Crow Search Optimization based Fuzzy C-Means Clustering for Optimal Centroid Initialization

S Parvathavarthini*
Assistant Professor, Department of Computer Technology,
Kongu Engineering College, Perundurai – 638 060 ,India
varthinis@gmail.com

Dr. N. Karthikeyani Visalakshi
Assistant Professor, Department of Computer Science,
Government Arts & Science College, Kangeyam, India
karthichitru@gmail.com

Dr. S. Shanthi
Assistant Professor (Sl. G), Department of Computer Applications,
Kongu Engineering College, Perundurai – 638 060, India
shanthis@kongu.ac.in

J. Madhan Mohan
Associate Professor, Department of Electronics and Communication Engineering,
Erode Sengunthar Engineering College, Perundurai, India
madhan79in@gmail.com

***Abstract*— Clustering is an unsupervised technique that segregates objects into several groups based on their qualities. In soft clustering algorithms like Fuzzy C-Means, the choice of initial cluster centers is done in a random fashion and this heavily influences the solution. Due to this random selection, there is a possibility of delay in convergence rate or there will be a chance of getting stuck in local optimal solution. To solve these problems, an optimization algorithm can be employed. Crow search is one of the emerging optimization algorithms that aim at attaining global optima and faster convergence rate using only two user-defined parameters. This work throws light on a novel work that combines crow search algorithm with Fuzzy C-Means clustering. The experimental analysis is done using benchmark datasets from UCI data repository and artificial datasets from the University of Eastern Finland. In order to evaluate the performance of the FCM-Crow Search algorithm, three aspects like error rate, objective function value and cluster validity indices are considered. The results of benchmark datasets are compared with K-Means, FCM-PSO and ACPSO algorithms and the proposed algorithm is found to be more efficient.**

Keywords— Data Mining, Clustering, Optimization, Fuzzy C-Means, Crow search, Hybrid FCM crow search, Awareness Probability, Flight Length

I. INTRODUCTION

Data Mining is the process of identifying some patterns and extracting valuable and useful information from large volumes of data. There are several techniques for mining the data such as clustering, classification, prediction, association rule mining, etc. Today's world is filled with enormous data. It is tedious to do all processing manually. Therefore intelligent data analysis techniques are to be designed to analyze data.

The goal of data clustering [1], also known as cluster analysis, is to discover the natural grouping(s) of a set of patterns, points, or objects. Clustering finds its application in various domains like market basket analysis, business intelligence, pattern recognition, medical image analysis, gene clustering, web document search, satellite image analysis, etc. Clustering algorithms are generally categorized into hard and soft. Hard clustering algorithms like K-Means designate an object to exactly one cluster. Soft clustering algorithms like Fuzzy C-Means allow an object to be a part of various clusters based on the membership value.

Clustering algorithms can be classified as hierarchical or partitional. Hierarchical clustering algorithms are further categorized into agglomerative and divisive based on the clustering fashion they follow. If each data point is considered as a cluster and then the similar pairs are merged, it is termed as agglomerative. If all the objects are put in a single cluster and then continuously divided to form smaller clusters, it is known as divisive. In partitional clustering, the data objects are simultaneously clustered by partitioning the data and mapping it on to the d-dimensional feature space.

The major difficulty with any clustering algorithm is that there is no prior knowledge available about the dataset. Fuzzy C-Means (FCM) developed by Bezdek et al [2] allows each of the data objects in the dataset to be a part of different clusters with different belongingness. So FCM can be effectively used to model real world scenarios.

While randomly choosing the seed points, the convergence time of the clustering algorithm increases. Instead, an optimization algorithm helps in selecting the best among the feasible solutions for a problem. It allows exploring the problem space subject to certain constraints and resulting in the minimization or maximization of the criterion function. This also avoids the problem of getting trapped into local minima. Also, an insight into the shape of the clusters is necessary before grouping the data. Clusters with arbitrary shapes or containing noisy data or outliers need to be clearly identified. So, there is a need for hybridizing FCM with optimization algorithms in order to efficiently cluster the objects.

There are quite a lot of nature-inspired metaheuristic optimization algorithms like Genetic Algorithm (GA), Ant Colony Optimization Algorithm (ACO), Simulated Annealing (SA), Particle Swarm Optimization (PSO), Tabu Search (TS), Cat Swarm Optimization (CSO), Artificial Bee Colony (ABC), Cuckoo Search Algorithm (CS), Gravitational Search Algorithm (GS), Firefly Algorithm (FA), Bat Algorithm (BA), Bacterial Foraging Algorithm (BFA), Wolf Search Algorithm (WSA), Krill Herd (KH), etc. Several researchers have combined FCM with these optimization methods. This paper presents a new way of combining FCM with one of the upcoming meta-heuristic algorithms inspired by the behavior of crows.

This paper is organized as follows: section 2 focuses on the FCM algorithm section 3 describes the existing works in the literature, section 4 elucidates the crow search algorithm, section 5 explains the proposed methodology, section 6 concentrates on the experimental study, section 7 illustrates the results of the proposed method and section 8 presents the conclusion of this research work.

II. Fuzzy C-Means Algorithm

FCM (Bezdek et al, 1984) is the most popular soft clustering algorithm. In fuzzy sets, the uncertainty in the dataset is preserved by representing the data as a combination of membership and non-membership values. Let D= {d_1, d_2… d_n} be the data set and D has to be partitioned into C clusters based on the features of the dataset. The data has to be fuzzified before proceeding with the execution of clustering algorithm.

A membership function μ_i (d_j) for the fuzzy representation is defined by

$$\overline{\mu_i}(d_j) = \frac{d_{ij} - \min(d_j)}{\max(d_j) - \min(d_j)} \quad (1)$$

Where i=1,2,…,n and j=1,2,…,t. Here n is the number of instances in the dataset and t is the number of attributes in each instance of the dataset. The initial task is to estimate the similarity between the datasets using any distance measure like Euclidean distance. The belongingness of an object d_i to the cluster c_j is given by

$$U_{ij} = \frac{1}{\sum_{r=1}^{c}\left(\frac{dis(,V_i)}{dis(,V_r)}\right)^{\frac{2}{m-1}}}, 1 \le i \le c, 1 \le j \le n, m = 2 \quad (2)$$

The objective function of FCM algorithm can be given as follows

$$J_m(x, y) = \sum_{i=1}^{c} \sum_{j=1}^{p} U_{ij}{}^{m} \left\| X_j - V_i \right\|, \ 1 \le m \le \infty \quad (3)$$

The centroids are updated using the following formula

$$\nu_j = \left(\sum_{i=1}^{n} (\mu_{ij})^m x_i\right) / \left(\sum_{i=1}^{n} (\mu_{ij})^m\right), \forall \ j = 1, 2, \ldots, c \quad (4)$$

The centroids are updated and again the membership values are computed. The process is repeated until the consecutive iterations produce the same centroids or until the objective function is saturated. Finally, the defuzzification process is done by finding the cluster to which the object has a higher membership value. This will serve as the index of the cluster for that object.

The main drawback of FCM algorithm is that it doesn't allow the user to thrive for a global solution. To avoid this problem, optimization algorithms can be run first and the best outcome of these algorithms can be given as input to the FCM algorithm.

III. LITERATURE REVIEW

Optimization is an applied science that investigates the best out of the possible values for the parameters that a problem may obtain under specified constraints (Corne et al, 1999), (Horst et al, 2000). Optimization produces a feasible solution to a problem. The two main phases in optimization algorithms are exploration and exploitation where exploration deals with searching of best local solutions and exploitation concentrates on reaching a global optimum solution [3].

Bio-inspired optimization algorithms have become the state of art and thus many researchers are attracted towards these algorithms. They have the ability to produce a near optimal solution for any problem. Several such algorithms have been designed by great researchers and are available in the literature.

Ant Colony Optimization (ACO) was proposed by Dorigo [4]. In this population based approach, the ants are the search agents that communicate using the pheromone trails. When more ants follow the same path, the shortest path can be found. Tulin Inkaya et al [5] introduced two objective functions called adjusted compactness and relative separation and used ACO to construct neighborhood which results in subclusters and thus reduced the dataset by considering the only points that lie on the boundaries.

Various applications of ACO are surveyed in different domains and also a modified ACO model is proposed by Chandra Mohan and Baskaran [6] for solving network routing problem. Berat Dogan and Korurek [7] outlined a new kernel FCM algorithm based on ACO and applied it for ECG beat dataset.

A novel image segmentation method for remote sensing images is designed by Qian Wang et al [8] which combined ACO with FCM so as to minimize the loss of information and reduce the sensitivity to noise. Kanade and Lawrence [9] utilized ACO to cluster the objects and reformulated the cluster centers using FCM and Hard C-Means and to determine the number of clusters in each dataset.

Eberhart [10] proposed particle swarm theory and described its applications in neural network training and robot task learning. Mohamed Alia et al [11] grouped the pixels in an image based on their frequency of occurrence and segmented MR images using FCM. The clusters are initialized with the help of Harmony search algorithm.

When a massive star collapses, a black hole is formed. It pulls the other objects that cross its boundary so that those objects vanish. In the black hole algorithm, a random population of stars is generated, the fitness is evaluated and the best candidate is selected to be the black hole. All the other candidates are moved towards the black hole by changing position in every iteration. If a star reaches a location with lower cost than the black hole, then their locations are exchanged. Hatamlou [12] explains how this blackhole optimization can be used for clustering.

The intelligent foraging behavior of honey bees is simulated and used for fuzzy clustering by Karaboga [13]. Employee bees collect nectar and share position of food with onlooker bees. The position of food source indicates the solution and the amount of nectar indicates the quality of a solution.

M. Krishnamoorthi and Natarajan [14] provided a new FCM operator to introduce the scout bee. If the solution is abandoned, a new bee position will be randomly assigned in ABC algorithm. The scout bee is added to the solution by processing the best fitness values obtained from the previous solution.

Bat algorithm is based on echo ranging behavior of bats. Bats use echo ranging to measure distance and the difference between their prey and background barriers. Ye et al [15] use bat algorithm to find the optimal thresholds for image segmentation by maximizing the fuzzy entropy.

Katarya et al [16] utilized grey wolf optimizer for feature reduction and created a recommender system for watching movies based on age, gender, etc. Mehrabian [17] proposed the Invasive Weed Optimization (IWO) which is based on weed colonization. The population is initialized; the seeds produced by each individual are calculated and are distributed around the parent. The seeds and weeds are sorted based on fitness. The fitter weeds go to the next iteration while the others are eliminated. Thus IWO is combined with kernel Possibilistic C-Means [18] to yield better results.

Krill herd is the idealization of herding of krill swarms in the sea. Jensi [19] improved krill herd optimization by introducing a global exploration operator. Li et al [20] used the elitism strategy i.e. instead of updating the positions of all the krill individuals, certain best krill individuals are retained in memory, and then all the krill are updated by three motions. Finally, certain worst krill individuals in the new population are replaced by the memorized best ones in the last generation. The best individual forms the initial centroids for FCM algorithm.

Adan Jose Garcia and Flores [21] reviewed the major nature-inspired meta-heuristic algorithms for finding the number of clusters in any dataset automatically. Also, the encoding schemes, cluster validity indices and proximity measures are discussed in this work. A Tabu search approach to the fuzzy clustering problem is proposed by Miguel Delgado [22]. Mane and Gaikwad [23] reviewed the effectiveness of hybridizing the traditional clustering algorithms with nature-inspired techniques.

Artificial Fish Swarm Algorithm (AFSA) is developed by Xiaoli Li in 2003. Fish swims towards the food and it always searches for a higher concentration of food. Fish assembles in groups to capture colonies and protect themselves from enemies. Si He et al [24] improved AFSA with adaptive visual and adaptive step by combining with FCM and resulting in a higher convergence rate. Sweta and Sahana [25] discussed the algorithms inspired by the insects like Honey bee, cockroach, firefly, glowworm, mosquito, superbug, termite and wasps.

In glowworm swarm optimization, each glowworm maps the objective function value at its current location into a luciferin value and constructs a neighborhood. Then, it starts moving towards a brighter glowing neighbor and the positions are updated and the radius is fixed. Huang et al [26] combined this with K-Means clustering algorithm to deal with non-convexity data set.

Satyasai Nanda and Panda [27] developed a detailed survey on nature inspired metaheuristic algorithms for partitional clustering and presented an overview of single objective and multi-objective algorithms for clustering. Also, some real life applications are discussed in this work.

Many of these algorithms need several parameters to be tuned by the user. But crow search has a relatively reduced number of parameters such as flight length and awareness probability. This leads to a simple and efficient algorithm. The application of Crow search optimization to clustering has not yet been discussed by any author.

IV. Crow Search Algorithm

Crows are renowned for their unity and intelligence. They have some special characteristics like recognizing faces, self-awareness and memorizing food sources. Based on these illustrious features, a new population-based metaheuristic algorithm [28] that simulates the behavior of these intelligent birds is designed in order to solve optimization problems and to achieve a global optimal solution.

Crows live in flocks and they make note of other birds to know where they hide food. They are stealthy by nature and are cautious in hiding their caches from being identified by other birds with a probability. A crow always tries to follow another to do thievery. Crows secure their caches from being filched by others. Based on all these unique traits, the Crow Search Algorithm (CrSA) has the goal of finding a better food source or hiding place. The algorithm is so simple that it needs to handle two parameters: Awareness Probability (AP) and Flight Length (FL).

Let D be the problem dimension and N be the population size. The position of the crow i at time t is given as $X^{i,it}=[x_1^{i,t}, x_2^{i,t}, \ldots, x_d^{i,t}]$ where i=1,2,…,N; t=1,2,…,itmax, and itmax is the number of iterations. The hiding position of crow A at time t is given by $m^{A,t}$.

Suppose crow B wants to visit its hiding place $m^{B,t}$, and if crow A chooses to follow crow B, this results in two possible states such as

- crow B is not aware of crow A following it and thus crow A reaches the hiding place of crow B
- crow B is conscious that it is being followed by crow A and thus changes its position to any random flight direction in the search space.

V. Hybrid FCM-Crow Search Algorithm

The purpose of combining a clustering algorithm with an optimization algorithm is twofold:

- Firstly, it aims at achieving global optima
- Secondly, the initial seed points can be chosen as a result of running the optimization algorithm and a faster move towards the solution can be made.

This section presents the novel hybrid FCM-Crow search algorithm. Here the cluster centers are the solutions and they are simulated as crows. The best crow to be taken as the initial seed is found and then the FCM algorithm is run to obtain the clusters. For a dataset with n clusters and m attributes, the size of the crow is m × n.

A. FCM_Crow Search Algorithm (FCM_CrSA)

Initialize the parameters like population or flock size N, the number of clusters C, Maximum number of iterations itmax, flight length fl and awareness probability AP. Initialize the position of the crows pos by generating a random matrix of cluster centers and encode the data objects.

The dataset is converted into fuzzy representation using (1). Initially, the crows do not have any experience. So their memory is initialized same as the initial position assuming that they have hidden the food at their initial position. For each, crow the distance measure is computed and the membership values of each object to various clusters is found using (2).Then the fitness of initial positions is evaluated using the objective function in (3). Now, assume that the crow B wants to visit its hiding place, then any crow A is randomly chosen to follow it. In this case, there are two possible variations in the behavior of crow B. If crow B is not aware of crow A following it and thus crow A reaches the hiding place of crow B, new position of B is computed using

$$x_{A,t+1} = x_{A,t+1} + r_A.\ fl^{A,t}.(m^{B,t} - x^{B,t}) \qquad (5)$$

If crow B is conscious that crow A is following, it chooses a random new position to fool crow A. The feasibility of new position is checked and position is updated only if it is feasible. Otherwise, no change to the position is made.

Once again, the fitness of new position of crows is evaluated. If the quality of the new position is not better than the earlier position, the memory is not updated. Otherwise, the memory of crows is updated using the following equation

$$m^{A,t+1} = x^{A,t+1} \qquad (6)$$

The process of random selection of a crow, generating new position, checking the feasibility of positions, evaluating fitness function of new positions and updating the memory are repeated until the stopping criterion for crow search algorithm is reached i.e. the maximum number of iterations are reached.

Finally, the best initial centroids are found based on the minimum fitness value obtained. The membership value is calculated using (2) and the fitness is obtained using (3). The cluster centers are updated using (4). The procedure is repeated until FCM converges i.e. either the same set of centroids are obtained for two consecutive iterations or if the objective function is stabilized. The highest membership value for an object to a cluster indicates its belongingness to that cluster. Similarly, the clusters for all the instances in the dataset are found.

B. Pseudocode for FCM-Crow search algorithm

Initialize the population of N crows, C clusters and maximum iterations itmax

Assign initial values for flight length and awareness probability.

Initialize the position of crows randomly with N×D dimension search space and Initialize the memory of the crows equivalent to the position of crows.

***While** run < maxruns*

  ***while** t < itmax*

   ***for** A = 1 : N*

        Calculate membership matrix using (2)

        Calculate the fitness of each crow using (3)

        Randomly choose one of the crows to follow

        ***If** $r_B >= AP^{B,t}$ calculate new position using (5)*

          Else $x^{A,t+1}$= a random position of search space

       end if

   end for

   Check the feasibility of new positions

   ***If** it is feasible, Evaluate the cost of new position and Update the memory*

     ***if** $f(x^{A,t+1})$ is better than $f(m^{i,t})$*

        update memory using (6)

else $m^{A,t+1}=m^{A,t}$

end if

end while

Find the best position of the crow that minimizes the fitness function

while *iter < maxiterations*

Calculate membership matrix using best position

Update cluster centers using (4)

end while

end while

C. Benefits of Proposed Algorithm

The benefits of proposed method are

- Global optimal solutions are achieved with well-separated and compact clusters
- Good solutions are memorized and the best solutions found are used to find the better positions
- A non-greedy algorithm in which the crow moves to a new position if the generated solution is not better than its current position
- Novel hybridization is applied to reduce convergence time and the algorithm is efficient in terms of validity indices, accuracy and fitness function

VI. EXPERIMENTAL STUDY

Experiments are conducted to measure the performance of the proposed algorithm against existing methods. The algorithm is implemented in MATLAB. The Initial values to be set include the number of crows=20, maximum iterations=50, flight length=2 and awareness probability=0.1. The experiment is repeated for 100 runs and the best values are selected.

In order to quantitatively evaluate the performance of the proposed method, six benchmark datasets from UCI repository repository [29] and two artificial datasets are taken. Various measures are computed and the results are compared with partition based and optimization based clustering algorithms like K-Means, FCM, FCMPSO and ACPSO. The comparison is done in three aspects such as error rate, objective function and cluster validity indices. The results show that the proposed algorithm is efficient.

The details of the dataset are shown in Table I. Iris, wine and glass have less number of instances and the remaining three datasets Wisconsin Breast Cancer (WBC), CMC and Vowel have more than 680 instances. Two artificial datasets with a large number of clusters are also taken from Speech and image processing unit, University of Eastern Finland [30] and internal indices are calculated for them. These datasets A1 and A2 shown in Table II have an ample number of instances viz. 3000 and 5250. They are considered to evaluate the outcome of the proposed method on voluminous data.

A. Error Rate and Objective Function

For each dataset, the classification error percentage is calculated. It is the percentage of wrongly classified objects in the test datasets. The error rate of the proposed FCM-crow search algorithm is computed by using the following formula

$$ER = \frac{No.\ of\ misclassified\ samples}{No.\ of\ instances\ in\ the\ dataset} \times 100 \qquad (7)$$

The error rates obtained by the FCM-Crow search algorithm are lesser when compared to the other algorithms in Table III. The results are compared to ACPSO [31] and are found to be good with five out of six datasets. For vowel dataset, the performance of ACPSO is slightly high than FCM-Crow search. It is also evident from the Table III that the average error rate considered from the 100 runs is also significant. The comparison of error rates is shown in Fig. 1. It is clear from the figure that the error rate is significantly reduced for all the datasets except for vowel dataset. The objective function values in Table IV outperform the other methods. For all the datasets, the best and worst values are improved. In the case of iris dataset, the mean and standard deviation values of proposed method are found to be the next best value after ACPSO and the best standard deviation is achieved in the case of wine dataset for ACPSO. The average values for objective function are higher for the proposed method than ACPSO as the crow search algorithm converges very faster. For vowel dataset, ACPSO shows a slightly decreasing error rate. The proposed method yields a lower error rate than the other algorithms.

TABLE I. DETAILS OF BENCHMARK DATASETS

Dataset	Number of clusters	Number of attributes	Number of Instances (size of each class)
Iris	3	4	150 (50,50,50)
Wine	3	13	178 (59,71,48)
Glass	6	9	214 (70,17,76,13,9,29)
WBC	2	9	683 (444,239)
CMC	3	10	1473 (629,333,511)
Vowel	6	3	871 (72,89,172,151,207,180)

TABLE II. DETAILS OF ARTIFICIAL DATASETS

Dataset	Number of clusters	Number of attributes	Number of Instances
A1	20	2	3000
A2	35	2	5250

TABLE III. COMPARISON OF BEST AND AVERAGE ERROR RATES OBTAINED

Dataset	Error Rates	K-Means[a] (in %)	FCM (in %)	FCM-PSO (in %)	ACPSO[a] (in %)	FCM-Crow Search (in %)
Iris	Best	10.67	10.33	10.33	8.00	**6.00**
	Average	17.8	13.22	11.67	9.80	**9.14**
Wine	Best	29.78	29.70	27.86	28.09	**25.82**
	Average	31.12	30.82	30.05	28.23	**27.14**
Glass	Best	43.11	41.22	39.25	N/A	**37.49**
	Average	45.72	43.89	40.91	N/A	**40.26**
Cancer	Best	3.95	3.66	3.66	**3.51**	**3.51**
	Average	4.08	4.02	4.31	**3.51**	3.56
CMC	Best	54.45	54.45	53.50	54.38	**51.24**
	Average	54.49	55.26	54.04	54.38	**53.07**
Vowel	Best	42.02	43.79	42.32	**41.10**	41.76
	Average	44.26	42.58	41.97	**41.69**	41.83

[a]The results of K-Means and ACPSO can be found in Li-et al (2012). N/A: data not available.

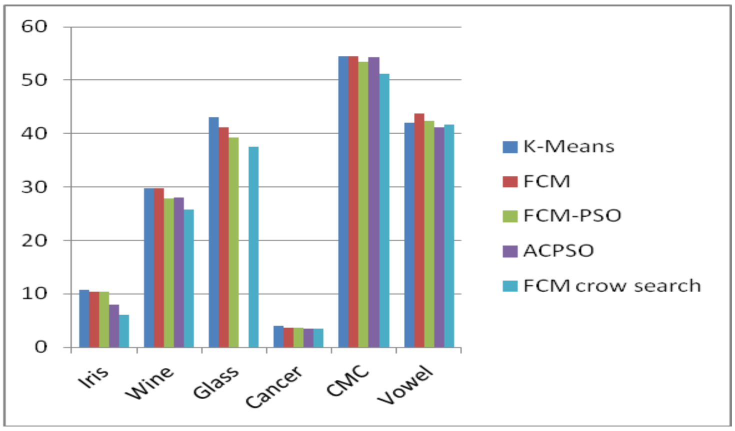

Figure 1. Comparison of error rates

TABLE IV. COMPARISON OF FITNESS VALUES

Dataset	Objective Values	K-Means[a]	FCM	FCM-PSO	ACPSO[a]	FCM-Crow Search
Iris	Best	97.33	97.02	97.01	96.66	**96.57**
	Worst	123.96	121.98	120.672	N/A	**104.785**
	Mean	106.5766	105.3439	105.3406	**96.66**	98.31
	Std	12.938	11.513	10.2	**0.0001**	3.06
Wine	Best	16555.68	16548.74	16389.29	16292.18	**16078.35**
	Worst	18294.85	18203.24	16976.93	N/A	**16032.54**
	Mean	17251.35	17043.78	16945.28	16292.31	**16091.66**
	Std	874.148	817.03	52.37	**0.03**	4.50
Glass	Best	215.73	218.65	213.84	N/A	**209.37**
	Worst	227.35	225.17	227.89	N/A	**229.03**
	Mean	218.70	219.66	215.43	N/A	**210.35**
	Std	2.456	6.23	4.87	N/A	**1.04**
Cancer	Best	2988.43	2979.35	2977.81	2964.39	**2957.02**
	Worst	2999.19	2956.84	2974.53	N/A	**2960.16**
	Mean	2988.99	2941.01	2976.04	2964.42	**2957.05**
	Std	2.469	2.25	1.48	0.03	**0.02**
CMC	Best	5703.20	5672.87	5627.10	5532.19	**5499.36**
	Worst	5704.57	5699.34	5618.42	N/A	**5504.17**
	Mean	5705.37	5673.81	5629.53	5532.20	**5499.19**
	Std	1.033	1.01	0.	**0.01**	**0.01**
Vowel	Best	149398.60	149536.39	149026.44	148970.84	**148895.27**
	Worst	162455.69	149592.65	149104.25	N/A	**148902.35**
	Mean	151987.98	149570.04	149097.83	149051.84	**148839.84**
	Std	3425.250	135.82	106.85	67.27	**34.17**

[a]The result of K-Means is found in Jensi and Jiji (2016) and ACPSO is found in Li- et al (2012).N/A: Data not available

B. *Validity Measures*

Cluster evaluation can correlate the structures found in the data with the externally provided class information and are used to check whether data consists of non-random structures. If the number of clusters for a dataset is not known, cluster evaluation helps in fixing the ideal number of clusters and assists in ranking the alternative clustering arrangements with regard to their quality.

Cluster validation is the predominant way of judging the performance of a clustering algorithm. There are three categories of validation indices such as internal indices, external indices and relative indices. In order to use external validity measures, there is a need for apriori knowledge about data [32].

Internal validation measures are based on two essential factors: separation and compactness. Separation indicates the degree with which a cluster is well-separated from others and compactness shows the relative closeness among the objects in a cluster. Thus it is essential to measure how far the objects in the dataset are clustered based on their intrinsic characteristics.

Four famous indices for measuring the cluster accuracy have been considered to evaluate benchmark datasets. Out of these, Rand Index, Adjusted Rand Index and F-Measure are the external indices and DB index is an internal measure.

A greater value closer to one indicates good performance in F-Measure, Adjusted Rand index and Rand indices. Lesser value results in good clusters in case of DB index. The performances of all the six algorithms have been evaluated using these indices.

In order to minimize the effect of the stochastic nature of FCM-Crow search on the indices and on the number of clusters, 100 independent runs are executed and finally, the best values from each run is considered for clear decision making.

1) Rand Index

If two instances with similar characteristics are assigned to the same cluster, a true positive (TP) decision is taken; a true negative (TN) decision assigns two dissimilar documents to different clusters. If two dissimilar documents are assigned to the same cluster, the state is said to be False Positive (FP) decision. A False Negative (FN) decision assigns two similar documents to different clusters. The Rand index [33] measures the percentage of decisions that are correct.

$$RI = \frac{TP + TN}{TP + FP + FN + TN} \quad (8)$$

The experiment results in Table V show that the Iris dataset produces the significant value for best and average rand index value as 0.9249 and 0.8579 respectively. The overall minimum value for the worst case is obtained as 0.3144 for the glass dataset which consists of spherical clusters and in the average case, the overall least value is for CMC dataset. The reason behind this may be the higher number of clusters in glass dataset and a large number of instances belonging to the CMC dataset.

TABLE V. BEST, WORST AND AVERAGE RAND INDEX VALUES

Dataset	Best	Worst	Average
Iris	0.9249	0.4289	0.8579
Wine	0.7549	0.3184	0.7211
Glass	0.7324	0.3144	0.7129
Cancer	0.8119	0.4632	0.8002
CMC	0.6620	0.3534	0.5935
Vowel	0.8292	0.3864	0.7416

The average case also shows significant values for Iris and cancer datasets, considerable results for wine, glass and vowel datasets and least values for CMC. Thus the results indicate the percentage of right decisions.

2) F-Measure

The F-Measure [34] is an external index. It is the harmonic mean of the precision and recall coefficients. If the precision is high and recall value is low, this results in a low F-measure. If both precision and recall are low, a low F-measure is obtained. On the other hand, if both are high, a high F-measure value is obtained.

F-Measure can be computed using the formula

$$F = \frac{2TP}{2TP + TN + FP} \tag{9}$$

Table VI shows the F-measure values for all six datasets. The iris and cancer datasets consist of elongated shape and spatially well-separated clusters leading to a higher value of F-measure such as 0.9258 and 0.8126 respectively.

The least average value is obtained as 0.5253 for the vowel dataset and the least value in the worst case is 0.3353 which is obtained for the glass dataset. This indicates that the higher the number of clusters, the F-measure value decreases to a great extent. The average values are more close to the best values.

TABLE VI. BEST, WORST AND AVERAGE F-MEASURE VALUES

Dataset	Best	Worst	Average
Iris	0.9258	0.4646	0.8439
Wine	0.7458	0.4712	0.7016
Glass	0.6980	0.3353	0.6095
Cancer	0.8126	0.4612	0.7512
CMC	0.6178	0.4091	0.6002
Vowel	0.6729	0.3594	0.5253

3) Adjusted Rand Index

Adjusted Rand Index was proposed by Hubert and Arabie[35]. The peculiarity of this measure is that it is not sensible to the number of clusters. Thus, this measure can be used to compare two partitions with varying cluster numbers and find the exact number of clusters. The range of permissible values falls within -1 to +1. A value of 1 indicates a perfect partition similar to the apriori class label. Negative values signify the inability to discriminate the clusters and the values near zero show the random solution. Let M={m_1,m_2, …., m_c} be the resulting partition of a clustering algorithm with C clusters and P={p_1,p_2,…,p_D} be the partition known from the class labels of benchmark datasets with D clusters, then ARI is calculated as

$$ARI = \frac{\sum_{i=1}^{C}\sum_{j=1}^{D}\binom{n_{ij}}{2}\binom{n}{2}^{-1}\sum_{i=1}^{C}\binom{n_i}{2}\sum_{j=1}^{D}\binom{n_j}{2}}{\frac{1}{2}\left[\sum_{i=1}^{C}\binom{n_i}{2}+\sum_{j=1}^{D}\binom{n_j}{2}\right]-\binom{n}{2}^{-1}\sum_{i=1}^{C}\binom{n_i}{2}\sum_{j=1}^{D}\binom{n_j}{2}} \tag{10}$$

Where C is the number of clusters, n_{ij} is the number of objects that belong to the group m_i and p_j, n_i is the number of objects that belong to the group m_i, n_j is the number of objects that belong to the group p_i and n is the total number of objects. Table VII shows the results of ARI and it should be noted that the higher values are achieved in the order of datasets that start with iris, cancer and wine that also contribute to the first three datasets with lower error rate. The least values in the worst case and average case are obtained in the case of higher cluster numbers that is, for the glass and vowel datasets. The higher value for the best case is 0.8306 which also has the highest value as 0.8208 for the average case.

TABLE VII. BEST, WORST AND AVERAGE ARI VALUES

Dataset	Best	Worst	Average
Iris	0.8306	0.0004	0.8028
Wine	0.7488	0.0015	0.7152
Glass	0.3020	0.0009	0.2596
Cancer	0.8157	0.0321	0.7259
CMC	0.5716	0.0301	0.4348
Vowel	0.4457	0.0215	0.3018

4) DB index

The Davis-Bouldin index [36] is based on a ratio of within cluster and between cluster distances. This shows good performance when the value is less.

The formula for DBIndex can be given as

$$\frac{1}{k}\sum_{i=1}^{k} max_j \frac{s(C_i)+s(C_j)}{d_c(C_i,C_j)} \tag{11}$$

Where k is the number of clusters, s(C) is the average distance among the instances in cluster C, $d_c(C_i, C_j)$ measures the distance between the centers of C_i and C_j.

Table VIII shows the results for DB index values and the overall best value is achieved by cancer dataset which has the least error rate also. The higher value in best case and average case is obtained for CMC dataset. The worst values are produced only for a very few number of runs out of the total 100 runs.

TABLE VIII. BEST, WORST AND AVERAGE DB INDEX VALUES

Dataset	Best	Average	Worst
Iris	0.1255	0.1856	0.9625
Wine	0.1223	0.1749	0.6654
Glass	0.1437	0.1638	0.8723
Cancer	0.0150	0.0927	0.5712
CMC	0.2337	0.3603	0.8416
Vowel	0.1585	0.2529	0.9989

C. Analysis of Artificial datasets

The artificial datasets are taken from Speech and image processing unit, School of Computing, University of Eastern Finland (Franti, 2015). Due to the large number of instances, more time is needed to run these datasets. Thus, 20 runs are executed for these datasets and the best, worst and average values for the Dunn index and the DB index are computed.

1) Dunn Index

Dunn Index [37] measures the ratio between the distances and diameter of the clusters. The larger the value, the performance is best.

$$I(C) = \frac{\min_{i\neq j}\{d_c(C_i,C_j)\}}{\max_{1\leq l\leq k}\{\Delta C_l\}} \tag{12}$$

Where k is the number of clusters, $d_c(C_i, C_j)$ measures the distance between the two nearest instances in C_i and C_j, ΔC_l is the diameter of the cluster C that is the distance between the two farthest instances in C.

Table IX shows the results for DB index and Dunn index values for the artificial datasets A1 and A2. A1 has a better result since it has spherically shaped clusters that have clear boundaries. But in A2, a cluster is completely surrounded by another cluster and this exposes an unclear structure which could be only partially found by the proposed method. Also, when the number of clusters and the number of instances increase in the dataset A2, there is a significant decrease in the resulting clustering structure. It is evident that the best value achieved for DB index is 0.3763 for the A2 dataset. However, the average values show better performance in case of the A1 dataset. The Dunn index value is obtained as 0.7466 and showed good performance for the artificial dataset A1.

TABLE IX. DBINDEX AND DUNN INDEX FOR ARTIFICIAL DATASETS

Dataset	Index values	DBIndex	DunnIndex
A1	Best	0.3982	0.7466
	Worst	0.9591	0.3110
	Average	0.4002	0.6983
A2	Best	0.3763	0.4957
	Worst	0.8228	0.2739
	Average	0.4219	0.4691

VII. CONCLUSION

The FCM Crow search algorithm is successfully implemented over six benchmark datasets and two artificial datasets. The proposed method works well for most of the datasets and is found to be efficient when compared with results of ACPSO. Also, the number of parameters for crow search is very minimal leading to reduced complexity. In future, the algorithm can be improved by altering the values for the parameters. The work can further be extended to combine crow search with Intuitionistic fuzzy c-means clustering algorithm.

References

[1] A.K Jain, M.N. Murty, P.J. Flynn, “Data clustering: a review”, ACM computing surveys (CSUR), vol. 31, no. 3, pp.264-323, 1999.
[2] J. C. Bezdek, R. Ehrlich, W. Full, “FCM: The fuzzy c-means clustering algorithm”, Computers & Geosciences, vol. 10, no. (2-3), pp. 191-203, 1984.
[3] E. Rashedi, H. Nezamabadi-Pour, S. Saryazdi, “GSA: a gravitational search algorithm”, Information sciences, vol. 179, no.13, pp.2232-2248, 2009.
[4] M. Dorigo, V. Maniezzo, A. Colorni, “Ant system: optimization by a colony of cooperating agents”, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 26, no. 1, pp. 29-41, 1996.
[5] T. İnkaya, S. Kayalıgil, N.E. Özdemirel, “Ant colony optimization based clustering methodology”, Applied Soft Computing, vol. 28, pp.301-311, 2015.
[6] B. C. Mohan, R. Baskaran, “A survey: Ant Colony Optimization based recent research and implementation on several engineering domains”, Expert Systems with Applications, vol. 39, no.4, pp.4618-4627, 2012.
[7] B. DoğAn, M. Korürek, “A new ECG beat clustering method based on kernelized fuzzy c-means and hybrid ant colony optimization for continuous domains”, Applied Soft Computing, vo1. 2, no. 11, pp.3442-3451, 2012.
[8] Wang, Q., Zhang, Q.P. and W. Zhou, “Study on Remote Sensing Image Segmentation Based on ACA–FCM”, Physics Procedia, vol. 33, pp.1286-1291, 2012.
[9] Kanade, M. Parag M, O.H. Lawrence, "Fuzzy ants and clustering", IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, vol. 37, no. 5, pp. 758-769, 2007.
[10] R. Eberhart, J. Kennedy, “A new optimizer using particle swarm theory”, Micro Machine and Human Science, Proceedings of the Sixth International Symposium, pp. 39-43, 1995.
[11] O. Moh'd Alia, R. Mandava, D. Ramachandran, M. E. Aziz, “Harmony search-based cluster initialization for fuzzy c-means segmentation of MR images”, TENCON 2009 IEEE Region 10 Conference, pp. 1-6, 2009.
[12] A. Hatamlou, “Black hole: A new heuristic optimization approach for data clustering. Information sciences”, vol. 222, pp.175-184, 2013.
[13] D. Karaboga, B. Basturk, “Artificial bee colony (ABC) optimization algorithm for solving constrained optimization problems”, International Fuzzy Systems Association World Congress, pp. 789-798, Springer Berlin Heidelberg, 2007.
[14] M. Krishnamoorthi, A. M. Natarajan, “Artificial bee colony algorithm integrated with fuzzy c-mean operator for data clustering”, Journal of Computer Science, vol. 9, no. 4, pp.404-412, 2013. doi:10.3844/ jcssp.2013.404.412
[15] Z. W. Ye, M. W. Wang, W. Liu, S. B. Chen, “Fuzzy entropy based optimal thresholding using bat algorithm”, Applied Soft Computing, vol. 31, pp.381-395, 2015.
[16] R. Katarya, O.P Verma, “Recommender system with grey wolf optimizer and FCM”, Neural Computing and Applications, pp.1-9, 2016.
[17] A. R. Mehrabian, C. Lucas, “A novel numerical optimization algorithm inspired from weed colonization”, Ecological Informatics, vol. 1, no. 4, pp. 355-366, 2006.
[18] X. Q. Zhao, J.H. Zhou, “Improved kernel possibilistic fuzzy clustering algorithm based on invasive weed optimization”, Journal of Shanghai Jiaotong University (Science), vol. 20, pp.164-170, 2015.
[19] R. Jensi, G.W. Jiji, “An improved krill herd algorithm with global exploration capability for solving numerical function optimization problems and its application to data clustering,”, Applied Soft Computing, vol. 46, pp.230-245, 2016.
[20] Z.Y. Li, J.H. Yi, G.G. Wang, “A New Swarm Intelligence Approach for Clustering Based on Krill Herd with Elitism Strategy”, Algorithms, vol. 8, no. 4, pp.951-964, 2015.
[21] A. Jose-Garcia, W. Gomez-Flores, “Automatic clustering using nature-inspired metaheuristics: A survey”, Applied Soft Computing, vol. 41, pp.192-213, 2016.

[22] M. Delgado, A. G. Skármeta, H.M. Barberá, "A tabu search approach to the fuzzy clustering problem", Fuzzy Systems, Proceedings of the Sixth IEEE International Conference, vol. 1, pp. 125-130, 1997.
[23] S. U. Mane, P.G Gaikwad, "Nature inspired techniques for data clustering", International Conference on Circuits, Systems, Communication and Information Technology Applications, pp.419- 424, 2014.
[24] S. He, N. Belacel, N., Hamam, H, Y Bouslimani, "Fuzzy clustering with improved artificial fish swarm algorithm", Computational Sciences and Optimization, International Joint Conference vol. 2, pp 317-321, 2009.
[25] Srivastava, S. and Sahana, S.K., 2017, "The Insects of Innovative Computational Intelligence", Advances in Computational Intelligence: Proceedings of International Conference on Computational Intelligence, pp. 176-186, 2015, Springer Singapore.
[26] Z. Huang, Y. Zhou, "Using glowworm swarm optimization algorithm for clustering analysis", Journal of Convergence Information Technology, vol. 6, no. 2, pp.78-85, 2011.
[27] S. J. Nanda, G. Panda, "A survey on nature inspired metaheuristic algorithms for partitional clustering", Swarm and Evolutionary Computation, vol. 16, pp.1-18, 2014.
[28] A. Askarzadeh, "A novel metaheuristic method for solving constrained engineering optimization problems: crow search algorithm", Computers & Structures, vol. 169, pp. 1-12, 2016.
[29] A. Asuncion, D. J. Newman, "UCI Repository of Machine Learning Databases. Irvine, University of California", 2007.
[30] P. Franti, "Speech and image processing unit, clustering datasets, School of Computing, University of Eastern Finland", 2015.
[31] C. Li, J. Zhou, P. Kou, J. Xiao, "A novel chaotic particle swarm optimization based fuzzy clustering algorithm", Neurocomputing, vol. 83, pp.98-109, 2012.
[32] N. K. Visalakshi, S. Parvathavarthini, K. Thangavel, "An intuitionistic fuzzy approach to fuzzy clustering of numerical dataset", Computational Intelligence, Cyber Security and Computational Models, pp. 79-87, 2014.
[33] W.M. Rand, "Objective criteria for the evaluation of clustering methods", Journal of the American Statistical Association, vol. 66, no. 336, pp.846-850, 1971.
[34] C.J. Van Rijsbergen, "Information retrieval", Dept. of computer science, university of glasgow. URL: citeseer. ist. psu. edu/vanrijsbergen79information. html, 1979.
[35] L.Hubert,P.Arabie, "Comparing Partitions", Journal of Classification, vol. 2, no. 1, pp.193-218, 1985.
[36] D. L. Davies, D.W. Bouldin, "A cluster separation measure", IEEE transactions on pattern analysis and machine intelligence", vol. 2, pp.224-227, 1979.
[37] J. C. Dunn, "Well-separated clusters and optimal fuzzy partitions", Journal of Cybernetics, vol. *4, no.* 1, pp. 95-104, 1974.