_k$ – the within group sum of squares of cluster k. he big value of the Calinski-Harabasz index, according to the above calculation, indicate superior clustering.

3. Fuzzy c- means clustering

FCM clustering is a preliminary unsupervised learning method based on fuzzy set theory to assign data points to multiple groups based on their similarity [2, 21]. Unlike rigorous traditional methods such as K-means, which each point needs to be assigned to only one group, FCM allows partial membership representation, which helps to better understand the relationship between data [5]. These algorithms reduce the objective function by:

$$J(U,V) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^{m} \left\|x_j - v_i\right\|^2 \tag{15}$$

c represents the number of clusters, while n denotes the number of data points. x_j Data point j, v_i cluster center i. Furthermore, u_{ij} denotes the number of members of data point j in cluster i, while m (where $m > 1$) is the ambiguity coefficient of the overlap between clusters Fuzzy C-means (FCM) algorithm iteratively updates cluster centers and membership criteria until there is no collision, making it particularly suitable for many applications such as image classification, medical analysis, and market segmentation [16]. but FCM is sensitive to noise and extraneous factors, which can affect the the magnitude of the outcome of the clustering process [28]. The FCM (Fuzzy Value-Based Cluster Analysis) algorithm works through an iterative process that includes the following steps [15]:

Step 1: Given the required number of data points and groups, initialise the membership matrix with random values.

Step 2:

- Determine the centroid.
- The formula for finding out the centroid (v) is:

$$\mathrm{v_{ij}} = \left(\sum_{\mathrm{k}=1}^{\mathrm{n}} \gamma_{\mathrm{ik}}^{\mathrm{m}} * \mathrm{x_k}\right) \div \sum_{\mathrm{k}=1}^{\mathrm{n}} \gamma_{\mathrm{ik}}^{\mathrm{m}} \tag{16}$$

where: γ_{ik}^{m} – Fuzzy membership value, m – Fuzziness parameter generally taken as 2, x_k – is the data point.

Step 3: The algorithm uses the Euclidean distance metric to calculate how far the data points are from their centers.

$$\mathrm{D(x,y)} = \sqrt{(\mathrm{x_1} - \mathrm{y_1})^2 + (\mathrm{x_2} - \mathrm{y_2})^2 + \cdots + (\mathrm{x_n} - \mathrm{y_n})^{\mathrm{n}}} \tag{17}$$

Step 4: Updating membership values.

$$\gamma_{ki} = \left(\sum_{j=1}^{n} \left(\frac{d_{k\,i}^2}{d_{k\,j}^2}\right)^{\frac{2}{m-1}}\right)^{-1} \tag{18}$$

where: $d_{k,i}$ – is the distance between data point k and centroid i, n – is the number of clusters.

Step 5: Iterate the steps until the values of subscriptions stabilize or the difference is under a tolerance limit.

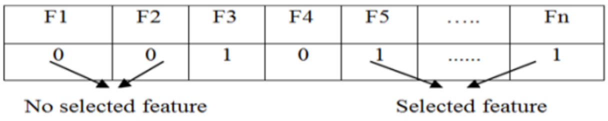
4. Proposed method

This work presents a new approach of Generalized Normal Distribution Optimization Algorithm (GNDOA) and Fuzzy C-Means Clustering (FCM), for optimizing this aspect of unsupervised data analysis. The hybrid method outlines the procedure as follows: GNDOA modeling capable diversity classification model implementation employs variance-based GNDOA-FCM optimization providing histogram appearance fine-tuning parameters to minimize the objective function while maintaining clustering quality, where used Calinsky-Harabaz index as criterion determine optimal number clusters balancing high cluster compactness and separation between them. The index measures cluster separation, which requires both intergroup and intra-group dispersion. The proposed method shows an efficient compromise between the two competing aspects of cluster separation and compactness by maximizing the CH index, which assist in identifying the best possible cluster structure for a Fuzzy C-Means based data set. The means approach adds different layers of clustering–while the Reflecting linkages GNDOA and FCM provides a solid foundation for unsupervised learning, thereby authenticity and reliability of cluster collection edge on complex datasets with great efficiency; can be proficiently used in various applications such as healthcare, financial and imaging applications. The steps described as following demonstrate how the proposed method work.

1. Start with a data set X of $N \times D$ dimensions, where N represents the number of observations and D denotes the number of features.
2. Specify the maximum number of clusters K_{max} to test in the cluster analysis.
3. Set up a population with a fixed number of solutions P, along with a set of combinations and selection orders for each number of clusters K.
4. Use the Fuzzy C-Means (FCM) clustering algorithm with the specified K and selected features to evaluate the fit of each solution in the population P.

5. Determine the optimal number of clusters $K_{optimal}$ by calculating the Calinski-Harabasz index for each clustering configuration, selecting the K that maximizes this index as the optimal cluster count.
6. Choose the optimal solution from the population P^* based on fitness criteria to form a new population X^*.
7. Apply the Generalized Normal Distribution Optimization Algorithm (GNDOA) to generate a new solution set P^*, using GNDOA evaluation operators such as status updates and penalty factors to refine the solution space.
8. Use the FCM algorithm with the updated data set X and the corresponding feature selection to evaluate the adequacy of the new solution set P^*.
9. Select the best solution from X based on fitness values, then update the population to X^{**}.
10. Repeat steps 6 through 8 until the termination criterion is met, which may include convergence to an optimal fitness value or a predefined number of iterations.
11. The final population is selected, with $K_{optimal}$ as the optimal subset for Fuzzy C-Means clustering, based on the highest quality value achieved.

The method introduced combines population-based search algorithms such as GNDOA operators and Fuzzy C-Means clustering method. The Calinski-Harabasz index simultaneously optimizes the number of clusters for C-Means clustering and is used to select features to cluster data more efficiently and improve resource efficiency analysis.



F1	F2	F3	F4	F5	…..	Fn
0	0	1	0	1	……	1

Fig. 2. Selection signal in the proposed algorithm

Figure 2 shows binary string representation of K and feature selection, with 1 indicating object selection and 0 indicating not selection.

5. Results and discussion

In order to test the performance of the introduced GNDOA-FCM algorithm, five public datasets from UCI were used. The data experimented is presented in the UCI Machine Learning Repository and summarized in table 1. We selected five realistic scenarios with varying numbers and dimensionalities of observations. A summary of each data set is presented in a table to highlight its features.

Table 1. Describe the characteristics of the dataset

The dataset	Case(N)	Distance(D)
Data-set(Yeast)	1484	8
Data-set(Semeion)	1593	265
Data-set(Biodeg)	1055	41
Data-set(Cmc)	1473	9
Data-set(Parkinsons)	195	23

We use silhouette scores, which are defined to measure the efficiency of this algorithm. Cross-validation method measures the quality of the clustering results and aims to maximize the inter-cluster differences by the point difference of each cluster of equal size, where higher values indicate that those clusters fit well in a specified model mean while lower or negative values indicate less clustering. In Clustering it is appropriate when many things have high shadow on their own, and hence silhouette width s(i) is defined as:

$$s(i) = \frac{b(i)-a(i)}{max\{b(i),\ a(i)\}} \tag{19}$$

a(i) Represents the average distance between (i) and all other data points in the same set Ci, and b(i) represents the average distance between (i) and all other data points in the other set Ci.

$$\mathrm{b(i)} = \min_{I\neq J} \frac{1}{|C_J|} \sum_{j\in J} d(i,j) \tag{20}$$

where $d(i,j)$ represents the distance from (i) to (j).

It can be observed in table 2 that compared with fuzzy C-means, GNDOA-FCM achieves a higher clustering accuracy over all data types according to the silhouette value. GNDOA-FCM is more competent for the well data sets due to its reliability dealing with complicated content. In addition, traditional clustering algorithms may struggle with multidimensional datasets. As a result, compared with C-means it has advantages of great clustering accuracy and providing an effective approach to cluster high-dimensional datasets.

Table 2. The clustering accuracy of GNDOA-FCM and C-Means algorithms based on the silhouette value results

The dataset	GNDOA-FCM	C-Means
Data-set(Yeast)	0.2625	0.2164
Data-set(Semeion)	0.1085	0.0991
Data-set(Biodeg)	0.7029	0.4520
Data-set(Cmc)	0.6047	0.4132
Data-set(Parkinsons)	0.6755	0.6754

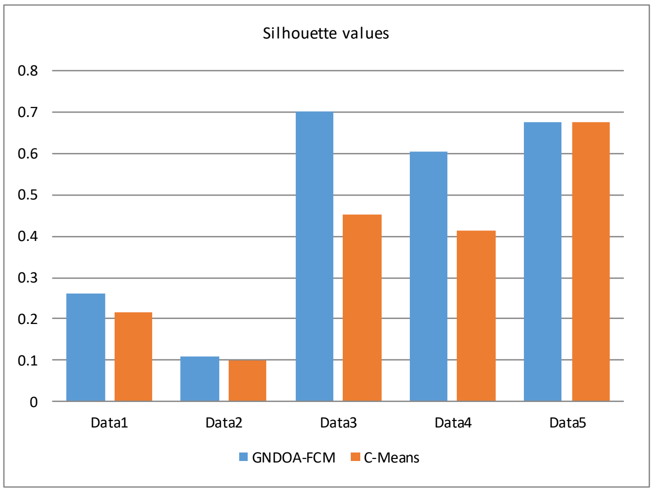


Fig. 3. The comparison of silhouette value results between GNDOA-FCM and C-Means algorithms

The data in table 3 reveal that the GNDOA-FCM algorithm outperforms other algorithms to find the optimal clusters number and identifies relevant features for all datasets. The GNDOA-FCM method clusters that is proved to be very closer match with expected size and also validated using Calinski-Harabasz index. Also, GNDOA-FCM has higher computational efficiency than the clustering method C in average.

Table 3. Compares the GNDOA-FCM and C-Means algorithms in terms of feature selection with optimal cluster

The dataset	GNDOA-FCM	C-Means	Optimal Cluster of number
Data-set(Yeast)	4	8	3
Data-set(Semeion)	126	265	2
Data-set(Biodeg)	21	41	2
Data-set(Cmc)	5	9	6
Data-set(Parkinsons)	10	23	4

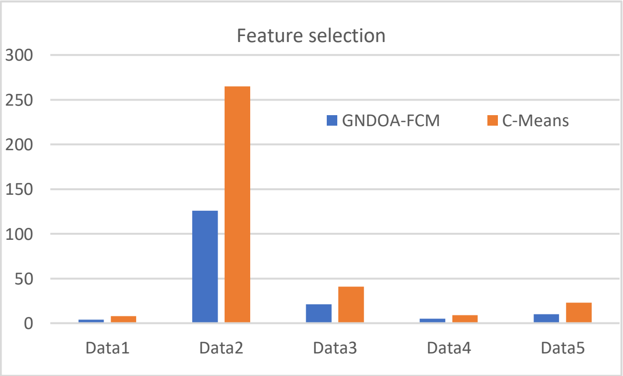


Fig. 4. The average feature selection comparison between GNDOA-FCM and C-means algorithms

6. Conclusion

This study presents a new approach to enhance fuzzy C-means clustering known as GNDOA. To maximize the cluster evaluation, we employed the Calinski-Harabasz index in order to identify the best number of clusters. Five different datasets were used to test the GNDOA-FCM. The two key metrics that compared the performances of the GNDOA-FCM testing technique are average number of features selected and inter-cluster distance. GNDOA-FCM method's efficiency to find optimal feature subsets and silhouette values is proved in Stream & Figures. The accuracy of measurement with this algorithm was better than classic fuzzy C-means. In future, the resilience of GNDOA-FCM in the bioinformatics and data analytical domains can be further investigated by using different sets of applications, which modify some primary assumptions of this algorithm. This is a critical development for the places where high-quality data exists and need to garner insights and drive wise decisions.

References

[1] Al-Kababchee S. G. M., Qasim O. S., Algamal Z. Y.: Improving penalized regression-based clustering model in big data. Journal of Physics: Conference Series 1897, 2021, 012036.

[2] Al Kababchee S. G., Algamal Z. Y., Qasim O. S.: Improving Penalized-Based Clustering Model in Big Fusion Data by Hybrid Black Hole Algorithm. Fusion: Practice & Applications 11(1), 2023, 70–76.

[3] Alqhtani S. M., et al.: Improved Brain Tumor Segmentation and Classification in Brain MRI with FCM-SVM: A Diagnostic Approach. IEEE Access 12, 2024, 61312–61335.

[4] Ashari I. F., et al.: Analysis of elbow, silhouette, Davies-Bouldin, Calinski-Harabasz, and rand-index evaluation on k-means algorithm for classifying flood-affected areas in Jakarta. Journal of Applied Informatics and Computing 7(1), 2023, 95–103.

[5] Cebeci Z., Yildiz F.: Comparison of k-means and fuzzy c-means algorithms on different cluster structures. Journal of Agricultural Informatics 6(3), 2015, 13–23.

[6] Cui A., et al.: Global and regional prevalence of vitamin D deficiency in population-based studies from 2000 to 2022: A pooled analysis of 7.9 million participants. Frontiers in Nutrition 10, 2023, 1070808.

[7] El Touati Y., Slimane J. B., Saidani T.: Adaptive Method for Feature Selection in the Machine Learning Context. Engineering, Technology & Applied Science Research 14(3), 2024, 14295–14300.

[8] Géron A.: Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, Inc. 2022.

[9] Haeri Boroujeni S. P., Pashaei E.: A hybrid chimp optimization algorithm and generalized normal distribution algorithm with opposition-based learning strategy for solving data clustering problems. Iran Journal of Computer Science 7(1), 2024, 65–101.

[10] Hasan F. M., et al.: Enhanced Unsupervised Feature Selection Method Using Crow Search Algorithm and Calinski-Harabasz. International Journal of Computational Methods and Experimental Measurements 12(2), 2024, 185–190.

[11] Hassani M., Seidl T.: Using internal evaluation measures to validate the quality of diverse stream clustering algorithms. Vietnam Journal of Computer Science 4, 2017, 171–183.

[12] Hussain I., Sinaga K. P., Yang M.-S.: Unsupervised multiview fuzzy c-means clustering algorithm. Electronics 12(21), 2023, 4467.

[13] Islam U. J., et al.: Dynamic exploration–exploitation trade-off in active learning regression with Bayesian hierarchical modeling. IISE Transactions 2024, 1–15.

[14] Ji B., et al.: A survey of computational intelligence for 6G: Key technologies, applications and trends. IEEE Transactions on Industrial Informatics 17(10), 2021, 7145–7154.

[15] Krasnov D., et al.: Fuzzy c-means clustering: A review of applications in breast cancer detection. Entropy 25(7), 2023, 1021.

[16] Munusamy S., Murugesan P.: Modified dynamic fuzzy c-means clustering algorithm–Application in dynamic customer segmentation. Applied Intelligence 50(6), 2020, 1922–1942.

[17] Possolo A.: Evaluating, Expressing, and Propagating Measurement Uncertainty for NIST Reference Materials. NIST Special Publication 260, 2020, 202.

[18] Rahman M. M., et al.: A review of machine learning methods of feature selection and classification for autism spectrum disorder. Brain sciences 10(12), 2020, 949.

[19] Raja V., et al.: Alleviation of cadmium-induced oxidative damage through application of zinc oxide nanoparticles and strigolactones in Solanum lycopersicum L. Environmental Science: Nano 11, 2024, 2633–2654.

[20] Song X., et al.: Evolutionary computation for feature selection in classification: A comprehensive survey of solutions, applications and challenges. Swarm and Evolutionary Computation 90, 2024, 101661.

[21] Tayyebi J., Hosseinzadeh E.: A fuzzy c-means algorithm for clustering fuzzy data and its application in clustering incomplete data. Journal of AI and Data Mining 8(4), 2020, 515–523.

[22] Tokat S., et al.: Fuzzy c-means clustering-based key performance indicator design for warehouse loading operations. Journal of King Saud University-Computer and Information Sciences 34(8), 2022, 6377–6384.

[23] Wang X., Xu Y.: An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index. IOP Conf. Ser.: Mater. Sci. Eng. 569(5), 2019, 052024.

[24] Yang K., et al.: Classification and evaluation of driving behavior safety levels: A driving simulation study. IEEE Open Journal of Intelligent Transportation Systems 3, 2022, 111–125.

[25] Zhan M.-F., et al.: Recent advances in statistical methodologies in evaluating program for high-dimensional data. Applied Mathematics-A Journal of Chinese Universities 37(1), 2022, 131–146.

[26] Zhang Y.: An improved generalized normal distribution optimization and its applications in numerical problems and engineering design problems. Artificial Intelligence Review 56(1), 2023, 685–747.

[27] Zhang Y., Jin Z., Mirjalili S.: Generalized normal distribution optimization and its applications in parameter extraction of photovoltaic models. Energy Conversion and Management 224, 2020, 113301.

[28] Zhao W., et al.: Comparison and application of SOFM, fuzzy c-means and k-means clustering algorithms for natural soil environment regionalization in China. Environmental Research 216, 2023, 114519.

B.Sc. Moatasem Mahmood Ibrahim
e-mail: moatasem.23csp150@student.uomosul.edu.iq

Moatasem M. Ibrahim received his B.Sc. degree in mathematics from University of Kirkuk, Kirkuk, Iraq, in 2018. He is studying for his M.Sc. in mathematics at University of Mosul, Mosul, Iraq, and was accepted in 2023. He is a teacher at Kirkuk Education Directorate. His research interests include applied mathematics.

https://orcid.org/0009-0006-5879-2196

Prof. Dr. Omar Saber Qasim
e-mail: omar.saber@uomosul.edu.iq

Omar S. Qasim received his B.Sc. degree in Mathematics from the University of Mosul, Mosul, Iraq, in 2003, and his M.Sc. degree in Mathematics from the University of Mosul, Mosul, Iraq, in 2006. He obtained his Ph.D. degree in Mathematical Science from the University of Mosul in 2010. He is currently a professor of mathematics at the University of Mosul. His research interests include applied mathematics, deep learning, bioinformatics, fuzzy logic, and optimization algorithms.

https://orcid.org/0000-0003-3301-6271

Dr. Talal Fadhil Hussein
e-mail: talal.math@uomosul.edu.iq

Talal F. Hussein received his B.Sc. degree in Mathematics from the University of Mosul, Mosul, Iraq, in 1998, and his M.Sc. degree in Mathematics from the University of Mosul, Mosul, Iraq, in 2002. He obtained his Ph.D. degree in Mathematical Science from the University of Mosul in 2013. He is currently a lecturer of mathematics at the University of Mosul. His research interests include applied mathematics, pattern recognition, mathematical modeling.

https://orcid.org/0009-0005-6276-9493