Developments and Trends in Intelligent Technologies and Smart Systems

Vijayan Sugumaran
Oakland University, USA

A volume in the Advances in Computational Intelligence and Robotics (ACIR) Book Series

Published in the United States of America by
IGI Global
Engineering Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA, USA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Names: Sugumaran, Vijayan, 1960- editor.
Title: Developments and trends in intelligent technologies and smart systems / Vijayan Sugumaran, editor.
Description: Hershey, PA : Engineering Science Reference, [2018] | Includes bibliographical references.
Identifiers: LCCN 2017020847| ISBN 9781522536864 (hardcover) | ISBN 9781522536871 (ebook)
Subjects: LCSH: Systems engineering. | Self-organizing systems.
Classification: LCC TA168 .D496 2018 | DDC 620/.0042--dc23 LC record available at https://lccn.loc.gov/2017020847

This book is published in the IGI Global book series Advances in Computational Intelligence and Robotics (ACIR) (ISSN: 2327-0411; eISSN: 2327-042X)

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter 7
Crow-Search-Based Intuitionistic Fuzzy C-Means Clustering Algorithm

Parvathavarthini S.
Kongu Engineering College, India

Karthikeyani Visalakshi N.
NKR Government Arts College for Women, India

Shanthi S.
Kongu Engineering College, India

Lakshmi K.
Kongu Engineering College, India

ABSTRACT

Data clustering is an unsupervised technique that segregates data into multiple groups based on the features of the dataset. Soft clustering techniques allow an object to belong to various clusters with different membership values. However, there are some impediments in deciding whether or not an object belongs to a cluster. To solve these issues, an intuitionistic fuzzy set introduces a new parameter called hesitancy factor that contributes to the lack of domain knowledge. Unfortunately, selecting the initial centroids in a random manner by any clustering algorithm delays the convergence and restrains from getting a global solution to the problem. To come across these barriers, this work presents a novel clustering algorithm that utilizes crow search optimization to select the optimal initial seeds for the Intuitionistic fuzzy clustering algorithm. Experimental analysis is carried out on several benchmark datasets and artificial datasets. The results demonstrate that the proposed method provides optimal results in terms of objective function and error rate.

DOI: 10.4018/978-1-5225-3686-4.ch007

INTRODUCTION

Data Mining pertains to the task of discovering hidden knowledge from a huge volume of data. The role of data mining has become inevitable because of the large volumes of data available in various fields. The world has become a village connected by global data. Due to their voluminous nature, these data cannot be dealt with manually. It is tedious to analyze these data manually and also difficult to identify the patterns associated with them.

Data Mining recognizes the patterns that are available in data with the help of several techniques like Classification, Clustering, Association rule mining, Prediction, etc. Classification is a supervised technique that categorizes data as belonging to which class. Prediction tries to guess the relationship between the variables in data objects and Association rule mining correlates the behavior of data with the outcome of events. Data Mining finds its applications in various fields like Biomedical research, Behavioral and social sciences, Earth sciences, Market Analysis, web search, Decision Support Systems, Buying pattern prediction, etc.

Need for Clustering

Clustering is an exploratory and descriptive data analysis technique that divides objects into several homogeneous groups based on their traits. Due to the increase in large multidimensional datasets, the need for summarizing, analyzing the qualitative and quantitative aspects of data has become unavoidable. Objects with similar features are put into a single cluster. Clustering algorithms should show the same performance irrespective of the number of instances in the dataset. There may be different types of attributes in the dataset. Many real-world problems have several constraints to be satisfied while clustering data. Application areas of clustering include but are not limited to Medical image processing, Pattern recognition, Spatial database technology, Information retrieval, Computer vision, etc.

Types of Clustering Algorithms

Clustering algorithms can be categorized into partitional, hierarchical, density-based and grid-based methods. (Jain, Murty & Flynn, 1999) Partitional algorithms tend to find spherical clusters based on distance measures and generally use mean or medoid to represent cluster center. Hierarchical methods perform multiple levels of decomposition either in top-down or bottom-up fashion which is termed as divisive or agglomerative respectively. They are distance-based or density- and continuity based methods. Density-based algorithms continuously form a cluster until the density in the neighborhood exceeds some threshold and are good in finding arbitrarily shaped clusters. Grid-based methods use a multi-resolution grid structure and are very fast in nature.

Clustering algorithms can be classified into hard and soft based on the allotment of objects. A hard clustering algorithm like K-Means allows an object to be assigned to exactly one cluster. In case of soft clustering methods like Fuzzy C-Means (FCM), (Bezdek, Ehrlich & Full, 1984) an object is allocated to multiple clusters based on the membership value of the object to each of those clusters. The non-membership value is obtained by subtracting the membership value from one.

Fuzzy Set and Intuitionistic Fuzzy Set

Fuzzy sets are designed to manipulate data and information possessing non-statistical uncertainties. A Fuzzy set is represented (Zadeh, 1965) as follows

$$FS = \{< x, \mu_{FS}(x) >| \, x \in X\} \tag{1}$$

where μ_{FS}: X → [0, 1] and ν_{FS}: X → [0, 1] and $\nu_{FS}(x) = 1 - \mu_{FS}(x)$. Here μ_{FS} is the membership value and ν_{FS} is the non-membership value.

An Intuitionistic Fuzzy Set (Atanassov, 2003) can be symbolized as below

$$IFS = \{< x, \mu_{IF}(x), \nu_{IF}(x) >| \, x \in X\} \tag{2}$$

where μ_{IF}: X → [0, 1] and ν_{IF}: X → [0, 1] and $\pi_{IF}(x) = 1 - \mu_{IF}(x) - \nu_{IF}(x)$ such that $0< \mu_{IF}(x) + \nu_{IF}(x)<1$ where π_{IF} is the hesitancy value used to represent the uncertainty.

An IFS is generally a triplet which consists of the membership, non-membership and hesitation degree out of which at least two values should be known in order to calculate the third parameter.

Fuzzy C Means Clustering

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

$$\mu_i(d_j) = \frac{d_{ij} - \min(d_j)}{\max(d_j) - \min(d_j)} \tag{3}$$

where $i = 1, 2, \ldots,$ n and $j = 1, 2, \ldots$ 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 data sets using any distance measure like Euclidean distance.

The belongingness of an object d_i to the cluster c_j is given by

$$U_{ij} = \frac{1}{\sum_{r=1}^{C} \left(\frac{dis(d_j^{'}, v_i)}{dis(d_j^{'}, v_r)} \right)^{\frac{2}{m-1}}}, 1 \le i \le C, 1 \le j \le n, m = 2 \tag{4}$$

The objective function of FCM algorithm can be given as follows

$$J_m(x,y) = \sum_{i=1}^{c}\sum_{j=1}^{p} U_{ij}^{m} \parallel X_j^{'} - C_i \parallel, 1 \leq m \leq \infty \tag{5}$$

The centroids are updated using the following formula

$$V_j = \sum_{i=1}^{n} (u_{ij})^m x_i \Big/ \sum_{i=1}^{n} (u_{ij})^m, \forall j = 1,2,...,C \tag{6}$$

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.

ISSUES IN FCM

FCM is a partitional clustering algorithm which initially puts all the objects in a single group and then data points are relocated between clusters in a flexible manner. At each iteration, the value of criterion function is reduced and when it is stabilized, the algorithm is said to be converged. There are several issues in clustering such as the structure of datasets is unknown, the clusters can be of arbitrary shapes and inability to deal with noisy or missing data. The FCM algorithm minimizes the intra-cluster distance well but it leads to local minimum results only.

Intuitionistic Fuzzy Clustering

Fuzzy clustering deals with uncertainty and fuzziness. Uncertainty arises because there is a hesitation in assigning membership value due to its imprecise nature and also it varies from person to person. To avoid such confusions, Atanassov (2003) introduced another higher order fuzzy set named the Intuitionistic fuzzy set.

Intuitionistic Fuzzy Set (IFS) is a special type of fuzzy set that provides one additional factor called hesitancy degree which means that it is unclear whether the object belongs to or not belongs to a cluster. It is an intermediate state between yes and no and hesitation indicates a state of 'may be'. Intuitionistic fuzzy clustering algorithms are proposed by several authors for clustering images (Ananthi, Balasubramaniam, & Lim, 2014; Bhargava et al., 2013; Chaira, 2011; Huang et al., 2015) and numeric data (Lin, 2014; Xu, & Wu, 2010). The vagueness in data can be well represented using this hesitancy degree.

Need for Optimization

All the clustering algorithms have a general practice of choosing the initial clusters randomly. However, this heavily influences the result of a clustering algorithm. This may lead to getting trapped in local minima. So, the strive for a global solution necessitates the hybridization of the clustering algorithm with some optimization techniques.

There are several swarm-based meta-heuristic algorithms available in literature. The list includes but is not limited to Particle Swarm optimization (Kennedy, Kennedy, Eberhart et al., 2001), Ant Colony Optimization (Dorigo, Maniezzo, & Colorni, 1996), Bee Colony optimization (Karaboga, 2005), Krill herd optimization (Gandomi, & Alavi, 2012), Artificial Fish swarm (Li, & Qian, 2003), Bat optimization (Yang, & Hossein Gandomi, 2012), Cuckoo search Optimization (Yang, & Deb, 2009), Black hole optimization (Hatamlou, 2013), etc. (Jose-Garcia, & Gomez-Flores, 2016). Crow search Optimization is a novel algorithm based on the behavior of crow. This paper combines Intuitionistic fuzzy clustering with crow search optimization so that global optimal solutions can be reached. The crow search algorithm is used to find the best initial seed for the Intuitionistic Fuzzy C-Means (IFCM) Algorithm. The focus of this paper is towards effective clustering of data with a faster convergence in the value of objective function.

Contributions in this Paper

In order to efficiently cluster large and real time datasets, our contributions include

- Developing a novel, hybrid, highly scalable clustering algorithm by combining crow search optimization (one of the recent swarm based techniques) with Intuitionistic Fuzzy C-Means clustering. No such work exists in the literature
- The application of Crow search optimization to clustering has not yet been discussed by any author.
- Combining the best features from Chaira (2011) and Xu (2010) method and thus relieving users from the burden of having domain knowledge and ability to deal with noisy data

This paper is organized as follows: Section 2 focuses on the preliminaries, Section 3 provides an overview of the related literature, Section 4 gives a glimpse of crow search algorithm, Section 5 explains the proposed methodology, Section 6 concentrates on experimental analysis and results, Section 7 gives the concluding remarks.

INTUITIONISTIC FUZZY C-MEANS CLUSTERING

The first and foremost task for IFCM algorithm (Chaira, 2011) is to convert crisp data into fuzzy data which in turn would be converted to Intuitionistic fuzzy data. This process involves the task of fixing the lambda value which is a value that varies for each dataset. Entropy is the amount of fuzziness present in any given dataset. The value of lambda is chosen as the one which maximizes the entropy value.

Yager-generating function can be used to create IFS. The crisp data is converted into fuzzy data using Equation (3). Then the fuzzy data is converted to Intuitionistic fuzzy data as follows:

$$\mu_i(d_j;\lambda) = 1 - (1 - \mu_i(d_j))^{\lambda} \tag{7}$$

$$\nu_i(d_j;\lambda) = 1 - (1 - \mu_i(d_j))^{\lambda(\lambda+1)} \; where \, \lambda \in [0,1] \tag{8}$$

The intuitionistic fuzzification converts the intermediate fuzzy dataset to intuitionistic fuzzy dataset. The hesitancy factor is calculated by summing up the membership and non-membership degrees and subtracting the sum from one.

The clustering procedure given by Xu, & Wu, (2010) is followed. The distance matrix is calculated based on the Intuitionistic fuzzy Euclidean distance. Then, the membership matrix is calculated. This membership value is used to calculate non-member-ship and hesitancy values. Using these values, the mass (weight) factor given to each attribute t is calculated. Using these mass values, the new centroids are calculated. The algorithm proceeds until either the objective function converges or there is no change in the centroids for the consecutive iterations.

The objective function of IFCM can be given as

$$J_m(x,y) = \sum_{i=1}^{c}\sum_{j=1}^{p} U_{ij}^{m} \parallel X_j^{'} - C_i \parallel, 1 \le m \le \infty \tag{9}$$

RELATED WORKS

To overcome the drawbacks of clustering algorithms, lot of researchers have combined them with the optimization techniques. But still the combination of Intuitionistic fuzzy clustering with the optimization algorithms is at its infant stage. There is not much research works published in this area. But fuzzy clustering based optimization has grown to a great extent and several noteworthy references can be found in the literature.

Cuckoo search optimization based fuzzy clustering defines egg laying radius (Amiri & Mahmoudi, 2016) for the eggs being laid by cuckoo and the best habitat is chosen and then fuzzy rules are applied to get optimal solutions that reduce the error rate. Binu (2015) compared the performance of various optimization algorithms like Genetic Algorithm, PSO and Cuckoo search over seven newly designed objective functions. When experimenting with large scale data, PSO-based methods are found to be efficient. Cobos et al., (2014) clustered web document search results by introducing a description-centric algorithm that exploits balanced Bayesian information criterion as the fitness function and thus the number of centroids can be deliberated automatically in advance. In order to retain the merits of both FCM and fuzzy PSO, Izakian & Abraham, (2011) proposed a novel algorithm that found a global solution with reduced execution time. A set of satellite images related to agriculture are segmented using FCM by Parvathavarthini, Visalakshi, & MadhanMohan (2011).

Rajabioun (2011) proposed Cuckoo Optimization Algorithm (COA) with an extension to cuckoo search by adding a parameter called Egg Laying Radius (ELR). This determines the maximum range within which the egg has to be laid. The surviving birds immigrate to a new habit and setup their nests.

The performance of the algorithm is verified by using it against standard benchmark datasets. Cuckoo search algorithm (Yang, & Deb, 2009, 2010) imitates the breeding behavior of the bird cuckoo. The authors utilized Levy flight distribution using Mantegna's algorithm to obtain new solutions and the algorithm is demonstrated with standard and stochastic test functions. Certain percentage of eggs are identified by the host bird and abandoned. The best nest to lay eggs is found and the algorithm proved to be efficient in arriving at an optimal result. A novel hybridization of cuckoo search algorithm with IFCM (Parvathavarthini, Karthikeyani, Shanthi, & Mohan, 2017) is proposed and experiments show that the resulting clusters are efficient.

Kanade and Hall (2007) utilized ACO to cluster the objects and reformulated the cluster centers using FCM and Hard C-Means to determine the number of clusters in each dataset. The intelligent foraging behavior of honey bees is simulated (Karaboga & Ozturk, 2010) and this algorithm is used for clustering. Employee bees collect nectar and share position of food with onlooker bees. Position of food source indicates the solution and the amount of nectar indicates quality of solution. In the black hole algorithm (Hatamlou, 2013), 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 black hole, then their locations are exchanged. The author explains how blackhole optimization can be used for clustering.

Krill herd optimization (Li, Yi, & Wang, 2015) is the idealization of herding of krill swarms in sea. The position of an individual krill is determined by three motions such as: movement induced by other krill individuals, foraging action, and random diffusion. The authors 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.

Jose-Garcia, & Gomez-Flores (2016) 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 paper. Kumutha, & Palaniammal (2014) converted PSO into fuzzy PSO and finally transformed it into Intuitionistic fuzzy PSO and combined this IF-PSO with FCM to yield faster convergence and thus reduce the computational complexity of IFCM algorithm.

A novel method for IF clustering using a multi-objective criterion function is developed by Chaira (2011) to segment CT scan brain images. IFS representation is generated with Yager type IF generator, the objective function is modified and the cluster center updation is incorporated by considering the hesitancy factor also. Shanthi, & Bhaskaran, (2011) utilized this clustering to classify mammogram images and built decision tree for effective diagnosis. Visalakshi, Thangavel, & Parvathi (2010) utilized IFCM algorithm for clustering distributed datasets.

A new clustering algorithm which considered the car data set is built by Xu & Wu (2010) for clustering both IFS and Interval-valued IFS. This proved to be more efficient with numerical datasets. A novel IF approach for Tumor/ hemorrhage detection is proposed in (Chaira, & Anand, 2011) and the images are edge detected by forming an IF divergence matrix, thresholding and thinning it. A robust IFCM and kernel version of IFCM is presented (Kaur, Soni, & Gosain, 2011) with a new distance metric incorporating the distance variation of data-points within each cluster. Krishnamoorthy, Sadasivam, Rajalakshmi, Kowsalyaa, & Dhivya (2017) used PSO to hide sensitive privacy information available in clusters. Bhargava et al. (2013) hybridized rough set with IFS in order to describe a cluster by its centroid and its lower and upper approximations. The method introduces modified Rough FCM with the membership of IFCM. Shanthi, & Bhaskaran. (2013) processed a set of mammogram images to detect and classify breast cancer by finding the region of interest and separating the affected part.

An image is represented as several fuzzy sets with the membership functions for symbolizing the foreground and background and then converted to IFS (Ananthi et al., 2014). The gray scale images are segmented using IFS. The entropy is calculated to find the threshold. The value that minimizes the entropy is taken as the threshold for segmenting the image. An artificial bee colony algorithm is designed (Naser, & Alshattnawi, 2014) for effectively grouping the social networks by collecting people with common interests. Sumathi, Sendhilkumar, & Mahalakshmi (2015) ranked the web pages using weighted page rank algorithm and utilized PSO for clustering the web users. Tripathy, Basu, & Govel (2014) segmented images by defining a spatial function which represents the degree of likeliness a pixel will have to each cluster. This value reaches its maximum for a cluster when most of the neighborhood pixels belong to the same cluster.

An evolutionary kernel IFCM (Lin, 2014) is introduced by maximizing the good points in the kernel space. Genetic Algorithm is used for selecting the parameters involved in this algorithm. A population of chromosomes is initialized, the fitness function is evaluated, roulette wheel selection is applied to choose chromosomes for reproduction, cross over and mutation is performed to get the next generation until the number of epochs are reached. The membership function proposed by Chaira is modified by (Huang et al., 2015) using neighborhood pixel tuning. The membership value is determined using a similarity measurement that represents the difference between the intensity of a pixel and the cluster and has no effective resistance to noise. Balasubramaniam, & Ananthi (2016) segmented nutrition deficiency in incomplete crop images using IFCM. The missing pixels in the incomplete images are imputed using IFCM algorithm. The resulting membership matrix efficiently portrayed the deficiency region of the crop.

CROW SEARCH ALGORITHM

Crows are well-known for their unity and intelligence. They have some special characteristics like self-awareness, recognizing faces and memorizing food sources. Crow Search Algorithm (CrSA) is a new population-based metaheuristic algorithm that simulates the behavior of these intelligent birds in order to solve optimization problems (Askarzadeh, 2016).

Crows live in flocks and they observe 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. To do thievery, a crow always tries to follow another. Crows defend their caches from being pilfered by others. Based on these characteristics, the algorithm 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 $i = 1, 2, \ldots, N$; 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

1. Crow B is not aware of crow A following it and thus crow A reaches the hiding place of crow B.

2. 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.

PROPOSED METHODOLOGY

The IFCM algorithm selects the initial seeds randomly and also it suffers from the problem of falling into local minima. Due to this, there is a delay in the convergence of the clustering algorithm. Instead of choosing random points, this work uses crow search optimization algorithm for selecting the best initial seeds for performing the IFCM clustering. The crow search algorithm uses very few parameters like flight length and awareness probability and thus has reduced complexity. Thus, the implementation of such simple and user-friendly metaheuristic algorithm leads to promising results.

Crow Search-Based Intuitionistic Fuzzy C-Means Algorithm (CrSA_IFCM)

The parameters like population or flock size N, number of clusters C, Maximum number of iterations itmax, flight length fl and awareness probability AP are initialized. The position of the crows denoted by pos is set by generating a random matrix of cluster centers and the data objects are encoded. Here the initial seed values are taken as the crows, the dataset is the search space, each position of the crow is a feasible solution, and the quality of a set of centroids is determined by the objective function. The encoding is done in such a way that the set of initial centroids are taken as the population. The optimal value for initial centroid is obtained as a result of running the crow search optimization algorithm.

The dataset is converted into fuzzy representation using Eq. (3). This in turn is converted into Intuitionistic fuzzy representation using Eq. (5) and (6). The lambda value for this conversion is fixed by a heuristic method and it varies for each dataset. The lambda value which maximizes the entropy is fixed for each dataset. The entropy value is found using the following formula

$$IFE = \frac{1}{N \times M} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \frac{2\mu_i\left(d_j\right)\nu_i\left(d_j\right) + \pi_i^2\left(d_j\right)}{\pi_i^2\left(d_j\right) + \mu_i^2\left(d_j\right) + \nu_i^2\left(d_j\right)} \tag{10}$$

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 are calculated as follows:

$$U_{ij} = \frac{1}{\sum_{r=1}^{C} \left(\frac{dis\left(d_j^{'}, v_i\right)}{dis\left(d_j^{'}, v_r\right)} \right)^{\frac{2}{m-1}}}, 1 \le i \le C, 1 \le j \le n, m = 2 \tag{11}$$

The fitness of initial positions is calculated using the objective function in Eq. (7). Assume that the crow B wants to visit its hiding place, then any crow A is randomly chosen to follow it. There are two possibilities now: either the crow may be aware of its follower or it is unaware. If crow B is conscious that crow A is following, it chooses a random new position to fool crow A. 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

$$x_{A,t+1} = x_{A,t+1} + r_A.fl^{A,t}.\left(m^{B,t} - x^{B,t}\right) \tag{12}$$

The feasibility of new position is then checked and position is updated only if it is feasible. Otherwise no change to the position is made. The fitness of new position of crows is evaluated again. If the quality of the new position is better than the earlier position, the memory of crows is updated using

$$m_{A,t+1} = x_{A,t+1} \tag{13}$$

Similarly, random followers are selected for all the crows and the search is continued until the maximum iteration (itmax) is reached for crow search algorithm. As an outcome of crow search algorithm, the best initial seeds are found that minimize the fitness function to a greater extent. Now, the IFCM algorithm is executed to find the membership function and the fitness value. In order to calculate the membership function, a mass value is assigned to each attribute initially.

Then, the mass values for each attribute are updated during every iteration using the following formula

$$ma_i\left(k+1\right) = \left\{\frac{u_{i1}(k)}{\sum_{j=1}^{n} u_{ij}(k)}, \frac{u_{i2}(k)}{\sum_{j=1}^{n} u_{ij}(k)}, \ldots, \frac{u_{in}(k)}{\sum_{j=1}^{n} u_{ij}(k)}\right\}, 1 \le i \le C \tag{14}$$

Finally, the centroids are updated as follows:

$$V_i = \left\{\left[d_{s,} \sum_{j=1}^{n} ma_j \mu_{Aj}\left(d_s\right), \sum_{j=1}^{n} ma_j \nu_{Aj}(d_s)\right], 1 \le s \le n\right\}, 1 \le i \le C \tag{15}$$

The IFCM algorithm is run till the maximum number of iterations is reached. The cluster index of the objects is found based on the highest membership value obtained. To obtain better results, the CrSA_IFCM algorithm is repeated for 100 runs and the results are found to be better than the existing methods.

Pseudocode for CrSA_IFCM

```
Create intuitionistic fuzzy representation of data
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 NxD dimension search space
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 Eq. (13)
        Calculate the fitness of each crow using Eq. (9)
        Randomly choose one of the crows to follow (for example B)
        If r_B >=AP^{B,t} calculate new position using Eq. (12)
        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 of the crows and
        Update the memory
     if f(x^{A,t+1}) is better than f(m^{i,t}) update memory using Eq. (13)
     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 the best position obtained above
      Calculate mass values using Eq. (14)
      Update cluster centers using Eq. (15)
     end while
end while
```

Benefits of Proposed Algorithm

The prominent benefits yielded from the above methodology are

- Well-separated and compact clusters are obtained
- Global optimal solutions are achieved
- 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
- Efficient in terms of validity indices, accuracy and fitness function

IMPLEMENTATION

The algorithm is developed using MATLAB Programming. There are several algorithm-specific parameters for the IFCM algorithm. The first and foremost being the lambda value which is computed based on the entropy or amount of fuzziness in each dataset. The value that maximizes the entropy is fixed for lambda. The second parameter is the mass (weight) value that is allotted for each attribute of the dataset. This plays a dominant role since the centroids are updated by considering the mass values. Initially, the mass values are equally distributed for all the attributes that is mass of each attribute is set to 1/n where n is the number of attributes. Then they are updated as the IFCM algorithm is executed. The IFCM algorithm is run up to a maximum of 20 iterations. CrSA_IFCM algorithm is repeated for 100 runs.

With respect to the crow search algorithm, the parameters include the number of crows = 20, maximum iterations = 50, and the values for flight length = 2 and awareness probability = 0.1 are taken from (Askarzadeh, 2016). The experiment is repeated for 100 runs and the best, worst and average values are selected for analysis. A salient characteristic of this algorithm is that it is highly adaptable with less number of attributes.

In this section, the algorithm is tested over eleven different datasets which include both real and artificial datasets. Experiments are done in three aspects, first one in terms of error rate, second in terms of cluster validity indices and the other in terms of objective function. The classification accuracy is higher when compared to the other algorithms.

Performance Analysis on Real Datasets

Real datasets are taken from UCI data repository (Asuncion, & Newman, 2007). The algorithm is tested over six types of real data like iris, wine, seed, Contraceptive Method Choice (CMC), glass and vowel. The details of the datasets are given in Table 1. The vowel and glass datasets have large number of clusters and CMC and vowel datasets have large number of instances.

Error Rate

For each dataset, the classification error percentage is calculated. It is the percentage of wrongly classified objects in the test datasets. The error rate is computed by comparing the cluster indices obtained by the CrSA_IFCM algorithm with that of the cluster indices of the benchmark datasets.

The classification error percentage is computed using the following formula

$$ER = \frac{\text{No. of misclassified samples}}{\text{No. of instances in the dataset}} \times 100 \tag{16}$$

Table 1. Benchmark datasets taken for the experiment

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)
Seed	3	8	210(70,70,70)
CMC	3	10	1473 (629,333,511)
Vowel	6	3	871 (72,89,172,151,207,180)
Glass	6	9	214 (70,17,76,13,9,29)

COMPARISON WITH STATE OF THE ART TECHNIQUES

The error rate of the proposed methodology is shown in Table 2 and is compared with the other optimization techniques like PSO, GSA, Blackhole, Cuckoo Optimization and Fuzzy cuckoo optimization algorithm. It is evident from the results that the proposed method shows a significant improvement in the error rate for the datasets iris and glass. In case of CMC and vowel datasets, there is slight decrease in the error rate. For wine and seed datasets, the CrSA_IFCM algorithm tends to have a minor increase in error rate.

Comparison With Xu Method Using the Car Dataset

To quantitatively evaluate the performance of the proposed method, the fitness values are calculated. For each run, the best criterion values are stored and finally the fitness function that produces a best partition by generating the minimum error rate is chosen. If many fitness values generate the same error rate, then the fitness function with the least standard deviation is chosen. Algorithms like PSO need four user-defined parameters such as inertia weight, upper bound for velocity, individual and social learning factors. But crow search achieves more accuracy only with the help of two user-defined parameters. From Table 3, it is known that the CrSA_IFCM produces a better result for the best, average and worst values of the objective function.

Xu, & Wu (2010) proposed the Intuitionistic Fuzzy C-Means Clustering algorithms for IFS and Interval valued-IFS. The experiment is conducted using the car dataset which contains the information of cars sold in the Guangzhou car market in Guangdong, China. Ten instances of data are taken and the result of the clustering algorithm is derived. These cars have six attributes such as Fuel economy, Aerod degree, Price, Comfort, Design and Safety and for the first iteration, a mass value is fixed for these attributes. The mass values for the six attributes are 0.15, 0.10, 0.30, 0.20, 0.15 and 0.10 respectively. The cars are to be classified into three categories. The results of proposed method are compared to Xu method in Table 4 and were found to have a closest match.

The instances R1, R6 which did not match with any specific cluster are grouped in the first cluster. Additionally, R5 which belongs to the second cluster is brought into the first cluster. R8 is moved into cluster 3 and it is found that all the items except the instance R5 are grouped as per Xu method. Thus, the accuracy is 90 percentage when compared to Xu IFCM.

Table 2. Comparison of error rate with COAC and FCOAC

Dataset	PSO	GSA	Blackhole	COAC	FCOAC	CrSA_IFCM
Iris	10.06	10.04	10.02	11.53	9.81	**6.67**
Wine	28.79	29.15	28.47	9.44	**6.18**	6.74
Seed	16.67	14.76	14.29	13.67	**10.13**	10.47
CMC	51.50	57.68	54.39	11.24	10.26	**10.18**
Vowel	42.39	42.26	41.65	14.11	12.13	**12.05**
Glass	41.20	41.39	36.51	35.21	33.35	**31.78**

Table 3. Comparison of objective function with COAC and FCOAC

Dataset	Objective function	COAC	FCOAC	CrSA_IFCM
Iris	Mincost	1.8429	1.7345	1.0670
	Avgcost	1.8869	1.6567	1.1132
	Maxcost	3.0229	2.4341	1.1643
Wine	Mincost	240.3569	238.4567	203.4671
	Avgcost	248.0417	231.0465	207.6163
	Maxcost	550.8276	548.6783	230.1835
Seed	Mincost	5.4613	5.1287	1.4312
	Avgcost	3.3368	3.1262	1.4623
	Maxcost	3.2680	3.0630	1.4763
CMC	Mincost	0.5681	0.4961	0.4159
	Avgcost	0.5787	0.3784	0.5003
	Maxcost	0.7676	0.6826	0.7026
Vowel	Mincost	558.3689	557.3098	529.3124
	Avgcost	582.4493	580.2345	571.3691
	Maxcost	715.8966	713.6643	708.9372
Glass	Mincost	50. 0535	49.1934	49.0137
	Avgcost	53.1587	51.9231	53.6831
	Maxcost	74.2647	68.2312	69.3928

Table 4. Comparison of car dataset result with Xu method

Cluster Id	IFCM Xu Method	CrSA_IFCM
1	R4, R9	R1, R4, R5, R6, R9
2	R5, R10	R10
3	R2, R3, R7	R2, R3, R7, R8
No significant membership of any cluster	R1, R6, R8	–

Also, the objective function value is minimized to a greater extent. The values of Partition Coefficient and Partition Entropy are 0.941 and 0.192 respectively. Also, the number of clusters is varied and in this case, the optimal values are achieved when there are three clusters.

Performance Analysis on Artificial Datasets

Artificial data from University of Eastern Finland site is tested to validate the performance of the algorithm with respect to various size, shapes and overlapping clusters. Four shape datasets given in Figure 1such as flame, jain, pathbased and compound datasets are considered for evaluation. All the data are two-dimensional which means that they contain two attributes. Jain and flame datasets have two clusters while pathbased and compound datasets have three and six clusters respectively. The class label associated with the datasets is provided so that error rate and validity measures are calculated. The number of instances present in the dataset is as follows: flame: 240, jain: 373, pathbased: 300 and compound: 399.

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 incorporate external validity measures, there is a need for apriori knowledge about data (Visalakshi, Parvathavarthini, & Thangavel, 2014). 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 Partition entropy (Halkidi, Batistakis, & Vazirgiannis, 2002) 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 Partition entropy. The performances of all the four datasets have been evaluated using these indices.

For the artificial datasets, the best error rate is achieved for the flame dataset. However, the results of the objective function values are presented as the fuzzy values which show some minute details and the fitness values are minimized as well. It can be seen that the error rate increases as the number of clusters increase. Even though Flame and Jain datasets have same number of clusters, there is substantial increase in the error rate since the number of instances in the Jain dataset is more than that of flame dataset. So, the error rate is directly proportional to the number of clusters and number of instances.

Rand Index

A true positive (TP) decision assigns two similar documents to the same cluster; a true negative (TN) decision assigns two dissimilar documents to different clusters. In general, two types of errors occur frequently. A false positive (FP) decision means that two objects with different features are assigned to the same cluster. A false negative (FN) decision assigns two objects with similar traits to different clusters. The Rand index (Rand, 1971) measures the percentage of decisions that are correct.

$$RI = \frac{TP + TN}{TP + FP + FN + TN} \tag{17}$$

Figure 1. Two dimensional artificial datasets with different shapes of clusters

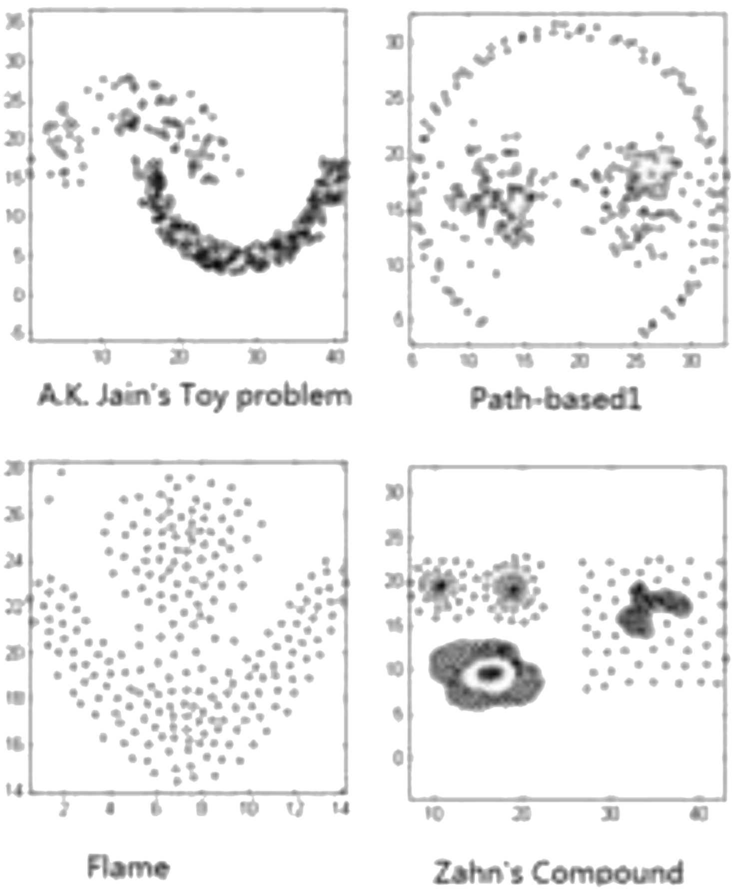

Table 5. Error rate for the artificial datasets

Dataset	Error rate	Objective Function Values			
		Best	Worst	Average	Std
Flame	12.9167	3.79	4.05	4.01	0.07
Jain	24.3968	10.45	11.07	10.87	0.12
Pathbased	25.3333	8.5628	8.9110	8.8505	0.1010
Compound	30.6015	2.2628	2.6333	2.5281	0.1142

F-Measure

The F-Measure (Van Rijsbergen, 1979) 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

$$F = \frac{2TP}{2TP + FP + TN} \tag{18}$$

Adjusted Rand Index

The peculiarity of Adjusted Rand Index (Hubert & Arabie, 1985) 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. 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.

$$ARI = \sum_{i=1}^{C}\sum_{j=1}^{D}\binom{n_{ij}}{2}\binom{n}{2}^{-1}\sum_{i=1}^{C}\binom{n_i}{2}\sum_{j=1}^{D}\binom{n_j}{2}\Bigg/\frac{1}{2}\left[\sum_{i=1}^{C}\binom{n_i}{2}+\sum_{j=1}^{D}\binom{n_j}{2}\right]-\binom{n}{2}^{-1}\sum_{i=1}^{C}\binom{n_i}{2}\sum_{j=1}^{D}\binom{n_j}{2} \tag{19}$$

Partition Entropy

Partition entropy (Dumitrescu, 1993) is an internal measure that involves only the membership values. The value ranges between 0 to number of clusters. The closer the value of PE to 0, the harder the clustering is. When the cluster structure is properly identified, partition entropy reaches its minimum value. It is calculated as follows

$$PE = -\frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}\mu_{ij}\log_a\left(\mu_{ij}\right) \tag{20}$$

Where N is the number of instances in the dataset and k is the number of clusters

Tables 6 and 7 present the index values that play a prominent role in measuring the clustering accuracy. The highest value is for ARI is achieved by flame dataset and the least value is for Jain dataset. In case of Rand index, compound dataset scores the maximum.

The experiments in Table 8 show that the Iris dataset produces the significant value for best rand index value. The overall minimum value is obtained for the glass dataset which consists of spherical clusters. The least optimal value for partition entropy is achieved by the CMC dataset.

Table 6. Best, Mean and standard deviation values for Adjusted Rand Index and Rand Index

Dataset	ARI	ARI Mean	ARI Std	Rand	Rand Mean	Rand Std
Flame	0.5482	0.4651	0.0268	0.7741	0.7369	0.0134
Jain	0.2607	0.2317	0.0003	0.6301	0.6231	0.0004
Pathbased	0.3696	0.3613	0.0053	0.7031	0.7000	0.0021
Compound	0.5412	0.5380	0.0370	0.8342	0.8323	0.0113

Table 7. Best, Mean and standard deviation values for Partition entropy and F-Measure

Dataset	Partition entropy	PE Mean	PE std	F-Measure	FM Mean	FM std
Jain	0.4778	0.4376	0.0005	0.8035	0.7709	0.0008
Flame	0.2866	0.2671	0.0073	0.8733	0.8478	0.0098
Pathbased	0.5881	0.5211	0.0136	0.6378	0.6317	0.0060
Compound	0.5250	0.5212	0.0085	0.7099	0.6816	0.0203

Table 8. Validation Indices for Real datasets

Dataset	ARI	Rand Index	Partition entropy	F-Measure
Iris	0.8176	0.9195	0.4207	0.9333
Wine	0.8253	0.8042	0.3865	0.8093
Seed	0.8607	0.8576	0.2672	0.8729
CMC	0.7890	0.7947	0.1986	0.7543
Vowel	0.7234	0.8152	0.3150	0.7211
Glass	0.5246	0.6730	0.5234	0.5907

CONCLUSION

Optimization techniques have become the need of the hour because of their ability to explore and exploit the problem space in order to achieve a near optimal solution. Traditional algorithms are suffering from local minima solutions which necessitate a fast move towards hybridization. To mine useful information from data and grouping them into clusters with a weightage for the hesitation makes sense in the current scenario. This work provides an innovative clustering algorithm by combining Intuitionistic Fuzzy Clustering with crow search optimization. Investigations on data show that the proposed method combines the benefits of IFCM and Optimization techniques. Since crow search being a non-greedy algorithm, the range and variety of the solutions that are generated is increased.

REFERENCES

Amiri, E., & Mahmoudi, S. (2016). Efficient protocol for data clustering by fuzzy Cuckoo Optimization Algorithm. *Applied Soft Computing*, *41*, 15–21. doi:10.1016/j.asoc.2015.12.008

Ananthi, V. P., Balasubramaniam, P., & Lim, C. P. (2014). Segmentation of gray scale image based on intuitionistic fuzzy sets constructed from several membership functions. *Pattern Recognition*, *47*(12), 3870–3880. doi:10.1016/j.patcog.2014.07.003

Askarzadeh, A. (2016). A novel metaheuristic method for solving constrained engineering optimization problems: Crow search algorithm. *Computers & Structures*, *169*, 1–12. doi:10.1016/j.compstruc.2016.03.001

Asuncion, A., & Newman, D. (2007). UCI machine learning repository.

Atanassov, K. T. (2003, September). Intuitionistic fuzzy sets: past, present and future. In *Proceedings of the EUSFLAT Conf.* (pp. 12-19).

Balasubramaniam, P., & Ananthi, V. P. (2016). Segmentation of nutrient deficiency in incomplete crop images using intuitionistic fuzzy C-means clustering algorithm. *Nonlinear Dynamics*, *83*(1-2), 849–866. doi:10.1007/s11071-015-2372-y

Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, *10*(2-3), 191–203. doi:10.1016/0098-3004(84)90020-7

Bhargava, R., Tripathy, B. K., Tripathy, A., Dhull, R., Verma, E., & Swarnalatha, P. (2013, August). Rough intuitionistic fuzzy c-means algorithm and a comparative analysis. In *Proceedings of the 6th ACM India Computing Convention* (p. 23). ACM. doi:10.1145/2522548.2523140

Binu, D. (2015). Cluster analysis using optimization algorithms with newly designed objective functions. *Expert Systems with Applications*, *42*(14), 5848–5859. doi:10.1016/j.eswa.2015.03.031

Chaira, T. (2011). A novel intuitionistic fuzzy C means clustering algorithm and its application to medical images. *Applied Soft Computing*, *11*(2), 1711–1717. doi:10.1016/j.asoc.2010.05.005

Chaira, T., & Anand, S. (2011). A novel intuitionistic fuzzy approach for tumour/hemorrhage detection in medical images.

Cobos, C., Muñoz-Collazos, H., Urbano-Muñoz, R., Mendoza, M., León, E., & Herrera-Viedma, E. (2014). Clustering of web search results based on the cuckoo search algorithm and balanced Bayesian information criterion. *Information Sciences*, *281*, 248–264. doi:10.1016/j.ins.2014.05.047

Dorigo, M., Maniezzo, V., & Colorni, A. (1996). Ant system: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics*, *26*(1), 29–41. doi:10.1109/3477.484436 PMID:18263004

Dumitrescu, D. (1993). Fuzzy measures and the entropy of fuzzy partitions. *Journal of Mathematical Analysis and Applications*, *176*(2), 359–373. doi:10.1006/jmaa.1993.1220

Gandomi, A. H., & Alavi, A. H. (2012). Krill herd: A new bio-inspired optimization algorithm. *Communications in Nonlinear Science and Numerical Simulation*, *17*(12), 4831–4845. doi:10.1016/j.cnsns.2012.05.010

Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2002). Cluster validity methods: Part I. *SIGMOD Record*, *31*(2), 40–45. doi:10.1145/565117.565124

Hatamlou, A. (2013). Black hole: A new heuristic optimization approach for data clustering. *Information Sciences*, *222*, 175–184. doi:10.1016/j.ins.2012.08.023

Huang, C. W., Lin, K. P., Wu, M. C., Hung, K. C., Liu, G. S., & Jen, C. H. (2015). Intuitionistic fuzzy c-means clustering algorithm with neighborhood attraction in segmenting medical image. *Soft Computing*, *19*(2), 459–470. doi:10.1007/s00500-014-1264-2

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of classification, 2*(1), 193-218.

Izakian, H., & Abraham, A. (2011). Fuzzy C-means and fuzzy swarm for fuzzy clustering problem. *Expert Systems with Applications*, *38*(3), 1835–1838. doi:10.1016/j.eswa.2010.07.112

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. [CSUR]. *ACM Computing Surveys*, *31*(3), 264–323. doi:10.1145/331499.331504

Jose-Garcia, A., & Gomez-Flores, W. (2016). Automatic clustering using nature-inspired metaheuristics: A survey. *Applied Soft Computing*, *41*, 192–213. doi:10.1016/j.asoc.2015.12.001

Kanade, P. M., & Hall, L. O. (2007). Fuzzy ants and clustering. *IEEE Transactions on Systems, Man, and Cybernetics. Part A, Systems and Humans*, *37*(5), 758–769. doi:10.1109/TSMCA.2007.902655

Karaboga, D. (2005). *An idea based on honey bee swarm for numerical optimization* (Technical report-tr06). Erciyes University, Engineering Faculty, Computer Engineering Department.

Karaboga, D., & Ozturk, C. (2010). Fuzzy clustering with artificial bee colony algorithm. *Scientific Research and Essays*, *5*(14), 1899–1902.

Kaur, P., Soni, A. K., & Gosain, A. (2011, November). Robust Intuitionistic Fuzzy C-means clustering for linearly and nonlinearly separable data. In *Proceedings of the 2011 International Conference on Image Information Processing (ICIIP)* (pp. 1-6). IEEE. doi:10.1109/ICIIP.2011.6108908

Kennedy, J. F., Kennedy, J., Eberhart, R. C., & Shi, Y. (2001). *Swarm intelligence*. Morgan Kaufmann.

Krishnamoorthy, S., Sadasivam, G. S., Rajalakshmi, M., Kowsalyaa, K., & Dhivya, M. (2017). Privacy Preserving Fuzzy Association Rule Mining in Data Clusters Using Particle Swarm Optimization. *International Journal of Intelligent Information Technologies*, *13*(2), 1–20. doi:10.4018/IJIIT.2017040101

Kumutha, V., & Palaniammal, S. (2014). Improved Fuzzy Clustering Method Based On Intuitionistic Fuzzy Particle Swarm Optimization. *Journal of Theoretical & Applied Information Technology*, *62*(1).

Li, X. L., & Qian, J. X. (2003). Studies on artificial fish swarm optimization algorithm based on decomposition and coordination techniques. *Journal of Circuits and Systems*, *1*, 1–6.

Li, Z. Y., Yi, J. H., & Wang, G. G. (2015). A New Swarm Intelligence Approach for Clustering Based on Krill Herd with Elitism Strategy. *Algorithms*, *8*(4), 951–964. doi:10.3390/a8040951

Lin, K. P. (2014). A novel evolutionary kernel intuitionistic fuzzy C-means clustering algorithm. *IEEE Transactions on Fuzzy Systems*, *22*(5), 1074–1087. doi:10.1109/TFUZZ.2013.2280141

Naser, A. M. A., & Alshattnawi, S. (2014). An Artificial Bee Colony (ABC) Algorithm for Efficient Partitioning of Social Networks. *International Journal of Intelligent Information Technologies*, *10*(4), 24–39. doi:10.4018/ijiit.2014100102

Parvathavarthini, S., Karthikeyani, N., Shanthi, S., & Mohan, J. M. (2017). Cuckoo-search based Intuitionistic Fuzzy Clustering Algorithm. *Asian Journal of Research in Social Sciences and Humanities*, *7*(2), 289–299. doi:10.5958/2249-7315.2017.00091.0

Parvathavarthini, S., & Visalakshi, N. K. & MadhanMohan J, Identification of optimal clusters by Segmenting Satellite Images. In *Proceedings of ICNICT '11*.

Rajabioun, R. (2011). Cuckoo optimization algorithm. *Applied Soft Computing*, *11*(8), 5508–5518. doi:10.1016/j.asoc.2011.05.008

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, *66*(336), 846–850. doi:10.1080/01621459.1971.10482356

Shanthi, S., & Bhaskaran, V. M. (2011). Intuitionistic fuzzy C-means and decision tree approach for breast cancer detection and classification. *European Journal of Scientific Research*, *66*(3), 345–351.

Shanthi, S., & Bhaskaran, V. M. (2013). A novel approach for detecting and classifying breast cancer in mammogram images. *International Journal of Intelligent Information Technologies*, *9*(1), 21–39. doi:10.4018/jiit.2013010102

Sumathi, G., Sendhilkumar, S., & Mahalakshmi, G. S. (2015). Ranking Pages of Clustered Users using Weighted Page Rank Algorithm with User Access Period. *International Journal of Intelligent Information Technologies*, *11*(4), 16–36. doi:10.4018/IJIIT.2015100102

Tripathy, B. K., Basu, A., & Govel, S. (2014), December. Image segmentation using spatial intuitionistic fuzzy C means clustering. In *Proceedings of the 2014 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)* (pp. 1-5). IEEE.

Van Rijsbergen, C. J. (1979). *Information retrieval*. Dept. Of Computer Science, University Of Glasgow.

Visalakshi, N. K., Parvathavarthini, S., & Thangavel, K. (2014). An intuitionistic fuzzy approach to fuzzy clustering of numerical dataset. In Computational Intelligence, Cyber Security and Computational Models (pp. 79-87). Springer India.

Visalakshi, N. K., Thangavel, K., & Parvathi, R. (2010). An Intuitionistic fuzzy approach to distributed fuzzy clustering. *International Journal of Computer Theory and Engineering*, *2*(2), 295–302. doi:10.7763/IJCTE.2010.V2.155

Xu, Z., & Wu, J. (2010). Intuitionistic fuzzy C-means clustering algorithms. *Journal of Systems Engineering and Electronics*, *21*(4), 580–590. doi:10.3969/j.issn.1004-4132.2010.04.009

Yang, X. S., & Deb, S. (2009, December). Cuckoo search via Lévy flights. In *Proceedings of the World Congress on Nature & Biologically Inspired Computing NaBIC '09* (pp. 210-214). IEEE.

Yang, X. S., & Deb, S. (2010). Engineering optimisation by cuckoo search. *International Journal of Mathematical Modelling and Numerical Optimisation*, *1*(4), 330–343. doi:10.1504/IJMMNO.2010.035430

Yang, X. S., & Hossein Gandomi, A. (2012). Bat algorithm: A novel approach for global engineering optimization. *Engineering Computations*, *29*(5), 464–483. doi:10.1108/02644401211235834

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353. doi:10.1016/S0019-9958(65)90241-X

KEY TERMS AND DEFINITIONS

Clustering: It is an unsupervised learning technique that groups the data objects without apriori knowledge of class labels.

Data Mining: It is one of the steps in Knowledge Discovery in Databases (KDD). It discovers interesting patterns or knowledge from huge volume of data.

Fuzzy Logic: Fuzzy logic can assign many real values between 0 and 1. There may be partial truth or different degrees of membership for a statement.

Optimization: It aims at finding the best among the feasible solutions to a problem based on either minimizing or maximizing the fitness function.