Published Online First: February, 2024
https://doi.org/10.21123/bsj.2024.9516
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Enhancing Fuzzy C-Means Clustering with a Novel Standard Deviation Weighted Distance Measure

Ahmed Husham Mohammed*[1], Marwan Abdul Hameed Ashour [2]

[1] Department of Statistics, College, College of Administration and Economics, University of Basrah, Basrah, Iraq.
[2] Department of Statistics, College, College of Administration and Economics, University of Baghdad, Baghdad, Iraq
*Corresponding Author.

Received 18/09/2023, Revised 24/11/2023, Accepted 26/11/2023, Published Online First 20/02/2024

Abstract

The aim of this paper is to present a new approach to address the Fuzzy C Mean algorithm, which is considered one of the most important and famous algorithms that addressed the phenomenon of uncertainty in forming clusters according to the overlap ratios. One of the most important problems facing this algorithm is its reliance primarily on the Euclidean distance measure, and by nature, the situation is that this measure makes the formed clusters take a spherical shape, which is unable to contain complex or overlapping cases. Therefore, this paper attempts to propose a new measure of distance, where we were able to derive a formula for the variance of the fuzzy cluster to be entered as a weight on the Euclidean Distance (WED) formula. Moreover, the calculation was processed partitions matrix through the use of the K-Means algorithm and creating a hybrid environment between the fuzzy algorithm and the sharp algorithm. To verify what was presented, experimental simulation was used and then applied to reality using environmental data for the physical and chemical examination of water testing stations in Basra Governorate. It was proven through the experimental results that the proposed distance measure Weighted Euclidean distance had the advantage over improving the work of the HFCM algorithm through the criterion (Obj_Fun, Iteration, Min_optimization, good fit clustering and overlap) when (c = 2,3) and according to the simulation results, c = 2 was chosen to form groups for the real data, which contributed to determine the best objective function (23.93, 22.44, 18.83) at degrees of fuzzing (1.2, 2, 2.8), while according to the degree of fuzzing (m = 3.6), the objective function for Euclidean Distance (ED) was the lowest, but the criteria were (Iter. = 2, Min_optimization = 0 and δ_{XB}) which confirms that (WED) is the best.

Keywords: Cluster, Distance measures, FCM, Fuzzy logic, Hybrid algorithm.

Introduction

Fuzzy algorithms are of significant importance in addressing many phenomena characterized by uncertainty, particularly in relation to the outcomes of laboratory testing aimed at determining concentrations of specific compounds. This particular category of difficulty frequently entails outcomes that are frequently accompanied by a condition of uncertainty regarding the precision of the results, the methodology employed for data collection, or the reliability of the laboratory equipment involved. Hence, the findings suggest that the application of conventional clustering techniques may not yield appropriate group determinations under these circumstances. Furthermore, the

Published Online First: February, 2024
https://doi.org/10.21123/bsj.2024.9516
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

advancement of software development plays a crucial part in the enhancement of optimization algorithms. This advancement has resulted in numerous contributions, particularly within the domain of artificial intelligence, wherein it has played a significant role. The foundation for the development and proposition of numerous algorithms, including hybrid algorithms, is in their ability to tackle complexity and interference, enhance performance, and minimize errors. Consequently, there have been significant advances in this domain. The authors put forth a hybrid technique that combines Principal Component Analysis (PCA) and Fuzzy C-Means (FCM) in order to investigate the characteristics of air pollution[1]. The objective of this work was to develop a model that enhances the outcomes of K-Means clustering through the utilization of principle components analysis (PCA) for the processing of multi-dimensional data[2]. The researchers were able to calculate the density distributions of unstable, neutron-rich exotic nuclei using binary cluster analysis, and they concluded that the calculated cross-sections of the nuclei interactions agreed with practical values[3]. The researchers developed a new algorithm by hybridizing the K-Means algorithm with the DBSCAN algorithm where this proposal addressed the problem of determining the number of initial clusters as well as cluster centers[4]. In this study, a novel methodology was introduced for the assessment of torsion in braiding columns during experimental trials the proposed technique involved the collection of three-dimensional acoustic emission data, which were subsequently analyzed and classified using fuzzy c-means (FCM) technology to identify and categorize instances of damage[5]. The study was conducted to examine the monitoring of water quality, which is regarded as a crucial aspect of safeguarding surface water. The present study focused on the examination of water samples collected from the Nile River, with a specific emphasis on the drinking water stations (CDWPs) situated in Cairo[6]. The study addresses the problem of medical imaging of cancer diseases by applying genetic algorithms and hazy and acute cluster algorithms using an ADF diffusion filter that improved the accuracy of the algorithms' results[7]. This study presented a data clustering technique using the modified Voronoi Fuzzy Algorithm (VFCA), which aims to divide the sensed area into a number of cells that were grouped using the Fuzzy C Mean Clustering (FCM) algorithm. The results also showed the efficiency of the modified algorithm compared to traditional clustering algorithms[8]. The study presented an explanation of how to improve the security and safety of transmitting information and messages over the Internet by hiding data within clusters by adopting the Least Significant Bit (LSB) method[9]. A hybrid algorithm was proposed to improve the work of the FCM fuzzy clustering algorithm by adopting the Tabu probabilistic heuristic algorithm and finding a global clustering based on the value of the objective function at its minimum, and the results showed the superiority of the Tabu-FCM hybrid algorithm[10]. The study addressed the problem of image retrieval by improving the accuracy and speed of the retrieval system by applying the Fuzzy C Means (FCM) clustering algorithm to reduce the search space and speed up the image retrieval process. The results showed a significant improvement in the accuracy and speed of the retrieval system[11]. They proposed a model that addresses the problem of the fuzzy exponent in the FCM algorithm by building a hybrid model (IT2-FCM-FTS) so that this model was able to improve the results of predictions in fuzzy time series[12]. They used the binary cluster analysis method with the aim of identifying groups of patients with a high risk and low risk of contracting Covid-19 disease, based on a set of vital factors. The study concluded that early detection of the disease can reduce cases of severe infection with the disease[13]. The study aimed to classify the pollution areas in the Orontes River in Syria into low-pollution areas and high-pollution areas through the use of hierarchical cluster analysis. The study achieved to classify these areas in the form of two clusters, and the study reached the identification of the pollutant elements in the water[14]. The study proposed a technique that processes images in a way that can collect prominent particles and separate them from the background of the image, taking some treatments, including removing outliers from prominent areas. This technique was applied to six sets of images that have backgrounds complicated[15]. The research Addressed the problem of distributing health human resources to healthcare sites and hospitals in Jakarta, Indonesia by adopting FCM and K-means algorithms through which they were able to find three clusters in the formation of hospitals[16].

Therefore, the contribution of this paper is that the researchers were able to provide a new contribution in addressing the formation of clusters and improving the performance of the Fuzzy Means

Published Online First: February, 2024
https://doi.org/10.21123/bsj.2024.9516
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Clustering algorithm by deriving a formula for the variance of the Fuzzy Clusters as a weight that enters the distance measure, and then developing a hybrid scenario to develop the Fuzzy Means Clustering Algorithm (FCM) Partitioned or membership degree matrix based on the K-means algorithm. Finally, this methodology was applied to the water sector by monitoring the levels of salt concentrations in the waters of Basra Governorate/Iraq, as this phenomenon is linked to an important goal of the sustainability goals set by the United Nations (UN) in the year (2017-2021), which is the goal (6 clean water and sanitation), noting that this goal negatively affects goal (14 life underwater).

Materials and Methods

K-Means Algorithm

It algorithm was proposed by Hartigan (1957), adopting the approach of dividing a number of solutions (S) that have (q) dimensions into K homogeneous[17]. The mean clustering method is considered one of the simple traditional methods through which clusters are formed by pre-determining the number of clusters (K) [18,19], the steps of the algorithm can be summarized as follows:

- Randomly determined the number of clusters (k) and cluster centers (v_k).
- Determine the partition matrix $[p_{ij}]$which is of order $n \times k$, and whose elements are:

$$p_{ik} = \begin{cases} 1 \; if \; s \in k \\ 0 \; if \; s \notin k \end{cases}$$

- Determine the type of distance D.
- Calculate the objective function (Obj_Fun), which is calculated according to the Eq :

$$Obj\_Fun(X; k; v) = \sum_{i=1}^{n} \sum_{k=1}^{c} p_{ij} \left\| X_{ij} - v_{jk} \right\|^2 ; j = 1,2, \dots, q \quad 1$$

- Stop condition $|Obj\_Fun^{I} - Obj\_Fun^{I-1}| < \epsilon$; since ϵ is a very small value, if the condition is not true, step 2 is returned.
- Updating the cluster centers v_k based on the distribution of cases within the clusters, is calculated according to the following formula:

$$v_k = \frac{\sum_{i=1}^{n_k} x_i}{n_k} \quad 2$$

- Then apply steps 3-5 and stop when the condition in step 5 is true.

Fuzzy Clustering

Fuzzy Logic

Fuzzy logic is an expression of the state of uncertainty that simulates and addresses problems of complexity and interference of data, modeling and measurement errors, etc. The foundation of this logic was laid by the Iranian scientist Zadeh in 1974, and this approach achieved great progress with the development of programming and artificial intelligence algorithms[20], the basis of this logic is based on two basic principles: the availability of belonging functions μ_A and the degree of each element's belonging to the comprehensive set (P), which is expressed as:

$$\mu_A: P \rightarrow \{0, 1\}; A \in P$$

Therefore, the fuzzy set can be defined as a set of ordered pairs of a number of elements (x) that belong to the comprehensive set, and the element belonging function is the function that determines the degree to which these elements belong to the fuzzy set (A)[21]:

$$A = \{(x, \mu_A(x); x \in P)\}$$

Fuzzy C-Means Clustering

The complexity and overlap facing many multi-dimensional data made traditional clustering algorithms useless, so it was necessary to develop clustering techniques that could deal with the overlap between clusters and form homogeneous groups according to fuzzy logic. The first algorithm to address this interference was Fuzzy C Means (FCM)[22], which aims to collect large data in the form of new, more homogeneous groups based on determining degrees of belonging to the targeted cases. This algorithm was developed by Dunn, Bezdek In 1974, by developing the partition matrix in the K-Means clustering algorithm while determining the degree of fuzzing, then the objective function was obtained, which represents the minimization sum of square errors[24,25] as in the following eq[23]**:**

$$Obj\_Fun(X, P, v) = \sum_{k=1}^{c} \sum_{i=1}^{n} p_{ik}^{m} \, \varphi(x_{ij} \, , v_{jk}) \quad 3$$

Where:

P: It is a matrix (k×n) and represents the membership degree for each element within the clusters.

$\boldsymbol{m}$**:** Fuzziness coefficient (fuzziness exponent), whose value is defined $\mathbf{1} < \boldsymbol{m} \leq \infty$

Published Online First: February, 2024
https://doi.org/10.21123/bsj.2024.9516
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

$\boldsymbol{x_{ij}}$: value of observation i at dimension j.
$\boldsymbol{v_{jk}}$**:** represents the center of cluster k of dimension j.
$\boldsymbol{\varphi(x_{ij}; v_k)}$: is a measure of similarity or difference of the observation value$\boldsymbol{x_{ij}}$ from its cluster center $\boldsymbol{v_{jk}}$**.**

In order for us to achieve the objective function, that must determine both the center of the fuzzy cluster $\boldsymbol{v_k}$and determine the partition matrix that contains the membership degree to each case p_{ik} and the degree of fuzziness (m) and (**c**) the number of clusters know its formulas as follows:

$$v_{jk} = \frac{\sum_{i=1}^{n} p_{ik}^{m} x_{ij}}{\sum_{i=1}^{n} p_{ik}^{m}} \quad 4$$

$\forall\, k = 1,2, \dots, c; j = 1,2, \dots, q$

Accordingly, the obtained Eq.4 represents the estimate of the cluster center[26]. Accordingly, the cluster center here has been weighted by the membership degree(p_{ik}), which also depends on the measure of similarity or dissimilarity, and accordingly the matrix can estimate membership degree according to the following equation[27]:

$$p_{ik} = \frac{h_k \varphi(x_i, v_k)}{\sum_{k=1}^{c} h_k \varphi(x_i, v_k)} \; ; \forall\, i = 1,2, \dots, n \quad 5$$

Where:
$\boldsymbol{h_k = \frac{n_k}{n}}$: represents the ratio of (sampling size), i.e. the number of elements in each cluster (n_k) to the total number of elements(n_k), and is the true condition $\sum_{k=1}^{c} h_k = 1$.
$\boldsymbol{\varphi(x_i, v_k)}$: Represents a measure of similarity or difference (distance measure).

Measure of Distance

The distance or similarity measure $\boldsymbol{\varphi(x_{ij}, v_{jk})}$ is a measure based on the formation of a similarity matrix for (n) cases and (q) variables. Accordingly, the degree of closeness between the points of each variable can be determined according to the cases to form the Proximate Matrix based on this in Measuring the distance based on the centers of the clusters[28, 29]. If it has a vector of variables, $X = [X_1 \quad X_2 \quad \dots \quad X_q]$ the information matrix that has a rank$(n \times q)$, and then a convergence matrix is as follows[30]**:**

$$D = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1q} \\ d_{21} & d_{22} & \dots & d_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \dots & d_{nq} \end{bmatrix} \quad 6$$

There are several types of similarity and difference metrics to determine distance but two traditional distance metrics will be displayed: Euclidean Distance (**ED**), Square Euclidean Distance (**SED**), and two other proposed measures are Weighted Euclidean Distance **(WED)** and Weighted Square Euclidean Distance (**WSED**), which will be explained later:

Euclidean Distance (ED)

$$d_{(w)ik} = \sqrt{\sum_{i=1}^{n} w_k^2 \left(x_{ij} - v_{jk}\right)^2} \quad 7$$

$\forall\, k = 1,2, \dots, c; \;\; w_k^2 = 1$
w is any weighted, that equals 1 in this case

Square Euclidean Distance (SED)

$$d_{(w)ik}{}^2 = \left(\sqrt{\sum_{i=1}^{n} w_k^2 \left(x_{ij} - v_{jk}\right)^2}\right)^2 \quad 8$$

$\forall\, k = 1,2, \dots, c; \;\; w_k^2 = 1$

Proposed Weighted Distance

The formula for the distance measure in the fuzzy average cluster algorithm was developed by deriving an approximate formula for the variance of the fuzzy cluster. Referring to Eq. 3, the researchers were able to derive this equation and find the weighted equation for the distance measure. To achieve this, the derivative amount was found $\frac{\partial obj(X,P,V)}{\partial w_k}$ and shown below:

$$Obj(X,P,V) = \sum_{k=1}^{c} \sum_{i=1}^{n} p_{ik}^{m} \, \varphi(x_{ik}, v_k)$$

$$Obj(X,P,V) = \sum_{k=1}^{c} \sum_{i=1}^{n} p_{ik}^{m} \; \boldsymbol{d_{(w)ik}^{2}}$$

$$= \sum_{i=1}^{n} p_{ik}^{m} \left[w_k^2 \left(x_{ij} - v_{jk}\right)^2\right]$$

$$\frac{\partial Obj}{\partial w_k} = 2w_k \left(\sum_{i=1}^{n} p_{ik}^{m} \left[\left(x_{ij} - v_{jk}\right)^2\right]\right)$$

$$= 2w_k \left(\sum_{i=1}^{n} p_{ik}^{m} \left[x_{ij}^2 - 2v_k x_{ij} + v_{jk}^2\right]\right)$$

$$= 2w_k \left(\sum_{i=1}^{n} p_{ik}^{m} x_{ij}^2 - 2v_k \sum_{i=1}^{n} p_{ik}^{m} x_{ij} + v_k^2 \sum_{i=1}^{n} p_{ik}^{m}\right)$$; Substitute the Eq 4

Published Online First: February, 2024
https://doi.org/10.21123/bsj.2024.9516
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

$$= \left(\sum_{i=1}^{n} p_{ik}^{m} x_{ij}^{2} - 2\left(\frac{\sum_{i=1}^{n} p_{ik}^{m} x_{ij}}{\sum_{i=1}^{n} p_{ik}^{m}}\right) \sum_{i=1}^{n} p_{ik}^{m} x_{ik} + \left(\frac{\sum_{i=1}^{n} p_{ik}^{m} x_{ij}}{\sum_{i=1}^{n} p_{ik}^{m}}\right)^{2} \sum_{i=1}^{n} p_{ik}^{m} \right)$$

$$= \left(\sum_{i=1}^{n} p_{ik}^{m} x_{ik}^{2} - \frac{\left(\sum_{i=1}^{n} p_{ik}^{m} x_{ij}\right)^{2}}{\sum_{i=1}^{n} p_{ik}^{m}} \right) \quad 9$$

Then the final equation of the derivative approaches the formula of the weighted variance with membership degree and the fuzziness exponent:

$$\phi_{jk}^{2} \cong \left(\sum_{i=1}^{n} p_{ik}^{m} x_{ij}^{2} - \frac{\left(\sum_{i=1}^{n} p_{ik}^{m} x_{ij}\right)^{2}}{\sum_{i=1}^{n} p_{ik}^{m}} \right) \quad 10$$

$\forall\, k = 1,2, \dots, c; j = 1,2, \dots, q$

The two Eq 7, 8 can therefore be interpreted to become the distance weighted by the inverse of the standard deviation of the $\frac{1}{\phi_{jk}}$ contract, so that, can set the proposed conventional weighted distance scale as follows:

Weighted Euclidean Distance (WED)

$$d_{ik} = \sqrt{\sum_{i=1}^{n} \frac{1}{\phi_k} \left(x_{ij} - v_{jk}\right)^{2}} \quad 11$$

$\forall\, k = 1,2, \dots, c$

$\boldsymbol{\phi_k}$: St. D. for the fuzzy clustering

Weighted square Euclidean Distance (WSED)

$$d_{ik} = \left(\sqrt{\sum_{i=1}^{n} \frac{1}{\phi_k} \left(x_{ij} - v_{jk}\right)^{2}} \right)^{2} \quad 12$$

$\forall\, k = 1,2, \dots, c$

Developed Fuzzy C-means Algorithm

The FCM cluster algorithm will be developed in two steps:

Step 1: Calculate the membership matrix degree from the original data by adopting the following steps:

The first step represents the initialization stage of the initial cluster centers, adopting the K-Means algorithm

1. Determining the number of clusters.

2. Determining the initial cluster centers randomly.

3. Calculating the distance according to formula 6.

4. Forming primary clusters.

5. Returning to step 2 and obtain cluster centers according to the clusters achieved in step (4).

6. Forming new clusters.

7. Applying the belonging function according to formula 5 to obtain the partitions matrix $(\boldsymbol{p_{ik}})$.

The Second Step: Hyper FCM Algorithm:

8. Entering the centers of the resulting clusters in step 5

9. Entering the matrix of primary affiliation scores resulting from step 7

10. Calculating the objective function according to formula 3.

11. Checking the condition$|Obj^{I+1} - Obj^{I}| < \varepsilon$.

12. If the condition is true it stops. If the condition is not true, the $(\boldsymbol{p_{ik}})$ matrix was updated according to Eq.13:

$$p_{ik} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d^{2}(x_{ij}, v_{jr})}{d^{2}(x_{ij}, v_{jk})} \right)^{\frac{1}{m-1}}} \quad 13$$

$\forall i, r$ where $r = 1,2, \dots, c$

13. Repeating steps 9-11, and the process of updating the$(\boldsymbol{p_{ik}})$ elements continue according to 12 until the stopping condition 11 is true.

Identification of the Validity of Fuzzy Cluster Xie-Beni Criteria (δ_{XB})

The cluster validity criterion (δ_{XB}) verifies the validity of the cluster structure based on the objective function and the fuzzing exponents. Accordingly, it is considered better than the partitions coefficient criterion, which depends only on the membership degrees of the partitions matrix. The best cluster structure is determined by determining the lowest value achieved for this criterion, and it is calculated to the following formula[31]:

$$\delta_{XB}(X, P, V) = \frac{\sum_{i=1}^{n} \sum_{k=1}^{c} p_{ij}^{m} \left\| x_{ij} - v_{k} \right\|^{2}}{n\left(\min_{i,k} \left(\left\| v_{rk} - v_{jk} \right\|^{2} \right) \right)} \quad 14$$

Published Online First: February, 2024
https://doi.org/10.21123/bsj.2024.9516
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Results and Discussion

The results will be discussed in two directions: the first is the experimental aspect (simulation) and the second is the applied aspect, which includes water sector data and its study through physical and chemical examination data, which represents one of the dimensions of the water sustainability goals. They will be presented and their results analyzed as follows:

Simulation Aspect

The simulation method is based on the study of different and complex phenomena in addition to testing the proposed mathematical formulas and developed algorithms with a view to demonstrating their suitability and conformity with reality, so the experimental aspect was built by adopting the following determinants:

- Determining the sizes of the experimental samples (n=20, 50, 200).
- Determining the dimensions of the variables (q=8, 15, 20).
- Determining the degrees of blurring (degree of overlap) m=(1.2, 2, 2.8, 3.6).
- Generating random variables assuming a uniform distribution using the RAND command.
- Converting the random variables generated in paragraph 1 to a normal distribution using the RANDN directive.
- Repeating the experiment to achieve improvement Iter = 100

In applying the above steps and adopting Matlab V.2023, the simulation results were obtained, as follows:

The results of Table 1 showed the superiority of the fuzzy cluster algorithm using the Weighted Euclidean Distance (WED) measure for all the specified experimental cases and all degrees of fuzzing. That noticed a large amount of improvement in the binary cluster case, where the algorithm was able to stop at **Iter = 2** and with the least improvement of the error **Min_Opt. = 0** In addition, it achieved the lowest Obj_Fun objective function and the best fuzzy cluster structure according to the $\boldsymbol{\delta_{XB}}$criterion, whose preference is determined according to the lowest value. It also achieved the best results when **c = 3** in terms of the **Obj_Fun** and the efficiency of the fuzzy cluster structure according to the $\boldsymbol{\delta_{XB}}$ criterion**.**

Table 1. A summary of the simulation results for the comparison between traditional and weighted distance measures when the sample size 20

HFCM		n	20				20				20			
		q	8				15				20			
Comparison			Obj.	Iter	Min Opt.	δ_{XB}	Obj.	Iter	Min Opt.	δ_{XB}	Obj.	Iter	Min Opt.	δ_{XB}
c	m	Dist.												
2	1.2	ED	84.03	10	3.37E-06	2.02	194.22	21	6.5E-06	4.18	277.67	12	5.9E-06	**5.30**
		SED	443.44	8	1.88E-06	16.04	2163.60	24	6.4E-06	54.54	4256.00	14	4.5E-06	**91.82**
		WED	33.68	2	0	0.89	68.88	2	0	1.57	99.81	2	0	**2.25**
		WSED	70.99	18	5.06E-06	1.76	264.32	12	3.1E-06	4.76	540.32	13	1.7E-08	**10.33**
2	2	ED	57.66	12	8.08E-06	2.03	119.67	31	8.8E-06	4.62	168.47	20	6.8E-06	**6.32**
		SED	361.16	23	8.72E-06	10.22	1552.40	42	8.5E-06	34.18	3040.00	31	7.2E-06	**64.74**
		WED	30.61	2	0.00E+00	1.13	56.25	2	0	1.89	78.66	2	0	**2.68**
		WSED	98.92	26	6.27E-06	2.21	338.48	36	8.0E-06	7.46	649.49	38	8.6E-06	**14.35**
2	2.8	ED	33.62	91	9.97E-06	2.15	68.73	15	7.6E-06	2.65	96.76	12	8.6E-06	**3.63**
		SED	237.74	39	7.94E-06	6.68	928.14	53	9.4E-06	27.45	1764.70	49	7.6E-06	**50.01**
		WED	24.19	2	0	0.99	42.43	2	0	1.52	58.64	2	0	**2.07**
		WSED	114.08	34	8.16E-06	2.52	338.11	100	1.3E-04	7.34	630.35	48	7.9E-06	**17.09**
2	3.6	ED	19.31	34	9.6E-06	1.26	39.48	12	8.1E-06	1.52	55.57	10	9.8E-06	**2.09**
		SED	146.27	54	9.1E-06	4.44	533.09	51	8.9E-06	21.12	992.59	100	**0.0042**	**42.50**
		WED	18.00	2	0	0.77	31.49	2	0	1.14	43.14	2	0	**1.53**
		WSED	117.60	60	8.4E-06	2.48	314.81	60	8.5E-06	8.28	580.40	100	**0.003**	**20.42**
3	1.2	ED	63.02	15	9.2E-06	1.48	165.85	35	8.4E-06	2.51	246.64	43	6.4E-06	**3.75**
		SED	247.39	16	4.2E-06	10.30	1758.00	23	8.5E-06	34.79	3458.60	33	9.2E-06	**63.35**
		WED	30.29	15	4.1E-06	0.51	74.56	29	3.0E-06	0.78	96.33	26	4.1E-06	**0.92**
		WSED	58.00	15	1.3E-06	0.97	339.35	28	5.5E-06	4.19	559.47	7	2.3E-07	**4.84**
3	2	ED	37.45	26	8.3E-06	0.88	79.78	30	7.4E-06	2.71	112.31	19	7.9E-06	**4.05**
		SED	206.73	33	8.2E-06	4.46	1063.40	56	7.2E-06	14.41	2116.70	100	**0.1016**	**32.37**
		WED	26.86	28	5.7E-06	0.57	55.99	28	9.6E-06	1.76	77.87	18	9.4E-06	**2.39**
		WSED	108.12	35	8.5E-06	1.79	522.70	100	0.0572	6.66	902.39	99	9.4E-06	**10.78**
3	2.8	ED	16.20	100	2.5E-05	0.97	33.13	14	6.8E-06	1.29	46.64	11	9.3E-06	**2.99**

Published Online First: February, 2024
https://doi.org/10.21123/bsj.2024.9516
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

		SED	114.51	44	7.1E-06	2.16	498.00	100	1.4E-05	6.91	882.40	**100**	0.0214	**31.17**
		WED	16.81	47	9.5E-06	0.86	34.45	14	6.9E-06	**1.09**	**51.34**	12	3.4E-06	**1.57**
		WSED	148.70	64	8.8E-06	2.73	529.95	100	2.1E-05	7.77	896.90	18	7.3E-06	**7.95**
3	**3.6**	**ED**	6.73	26	9.1E-06	0.58	13.76	11	4.6E-06	0.65	19.37	9	9.3E-06	**0.98**
		SED	53.33	100	**1.5E-05**	1.10	252.83	100	0.0534	2.78	351.02	100	**6.3E-05**	**13.36**
		WED	9.86	28	7.9E-06	0.52	20.86	11	6.5E-06	0.66	30.03	10	3.8E-06	**0.92**
		WSED	139.59	67	8.8E-06	2.47	507.64	25	9.8E-06	6.60	747.89	100	**3.3E-04**	**19.02**

The results of Table 2 showed the superiority of the fuzzy cluster algorithm using the (**WED**) measure for all the specified experimental cases and all degrees of fuzzing. A significant improvement has been shown in the binary cluster case. Where the algorithm was able to stop at **Iter = 2** and with the least improvement of the error Min_Opt. = 0 In addition, it achieved the lowest **Obj_Fun** and the best validity of the fuzzy cluster structure according to the $\boldsymbol{\delta_{XB}}$ criterion, whose preference is determined according to the lowest value. It also achieved the best results when **c = 3** in terms of the objective function Obj_Fun and the validity of the fuzzy cluster structure according to the $\boldsymbol{\delta_{XB}}$ criterion, except for one case in which **FCM$_{(ED)}$** was the best at the **(m = 3.6),** but the cluster structure according to **FCM$_{(WED)}$** was the best.

Table 2. a summary of the simulation results for the comparison between traditional and weighted distance measures when a sample size 50

HFCM		**n**												
		q			**8**				**15**				**20**	
Comparison			**Obj.**	**Iter**	**Min Opt.**	$\boldsymbol{\delta_{XB}}$	**Obj.**	**Iter**	**Min Opt.**	$\boldsymbol{\delta_{XB}}$	**Obj.**	**Iter**	**Min Opt.**	$\boldsymbol{\delta_{XB}}$
c	**M**	**Dist.**												
2	**1.2**	**ED**	282.23	9	9.9E-06	3.66	556.01	16	8.9E-06	6.77	750.1	16	8.6E-06	**9.52**
		SED	2043.8	9	1.4E-06	31.24	7344.9	51	8.8E-06	87.21	13656	29	9.7E-06	**148.70**
		WED	81.05	2	0	1.09	**160.1**	**2**	**0**	**2.31**	215.81	2	0	**3.55**
		WSED	166.59	11	9.3E-06	2.05	597.23	47	9.8E-06	6.32	1115.3	51	8.6E-06	**11.51**
2	**2**	**ED**	175.32	50	8.8E-06	8.72	326.3	12	9.2E-06	8.03	435.43	10	7.7E-06	**9.68**
		SED	1483.1	25	7.8E-06	20.27	4794.8	50	9.8E-06	76.63	8389.2	100	3.3E-02	**191.89**
		WED	64.33	2	0	1.68	**122.18**	**2**	**0**	**2.88**	162.72	2	0	**3.56**
		WSED	97.4	16	9.2E-06	2.69	668.14	43	8.0E-06	10.60	1141.2	100	2.1E-03	**25.35**
2	**2.8**	**ED**	100.7	18	8.7E-06	5.01	187.41	9	8.5E-06	4.61	250.10	8	5.1E-06	**5.56**
		SED	889.02	26	7.3E-06	13.91	2727.4	62	8.5E-06	70.90	4817.40	31	6.1E-06	**111.27**
		WED	46.69	2	0	1.43	**90.2**	**2**	**0**	**2.15**	121.21	2	0	**2.66**
		WSED	186.14	24	6.9E-06	2.90	622.53	56	9.0E-06	15.32	1125.40	28	6.5E-06	**25.02**
2	**3.6**	**ED**	57.84	14	7.8E-06	2.88	107.64	8	7.5E-06	2.65	143.65	7	6.9E-06	**3.19**
		SED	514.04	29	8.9E-06	9.70	1566.5	33	8.0E-06	43.16	2766.90	22	5.8E-06	**66.35**
		WED	33.23	2	0	1.08	**66.05**	**2**	**0**	**1.58**	88.02	2	0	**1.94**
		WSED	164.19	27	7.7E-06	2.89	579.57	31	8.2E-06	14.26	1032.80	20	9.6E-06	**22.96**
3	**1.2**	**ED**	249.68	65	8.7E-06	2.44	503.39	45	9.8E-06	4.09	688.31	42	9.0E-06	**5.92**
		SED	1654.3	93	8.2E-06	19.79	6171	100	**1.3E-04**	52.21	11904	94	7.4E-07	**90.61**
		WED	74.76	39	6.7E-06	0.61	153.76	42	**8.6E-06**	1.20	208.32	70	9.2E-06	**1.62**
		WSED	141.87	67	7.4E-06	1.09	590.9	100	**2.0E-04**	4.18	1067.50	93	9.0E-06	**5.52**
3	**2**	**ED**	116.88	47	9.0E-06	5.02	217.53	12	4.8E-06	4.69	290.30	10	3.8E-06	**5.70**
		SED	1000.5	100	3.0E-02	9.44	3222.8	100	**4.0E-04**	36.05	5593.90	100	**0.0193**	**107.24**
		WED	54.3	42	9.7E-06	2.21	100.27	11	7.6E-06	2.01	**134.89**	**9**	**8.2E-06**	**2.45**
		WSED	211.49	70	9.6E-06	1.49	668.14	100	**6.1E-04**	6.92	1155.40	100	**6.9E-03**	**18.71**
3	**2.8**	**ED**	48.54	16	9.8E-06	2.31	90.33	9	2.3E-06	2.06	120.55	7	9.9E-06	**2.52**
		SED	437.22	100	3.3E-03	5.05	1314.6	55	8.6E-06	29.07	2321.90	28	7.5E-06	**47.88**
		WED	32.99	16	7.2E-06	1.34	62.53	**8**	8.5E-06	**1.26**	**84.29**	**7**	**7.4E-06**	**1.53**
		WSED	201.34	57	9.7E-06	2.15	625.84	52	9.8E-06	12.57	1108.00	27	7.4E-06	**20.11**
3	**3.6**	**ED**	20.15	13	4.4E-06	1.32	37.51	7	9.7E-06	0.97	**50.06**	**7**	**1.4E-06**	**1.05**
		SED	180.12	100	2.8E-03	3.33	545.88	29	9.4E-06	13.49	964.16	20	5.6E-06	**19.56**
		WED	20.12	13	4.9E-06	0.82	38.5	8	**1.6E-06**	**0.77**	53.55	7	1.5E-06	**0.97**
		WSED	172.47	100	8.5E-03	1.67	567.09	30	7.3E-06	11.39	1102.60	20	6.5E-06	**20.01**

The results of Table 3 showed the superiority of the fuzzy cluster algorithm using the (WED) measure for all the specified experimental cases and all fuzziness exponents. A significant improvement was shown in the binary cluster case where the algorithm was able to stop at **Iter = 2** and with the least improvement of the error **Min_Opt. = 0**. In addition, it achieved the lowest Obj_Fun and the best validity for the fuzzy cluster structure according to the $\boldsymbol{\delta_{XB}}$ criterion, whose preference is determined according to the lowest value. It also achieved the

Published Online First: February, 2024
https://doi.org/10.21123/bsj.2024.9516
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

best results when **c = 3** for all specified criteria and all (m).

The results indicated that the weighted methods using the standard deviation of the clusters contributed significantly to smoothing the data, improving the analysis results, and forming clusters that have the ability to contain the cases. In addition, the simulation results showed that the fuzzy cluster when **c = 2** is the best because it achieved the best results in terms of **Iter.** The minimum amount of improvement is **Min_Opt**.

Table 3. a summary of the simulated results for the comparison between traditional and weighted distance measures when the sample size 200.

HFCM		n	200				200				200			
		q	8				15				20			
Comparison			Obj.	Iter	Min Opt.	δ_{XB}	Obj.	Iter	Min Opt.	δ_{XB}	Obj.	Iter	Min Opt.	δ_{XB}
c	m	Dist.												
2	1.2	**ED**	1297.20	53	9.3E-06	5.31	2598.40	95	9.8E-06	14.62	3456.70	100	1.9E-05	**39.91**
		SED	11637	100	7.1E-05	46.81	43037	97	9.5E-06	159.47	74254	100	3.2E-05	**273.73**
		WED	**143.51**	**2**	**0**	**0.70**	286.30	2	0	2.47	378.62	2	0	**3.76**
		WSED	138.26	76	9.2E-06	0.49	519.91	96	9.9E-06	1.80	875.01	80	9.6E-06	**2.33**
2	2	**ED**	769.27	18	8.5E-06	12.97	1495.90	9	4.0E-06	20.47	1985.40	8	1.2E-06	**24.33**
		SED	7589.30	100	1.3E-03	36.21	25690	41	8.2E-06	355.61	43514	21	9.9E-06	**548.06**
		WED	110.14	2	0	1.30	216.10	2	0	2.88	287.15	2	0	**3.48**
		WSED	152.21	54	8.8E-06	0.72	534.74	31	7.5E-06	7.32	897.68	17	4.7E-06	**11.00**
2	2.8	**ED**	441.83	11	6.4E-06	7.45	1495.90	7	9.1E-06	11.75	1140.30	6	5.6E-06	**13.97**
		SED	4349.10	100	2.2E-04	37.33	14755	20	5.4E-06	209.98	24992	13	9.3E-06	**314.29**
		WED	**81.56**	**2**	**0**	**1.07**	161.71	2	0	2.17	216.14	2	0	**2.64**
		WSED	144.97	44	2.0E-06	1.23	520.77	16	6.5E-06	7.13	895.84	11	5.3E-06	**10.98**
2	3.6	**ED**	253.77	10	8.9E-06	4.28	493.45	7	2.1E-06	6.75	654.93	6	2.9E-06	**8.02**
		SED	2476.10	100	7.2E-05	52.27	8474.40	16	6.4E-06	122.62	14354	12	8.5E-06	**184.24**
		WED	59.92	2	0.0E+00	0.81	120.34	2	0	1.62	161.23	2	0	**1.97**
		WSED	132.57	87	9.1E-06	2.23	502.24	14	3.6E-06	6.87	866.70	11	2.4E-06	**10.62**
3	1.2	**ED**	1156.80	69	8.3E-06	3.06	2389.80	100	**2.7E-04**	8.11	3187.50	100	1.4E-05	**30.40**
		SED	9624.00	100	0.1601	27.10	38460	100	**0.0149**	91.94	66784	100	4.3E-04	**150.45**
		WED	**163.74**	**62**	**8.6E-06**	**0.40**	333.42	68	8.4E-06	0.99	445.87	64	9.6E-06	**4.25**
		WSED	186.46	65	8.9E-06	0.41	732.79	100	**1.9E-04**	1.42	1288.50	100	7.8E-05	**2.16**
3	2	**ED**	512.85	17	9.4E-06	7.38	997.24	9	2.6E-06	11.47	1323.60	7	8.3E-06	**13.49**
		SED	5117.10	100	9.0E-06	13.76	17126	39	7.9E-06	195.13	29009	21	5.8E-06	**293.88**
		WED	109.81	15	9.7E-06	1.51	216.58	8	3.3E-06	2.42	287.42	7	1.8E-06	**2.88**
		WSED	233.45	48	7.5E-07	0.61	800.30	31	7.1E-06	8.94	1367.60	17	7.3E-06	**13.68**
3	2.8	**ED**	212.96	11	9.3E-06	3.31	414.10	7	4.7E-06	5.00	549.61	6	5.2E-06	**5.82**
		SED	2102.00	100	3.0E-01	13.29	7111.60	19	5.5E-06	84.27	12046	14	3.2E-06	**126.07**
		WED	68.16	11	2.9E-06	0.94	137.75	7	1.6E-06	1.54	183.11	6	1.7E-06	**1.83**
		WSED	215.96	100	2.8E-02	1.11	780.21	16	8.9E-06	8.72	1329.70	12	4.8E-06	**13.30**
3	3.6	**ED**	88.43	10	3.0E-06	1.56	171.95	6	7.2E-06	2.22	228.22	6	1.1E-06	**3.09**
		SED	862.83	100	2.7E-05	13.97	2953.10	15	4.1E-06	37.08	5002	12	2.6E-06	**66.80**
		WED	41.56	9	6.5E-06	0.57	85.21	6	3.6E-06	0.95	**118.18**	**6**	**5.3E-07**	**1.18**
		WSED	189.65	95	9.6E-06	2.61	720.86	14	3.3E-06	8.05	1247.50	11	2.7E-06	**12.48**

Practical Aspect

In order to make use of this HFCM algorithm and verify the efficiency of the proposals, these algorithms were implemented on the data of physical and chemical tests for the water testing stations in Basra Governorate, which amount to 8 stations, namely the Shatt al-Arab stations(H_1 ،H_2 ،H_2B ،H_3 ، H_4), and the Qurna station. For the Tigris River (T_{34}), the city's two stations for the Euphrates River (E_{20} ،E_{21}), as this sector is considered one of the important sectors that affect aquatic and human life and the production sectors and industries of various kinds. It is also considered one of the sustainability goals presented by the United Nations in its report (2017-2021). Which included goal 6 (clean water and sanitation) among 17 goals.

Data was collected from eight water testing stations on a monthly basis for the years (2010 - 2021).

It is clear from the results of Table 4 that the new proposed formula for the Euclidean distance measure weighted by the fuzzy cluster standard deviation (WED) has achieved the best results compared to the traditional distance measures (SED, ED) through the approved standards, which will be explained as follows:

Published Online First: February, 2024
https://doi.org/10.21123/bsj.2024.9516
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Table 4. Summary of comparison results of the HFCM cluster average algorithm for environmental data for water testing stations in Basra Governorate

HFCM		No. of clusters c = 2			
Comparison M	Dist.	Obj.	Iter	Min Opt.	δ_{XB}
1.2	**ED**	46.12	10	3.6E-06	**1.34**
	SED	338	7	6.9E-06	**27.42**
	WED	**23.93**	**2**	**0**	**1.09**
	WSED	132.10	7	8.9E-06	**6.26**
2	**ED**	32.76	7	9.1E-06	**1.61**
	SED	176	10	1.9E-06	**8.63**
	WED	**22.44**	**2**	**0**	**0.78**
	WSED	118.85	9	3.4E-06	**2.70**
2.8	**ED**	19.26	14	5.6E-06	**1.40**
	SED	152	13	7.0E-06	**6.79**
	WED	**18.83**	**2**	**0**	**0.62**
	WSED	179.78	12	8.4E-06	**4.10**
3.6	**ED**	**11.01**	16	8.5E-06	**1.00**
	SED	112	16	6.7E-06	**5.67**
	WED	**15.10**	**2**	**0**	**0.47**
	WSED	228.68	16	3.5E-06	**5.21**

When c = 2 m = 1.2:

The WED measure achieved the best results. The value of the objective function Obj_Fun, which represents the amount of error (**23.93**), was the lowest compared to other measures. In addition, the number of replicates achieved the best results at **(Iter. = 2)** and the least amount of error improvement was equal to **0.** As for the validity criterion of the fuzzy cluster structure, it was the best because it also achieved the lowest value $\boldsymbol{\delta_{XB} = 1.09}$, and this is evidence that the (WED) measure has contributed to improving the work of the FCM algorithm and reaching the best results with the least number of iterations compared to other measures, which can also determine the importance of each method by adopting the **AverageMax** criterion, which determines the degree of overlap between clusters. It is clear from **Fig 1** that the FCM algorithm does not show overlap and that the fuzzy cluster is moving towards the clear cluster since AverageMax achieved the highest **90%** for all metrics.

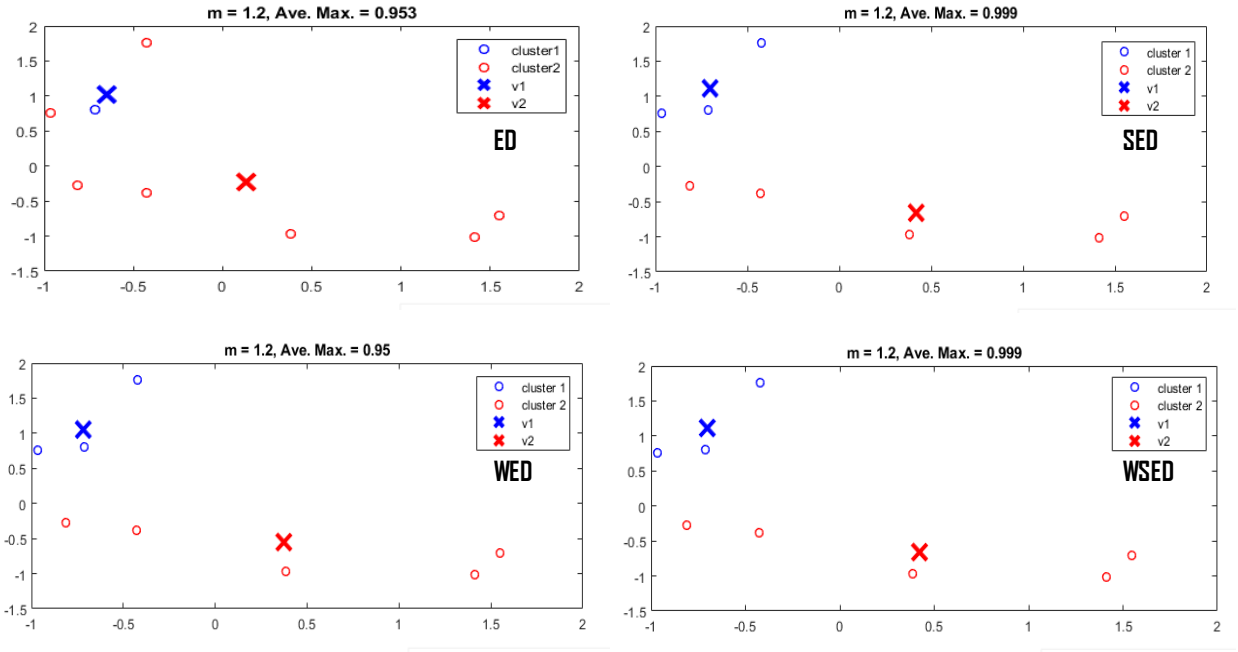

Figure 1. The stations overlap between the clusters when the fuzziness exponent (1.2)

When c = 2, m = 2

The WED measure achieved the best results. The value of the **Obj_Fun**, which represents the amount of error **22.44**, was the lowest compared to other measures. In addition, the number of replicates achieved the best results at (**Iter. = 2**) and the least amount of error improvement was equal to **0.** As for the validity criterion of the fuzzy cluster structure, it was the best because it also achieved the lowest value

Published Online First: February, 2024
https://doi.org/10.21123/bsj.2024.9516
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

$\boldsymbol{\delta_{XB} = 0.78}$, and this is evidence that the (WED) measure has contributed to improving the work of the HFCM algorithm and reaching the best results with the least number of iterations compared to other measures, it is clear that from **Fig 2** and according to the results of the **AverageMax** criterion that the HFCM algorithm has shown that the fuzzy cluster is appropriate at the degree of fuzzing **m = 2** for the **ED** and **WED** metrics, which achieved **0.77, 0.82**, respectively. As for the two metrics **SED** and **WSED**, the fuzzy cluster had a crisp cluster, where the AverageMax value is higher than **90%**.

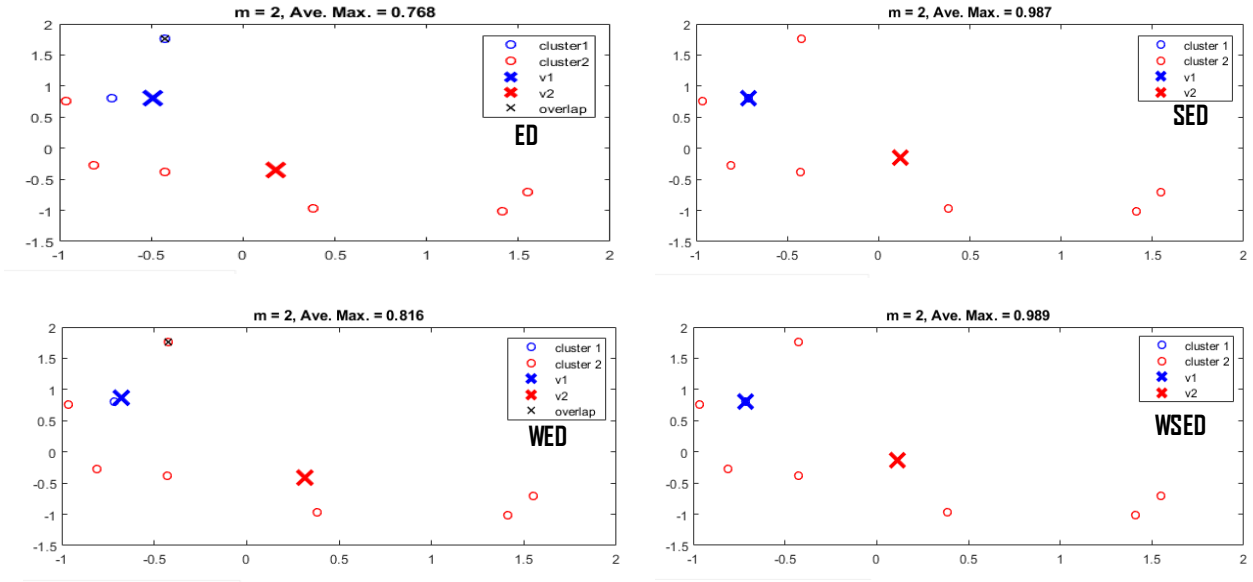

Figure 2. The stations overlap between the clusters when the fuzziness exponent (2)

When c = 2, m = 2.8

The **WED** measure achieved the best results. The value of the **Obj_Fun**, which represents the amount of error **18.83**, was the lowest compared to other measures. In addition, the number of replicates achieved the best results at (**Iter. = 2**) and the least amount of error improvement was equal to **0**. As for the validity criterion of the fuzzy cluster structure, it was the best because it also achieved the lowest value $\boldsymbol{\delta_{XB} = 0.62}$, and this is evidence that the (**WED**) measure has contributed to improving the work of the HFCM algorithm and reaching the best results with the least number of iterations compared to other measures, it shows us from **Fig 3** and according to the results of the **AverageMax** criterion that the HFCM algorithm has shown that the fuzzy cluster is appropriate at the degree of fuzzing **m = 2** for the **ED** and **WED** metrics, which achieved **0.65, 0.75**, respectively. As for the two measures **SED** and **WSED**, the fuzzy cluster had a crisp cluster, where the AverageMax value is higher than **90%**.

Published Online First: February, 2024
https://doi.org/10.21123/bsj.2024.9516
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

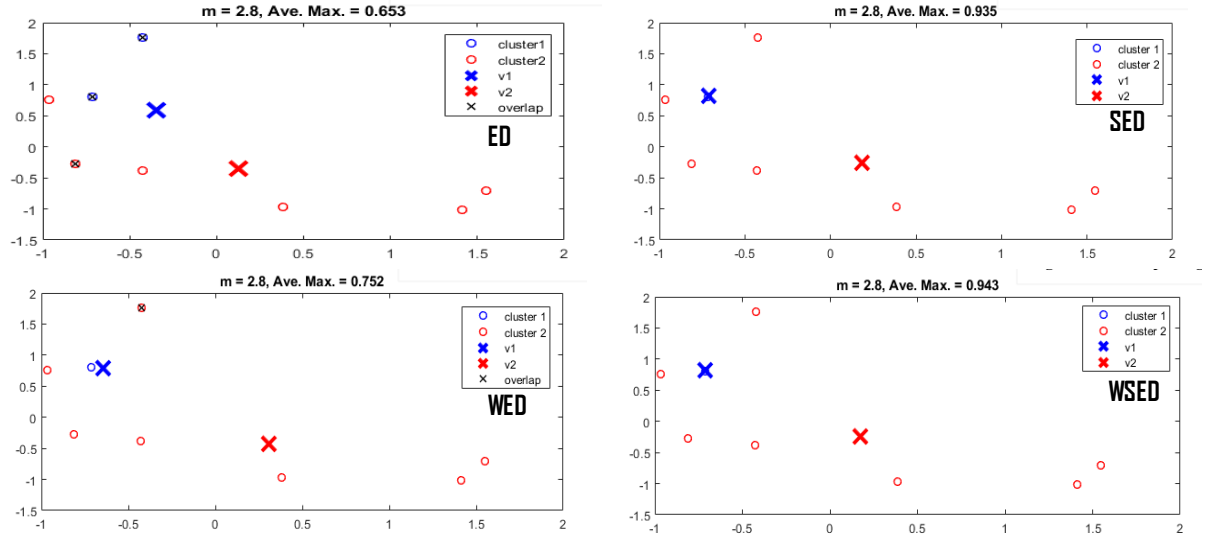

Figure 3. The stations overlap between the clusters when the fuzziness exponent (2.8)

When c = 2, m = 3.6

The results showed that the (ED) measure achieved the best lowest value for the objective function **Obj_Fun**, which represents the amount of error **11.01**, which is the lowest compared to the other measures, but the (WED) measure was the best with respect to the **(Iter., Min_Opt, δ_{XB})** criteria. Therefore, preference can go to the (WED) measure, as it achieved the best results in most criteria, and therefore it contributed to improving the work of the HFCM algorithm and reaching the best results.

As it is clear from **Fig 4** and according to the AverageMax criterion, the fuzzy cluster according to the HFCM algorithm was adequate according to ED and WED, but the percentage of overlap was less according to the (WED) measure, which achieved a percentage of 0.71. This is evidence that it suffers from uncertainty at a lower percentage than this is the case in the case of (ED), which achieved a percentage of 0.60, and this is clear from the many cases of overlap. As for the two measures (SED and WSED), they achieved the highest 85%. The figure shows that the fuzzy cluster has clear tendencies.

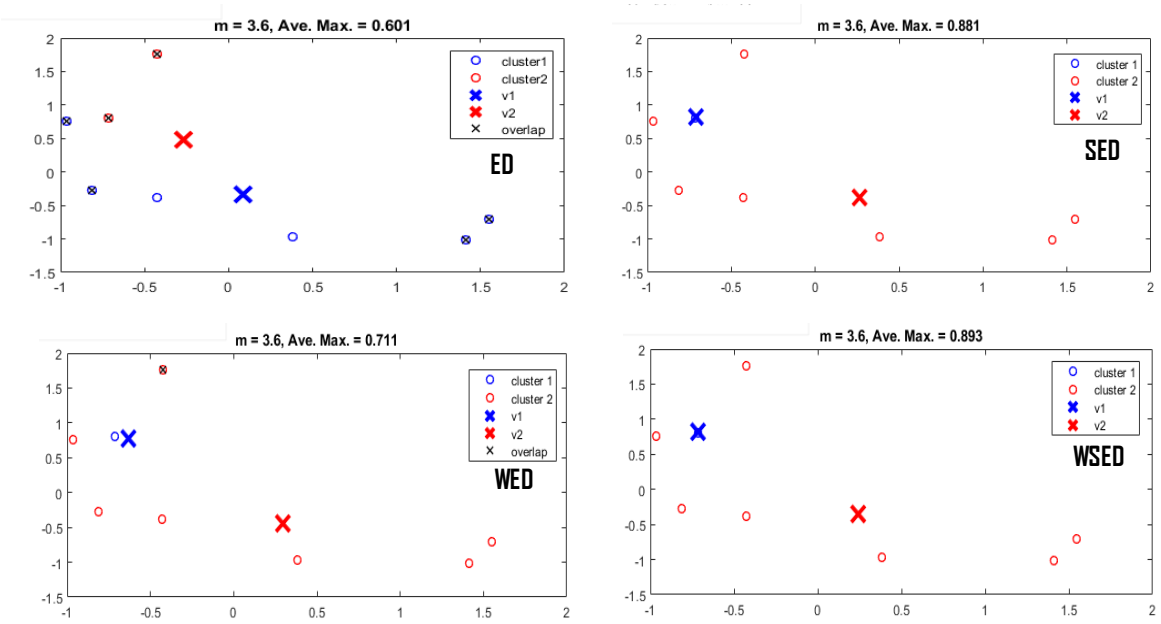

Figure 4. The stations overlap between the clusters when the fuzziness exponent (3.6)

Published Online First: February, 2024
https://doi.org/10.21123/bsj.2024.9516
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Conclusion

Based on the analysis of the experimental findings and the observed impact of the Weighted Euclidean Distance (WED) measure on enhancing the performance of the Fuzzy Mean Clustering algorithm, it can be inferred that the utilization of WED leads to improved accuracy and validity of the resultant cluster structure, particularly in cases when the number of clusters (k) is set to 2. Therefore, the Euclidean Distance (ED) measure ranked second in terms of significance. Hence, researchers propose the utilization of these metrics and their incorporation through the fuzzy clustering technique to ascertain their significance within the domain of clustering. This approach aims to establish a hybrid framework that amalgamates the fuzzy clustering algorithm with artificial intelligence algorithms, thereby mitigating errors, enhancing algorithmic performance, and facilitating its application across diverse domains.

Acknowledgment

The authors express their gratitude and thanks to the engineers of the Technical Department at the Iraqi Ministry of Environment for facilitating the task of obtaining data.

Authors' Declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images, that are not ours, have been included with the necessary permission for re-publication, which is attached to the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee in University of Baghdad.

Authors' Contribution Statement

A. H. M. contributed to the design, acquisition of data, analysis, interpretation of the results, and deriving a new weight for the variance of the fuzzy cluster to modify the distance measure and then to improve the work of the FCM algorithm. M. A. A. contributed to the conception of the idea of the research, drafting the manuscript, revision, and proofreading.

References

1. Dogruparmak SC, Keskin GA, Yaman S, Alkan A. Using Principal Component Analysis and Fuzzy C–means Clustering for the Assessment of Air Quality Monitoring. Atmos Pollut Res. 2014; 5(4): 656-663. https://doi.org/10.5094/APR.2014.075
2. Al-Mousa Y, Al-Jasem A, Dahhand ML. Improve the Result of K-Means Algorithms Using Factor Analysis. Res. J. Aleppo Univ. 2015; (16): 1-22. https://www.academia.edu/23149964
3. Kareem MA, Hamoudi AK, Abdullah AN. Elastic Electron Scattering From 11Li and 12Be Exotic Nuclei in the Framework of the Binary Cluster Model. Iraqi J Sci. 2016; 57(4B): 2664-2676.
4. Hussein Y, Abdel Jalil S. Proposed KDBSCAN Algorithm for Clustering. Iraqi J Sci. 2018; 59(1A): 173-178. https://doi.org/10.24996/ijs.2018.59.1A.18
5. Zhao G, Zhang L, Tang C, Hao W, Luo Y. Clustering of AE Signals Collected During Torsional Tests of 3D Braiding Composite Shafts Using PCA and FCM. Compos B Eng. 2019; 161: 547-554. https://doi.org/10.1016/j.compositesb.2018.12.145
6. Hamed MAR. Application of Surface Water Quality Classification Models Using Principal Components Analysis and Cluster Analysis. J geosci. environ. prot. 2019; 7(6): 26-41. https://doi.org/10.4236/gep.2019.76003
7. Abbas WA. Genetic Algorithm-Based Anisotropic Diffusion Filter and Clustering Algorithms for Thyroid Tumor Detection. Iraqi J Sci. 2020; 61(5): 1016-1026. https://doi.org/10.24996/ijs.2020.61.5.10
8. Shiltagh NA, Hussein MA. Data Aggregation in Wireless Sensor Networks Using Modified Voronoi Fuzzy Clustering Algorithm. J Eng. 2015; 21(4): 42-60. https://doi.org/10.31026/j.eng.2015.04.03
9. Mazhar AN, Naser EF. Hiding the Type of Skin Texture in Mice Based on Fuzzy Clustering Technique. Baghdad Sci J. 2020; 17(3(Suppl.)): 967-

Published Online First: February, 2024
https://doi.org/10.21123/bsj.2024.9516
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

972. https://doi.org/10.21123/bsj.2020.17.3(Suppl.).0967
10. Yaqoob AF, Al-Sarray B. Finding Best Clustering For Big Networks with Minimum Objective Function by Using Probabilistic Tabu Search. Iraqi J Sci. 2019; 60(8): 1837-1845. https://doi.org/10.24996/ijs.2019.60.8.21
11. Abdul-Samad ST, Kamal S. Image Retrieval Using Data Mining Technique. Iraqi J Sci. 2020; 61(8): 2115-2125. https://doi.org/10.24996/ijs.2020.61.8.26
12. Yin Y, Sheng Y, Qin J. Interval Type-2 Fuzzy C-means Forecasting Model for Fuzzy Time Series. Appl Soft Comput. 2022 November; 129: 1-7. https://doi.org/10.1016/j.asoc.2022.109574
13. Mohammed SK, Taha MM, Taha EM, Mohammad MNA. Cluster Analysis of Biochemical Markers as Predictor of COVID-19 Severity. Baghdad Sci J. 2022; 19(6(Suppl.)): 1423-1429. https://doi.org/10.21123/bsj.2022.7454
14. Khouri L, Al-Mufti MB. Assessment of Surface Water Quality Using Statistical Analysis Methods: Orontes River (Case study). Baghdad Sci J. 2022; 19(5): 981-989. https://doi.org/10.21123/bsj.2022.6262
15. Nawaz M, Qureshi R, Teevno MA, Shahid AR. Object Detection and Segmentation by Composition of Fast Fuzzy C-mean Clustering Based Maps. J Ambient Intell Humaniz Comput. 2023; 14(6): 7173–7188. https://doi.org/10.1007/s12652-021-03570-6
16. Setiawan KE, Kurniawan A, Chowanda A, Suhartono D. (Eds.). Clustering Models for Hospitals in Jakarta Using Fuzzy C-means and K-means. Procedia Comput Sci.. 2023; 216: 356–363. https://doi.org/10.1016/j.procs.2022.12.146
17. Hartigan JA, Wong MA. A K-Means Clustering Algorithm. J R Stat Soc Ser C Appl Stat. 1979; 28(1): 100-108. https://doi.org/10.2307/2346830
18. Kadhum IJ, Mohammed AS. Classification & Evaluation of Evidence of Deprivation in Iraq (2009) by using Cluster analysis. J Econ Adm Sci. 2015; 21(82): 391-411. https://doi.org/10.33095/jeas.v21i82.630
19. Ning Z, Chen J, Huang J, Sabo UJ, Yuan Z, Dai Z. WeDIV – An improved k-means clustering algorithm with a weighted distance and a novel internal validation index. Egypt Inform J. 2022; 23(4): 133-144. https://doi.org/10.1016/j.eij.2022.09.002
20. Ashour MA. Optimum Cost of Transporting Problems with Hexagonal Fuzzy Numbers. J Southwest Jiaotong Univ. 2019; 54(6): 1-7. https://doi.org/10.35741/issn.0258-2724.54.6.10
21. Arora HD, Naithani AA. New Definition for Quartic Fuzzy Sets with Hesitation Grade Applied to Multi-Criteria Decision-Making Problems Under Uncertainty. Decis. Anal. J. 2023; 7: 1-10. https://doi.org/10.1016/j.dajour.2023.100239
22. Murfi H, Rosaline N, Hariadi, N. Deep Autoencoder-Based Fuzzy C-means for Topic Detection. Array. 2022; 13: 1-9. https://doi.org/10.1016/j.array.2021.100124
23. El-Zaghmouri B, Abu-Zanona M. Fuzzy C-Mean Clustering Algorithm Modification and Adaptation for Applications and Adaptation for Applications. WCSIT. 2012; 2(1): 42-45.
24. Javadi S, Rameez M, Dahl M, Pettersson MI. Vehicle Classification Based on Multiple Fuzzy C-Means Clustering Using Dimensions and Speed Features. Procedia Comput Sci.. 2018; 126: 1344–1350. https://doi.org/10.1016/j.procs.2018.08.085
25. Hameed SM, Mohammed MB, Attea BA. Fuzzy Based Spam Filtering. Iraqi J Sci. 2015; 56(1B): 506-519.
26. Goyal LM, Mittal M, Sethi JK. Fuzzy Model Generation Using Subtractive and Fuzzy C-Means Clustering. CSI trans ICT. 2016; 4(2-4): 129–133. https://doi.org/10.1007/s40012-016-0090-3
27. Oliveira JV, Pedrycz W, editors. Advances in Fuzzy Clustering and its Applications. 1st ed. The Atrium, Southern Gate, Chichester: John Wiley & Sons Ltd; 2007. 454p. https://doi.org/10.1002/9780470061190
28. Abdulghafoor SA, Mohamed LA. Using Some Metric Distance in Local Density Based on Outlier Detection Methods. J Posit. Psychol. Wellbeing. 2022; 6(1): 189-202.
29. Ahmad MR, Afzal U. Mathematical Modeling and AI Based Decision Making for COVID-19 Suspects Backed by Novel Distance and Similarity Measures on Plithogenic Hypersoft Sets. Artif Intell Med. 2022; 132: 1-8. https://doi.org/10.1016/j.artmed.2022.102390
30. Wierzchon ST, Kłopotek MA. Modern Algorithms of Cluster Analysis. 1Ed ed, Springer, Cham; 2018; 34.
31. Mota VC, Damasceno FA, Leite DF. Fuzzy Clustering and Fuzzy Validity Measures for Knowledge Discovery and Decision Making in Agricultural Engineering. Comput Electron Agric. 2018; 150: 118-124. https://doi.org/10.1016/j.compag.2018.04.011

Published Online First: February, 2024
https://doi.org/10.21123/bsj.2024.9516
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

تحسين خوارزمية عنقدة المتوسطات الضبابية باستعمال مقياس جديد للمسافة الموزون بالانحراف المعياري

احمد هشام محمد[1]، مروان عبد الحميد عاشور[2]

[1]قسم الاحصاء، كلية الادارة والاقتصاد، جامعة البصرة، البصرة، العراق.
[2]قسم الاحصاء، كلية الادارة والاقتصاد، جامعة بغداد، بغداد، العراق.

الخلاصة

ان الهدف الرئيس من هذه الورقة هو تقديم منهج جديد لمعالجة خوارزمية عنقدة المتوسطات الضبابية FCM التي تُعد من اهم واشهر الخوارزميات التي عالجت ظاهرة عدم اليقين في تشكيل العناقيدعلى وفق نسب التداخل، ان من اهم المشكلات التي تواجه هذه الخوارزمية هو اعتمادها بشكل اساسي على مقياس المسافة الاقليدية وبطبيعة الحال هذا المقياس يجعل العناقيد المشكلة تاخذ الشكل الكروي الذي يكون غير قادر على احتواء الحالات المعقدة او المتداخلة، لذا حاولنا من خلال هذه الورقة اقتراح مقياس جديد للمسافة حيث تمكنا من اشتقاق صيغة لتباين العنقود الضبابي ليدخل كوزن على صيغة المسافة الاقليدية علاوة على ذلك تم معالجة حساب مصفوفة التقسيمات من خلال استعمال خوارزمية K-Means وخلق بيئية هجينة بين الخوارزمية الضبابية والخوارزمية الحادة، وللتحقق مما تم عرضه تم استعمال المحاكاة التجريبية ومن ثم تطبيقها على الواقع باستعمال البيانات البيئية للفحص الفيزيائي والكيميائي لمحطات فحص المياه في محافظة البصرة، وقد اثبتت النتائج التجريبية ان مقياس المسافة المقترح (WED) كانت له الافضلية على تحسين عمل خوارزمية HFCM من خلال معيار (Obj_Fun, Iteration, Min_optimization, Good fit clustering and overlap) عندما (k = 2,3)، وبموجب نتائج المحاكاة تم اختيار c = 2 وتشكيل العناقيد للبيانات الحقيقية وتمكنت من ايجاد افضل دالة هدف (23.93، 22.44، 18.83) عند درجات التضبيب (1.2، 2، 2.8) بينما حسب درجة التضبيب (m = 3.6) كانت دالة الهدف لمقاس (ED) هي الاقل ولكن كانت المعايير (Iter. =2 ، Min_optimization = 0، δ_{XB}) تؤكد على ان (WED) هي الافضل.

الكلمات المفتاحية: العنقدة، مقاييس المسافة، عنقدة المتوسطات الضبابية، المنطق الضبابي، الخوارزمية الهجينة.