Granular Computing
https://doi.org/10.1007/s41066-021-00259-1

ORIGINAL PAPER

Generalized intuitionistic fuzzy c-means clustering algorithm using an adaptive intuitionistic fuzzification technique

Meenakshi Kaushal[1] · Q. M. Danish Lohani[1]

Received: 18 December 2020 / Accepted: 24 February 2021

Abstract

In real-world scenario, mostly, the datasets are either imprecise or uncertain in their original form. Due to this reason, the clustering of such datasets is unsatisfactory and we often get compromised results. The information present in the dataset can be made precise and useful with the popular generalization of fuzzy sets known as Atanassov intuitionistic fuzzy sets (AIFSs). Therefore, the AIFS-based *c*-means clustering algorithm becomes a convenient way for clustering uncertain and vague datasets. In the paper, we propose an intuitionistic fuzzy-based algorithm, namely Generalized Intuitionistic Fuzzy *c*-Means (G-IFCM) clustering algorithm which uses an adaptive AIFS Euclidean distance measure in its criterion function to cluster the dataset under intuitionistic fuzzy environment. The proposed intuitionistic fuzzification method incorporates a technique to transform the dataset into AIFS and maintains its original structure that tends to change during any fuzzification process. Further, simulation experiments are also conducted on few UCI machine learning repository datasets using G-IFCM and its performance is compared with some known clustering algorithms. The efficacy of the algorithm is tested with some popular benchmark indexes that check the cluster validity and clustering performance of G-IFCM. Finally, the computation suggests efficient performance of G-IFCM where the best choice of fuzzy factor, m probably lies in the interval [1, 2] that indicates reduction in the expense of computation.

Keywords Intuitionistic fuzzy set · Fuzzy clustering · IFCM clustering · Distance measure

1 Introduction

The solution of a real-world problem is obtained by constructing a model on the basis of its information. The relevant information gathered is then encoded in form of a dataset. There are cases when the ambiguity present in a dataset makes them very much imprecise and vague in nature such as linguistic variables. The fuzzy set theory proposed by Zadeh (1965) gave a new direction to researchers to formalize imprecise datasets like linguistic concepts discussed in Zadeh (1996) and Hellmann (2001) with wide range of applications in various fields by Chen and Huang (2003), Anderson et al. (2009), Chen (1996), Chen and Jong (1997), Castro et al. (2007), Chen et al. (1990), and Logambigai and Kannan (2016). A fuzzy set is fundamentally defined by a membership function whose key role is to grade the whole dataset with values lying between zero and one. In literature, there are various types of membership functions such as piecewise linear functions (triangular membership function proposed by Pedrycz (1994), trapezoidal membership function by Barua et al. (2014)) and non-linear functions such as gaussian membership function by Wu (2012), sigmoidal membership function by Mandal et al. (2012) and polynomial-based functions introduced by Narimani and Lam (2010). These membership functions are mainly used to cope up with uncertainty present in the linguistic terms of input and the output of fuzzy logic system (FLS). The study of fuzzy membership functions has explored FLS with applications like stability analysis of system control discussed by Li et al. (2020), by Li and Mehran (2019), edge detection by Singh et al. (2019), image recognition by Priya (2019) and neural networks has been done by Bede (2019).

✉ Meenakshi Kaushal
meenakshikaushal247@gmail.com

Q. M. Danish Lohani
danishlohani@cs.sau.ac.in

[1] Department of Mathematics, South Asian University, New Delhi 110021, India

Clustering is also one of the widely used unsupervised techniques in applications domain of fuzzy theory. It organizes the data into clusters such that most similar data-items are closest to each other and the dissimilar data-items are far from each other. In any fuzzy clustering algorithm, the selection of an optimal membership function plays a crucial role in cluster formation of imprecise data and is discussed by Aziz (2009). Fuzzy *c*-means (popularly known as FCM) proposed by Bezdek et al. (1984) is a widely used fuzzy clustering technique. Several extensions of FCM have been introduced such as intuitionistic fuzzy *c*-means (IFCM) by Xu and Wu (2010), interval type-2 fuzzy *c*-means by Qiu et al. (2013), rough fuzzy *c*-means introduced by Ji et al. (2012), hesitant fuzzy *k*-means proposed by Chen et al. (2014) and many more. These variants of FCM possess their own membership functions based on the criterion function which affects the performance of clustering accordingly. Their significant role can be seen in many real-world applications such as data mining by Kunwar et al. (2019), pattern recognition given by Kumar et al. (2019b) and image segmentation given by Chakraborty et al. (2020).

The Atanassov's intuitionistic fuzzy set (AIFS) introduced by Atanassov (1986) is one of the popular generalization of fuzzy sets with various useful applications by Chen and Chang (2016), Zeng et al. (2019) and Chen et al. (2016). The three main components that defines an AIFS are membership function, non-membership function and the hesitancy function. Besides a traditional fuzzy set, an AIFS can efficiently model the ambiguity and uncertainty of a real-world phenomenon by the exploitation of its hesitant factor. Due to this reason, it became a popular choice for clustering ambiguous datasets. Several researchers diversified the intuitionistic fuzzy set-based clustering algorithms for real world applications (Iakovidis et al. 2008; Chaira 2011; Huang et al. 2015; Verma et al. 2016; Zhou et al. 2016; Lin 2013; Zhao et al. 2019). An intuitionistic fuzzy clustering uses a novel similarity measure based on membership values and non-membership values in FCM to cluster intuitionistic fuzzy image representations (see Iakovidis et al. 2008). In the similar fashion, Chaira (2011) proposed another version of IFCM to perform medical segmentation by adding the hesitation degree of AIFS in the membership matrix. An entropy-based objective function was also introduced to enhance the performance of its membership function which improved overall clustering results (see Chaira 2011). As an extension of the IFS algorithm by Chaira (2011), another entropy-based objective function was proposed by Huang et al. (2015) to cope issues of noise/outliers using neighborhood pixel tuning in improving the clustering results of different regions of magnetic resonance imaging. Some recent IFCM algorithms have been discussed by Verma et al. (2016), Zhou et al. (2016) and Lin (2013). Besides AIFS, type-2 fuzzy set has been used for clustering uncertain vague data by Zhao et al. (2019), Nguyen et al. (2013), and Melin and Castillo (2014)). The combination of interval type-2 fuzzy sets with IFCM in handling noisy data have improved the results of color image segmentation by Zhao et al. (2019).

Mostly, the datasets in real-world applications are either imprecise or uncertain in nature. This ambiguity in dataset leads to poor clustering results. Atanassov intuitionistic fuzzy set-based approaches is one way to efficiently handle such uncertain situation. The information present in the dataset can be made precise and useful with the help of AIFSs. The given dataset is represented in the form of AIFS using a data fuzzification technique. However, there are chances where a data fuzzification technique can distort the original structure of a dataset. Generally, in any AIFS-based clustering technique, the normalization technique is used to generate membership value of the dataset that simply scales down the dataset to values in [0, 1] without changing or deforming the dataset. Nevertheless, there still does not exist any method which is adaptive in nature and can generate AIFS dataset by maintaining its original information. Therefore, in the paper, we have tried to work on an adaptive data intuitionistic fuzzification technique that is a requisite for better performance of any real-world dataset clustering under uncertainty.

Therefore, the paper gives the following contributions:

1. It provides a novel adaptive intuitionistic data fuzzification techniques which generate the dataset in the form of Atanassov intuitionistic fuzzy sets (AIFSs).
2. A new AIFS-based clustering algorithm called generalized intuitionistic fuzzy *c*-means (G-IFCM) clustering algorithm has been proposed which uses an adaptive AIFS-based Euclidean distance to cluster ambiguous datasets.
3. The optimization of proposed clustering algorithm is obtained under tuning of few parameters, namely fuzzy factor m, Intuitionistic fuzzy parameter α, and a tuning parameter β used in its criterion function and proposed fuzzification techniques.
4. Lastly, the analysis of G-IFCM clustering algorithm is done with study of obtained clusters using some benchmark measuring indexes such as clustering accuracy (CA), partition coefficient (PC), partition index (SC), Dunn index (DI), and Xie–Beni index (XB). The results are compared with some known clustering algorithms.

The remaining paper is organized in four sections. In Sect. 2, the basic preliminaries related to the work are presented. Section 3 proposes Generalized Intuitionistic fuzzy *c*-Means (G-IFCM) clustering algorithm using intuitionistic data fuzzification technique. The experimental results on UCI machine learning datasets are presented in Sect. 4. Finally, the conclusion is stated in Sect. 5.

2 Related preliminaries

In this section, we give the definition of intuitionistic fuzzy sets and some related mathematical terms. The notations used in the rest of the paper are descripted in Table 1:

Definition 2.1 (Atanassov 1986) *Atanassov Intuitionistic Fuzzy Set* (*AIFS*): An Atanassov intuitionistic fuzzy set $\hat{A}$ in the universe of discourse $X = \{x_1, x_2, \ldots, x_P\}$ is of the form

$$\hat{A} = \{\langle x, \mu_{\hat{A}}(x), v_{\hat{A}}(x)\rangle | x \in X\}. \tag{1}$$

The membership value $\mu(x)$ is assigned by the mapping $\mu_{\hat{A}} : X \rightarrow [0,1]$ and non-membership value $v(x)$ is assigned by the mapping $v_{\hat{A}} : X \rightarrow [0,1]$ to each element $x \in X$ with respect to $\hat{A}$, such that

$$0 \leq \mu_{\hat{A}}(x) + v_{\hat{A}}(x) \leq 1. \tag{2}$$

Atanassov has introduced the intuitionistic index Atanassov (1986) called hesitancy defined as:

$$\pi_{\hat{A}}(x) = 1 - \mu_{\hat{A}}(x) - v_{\hat{A}}(x). \tag{3}$$

If $\pi_{\hat{A}}(x) = 0$, then $v_{\hat{A}}(x) = 1 - \mu_{\hat{A}}(x)$. Therefore, $\hat{A}$ reduces to a fuzzy set.

Definition 2.2 (Bustince et al. 2000; Chaira 2011) *Intuitionistic fuzzy generator*: An intuitionistic fuzzy generator can be defined as a function $F : [0,1] \rightarrow [0,1]$ satisfying the characteristics below:

1. $F(x) \leq 1 - x$ for all $x \in [0,1]$,
2. $F(0) = 1$ and $F(1) = 0$.

If the function F depicts continuity and behaves as decreasing (increasing), then intuitionistic fuzzy generator, F is called continuous, decreasing (increasing). The Yager's generating function $(I(\mu(x)))$ introduced by Yager (1980) is used to compute the complement of intutionistic fuzzy generator. The fuzzy complement function is given as:

$$I(\mu(x)) = J^{-1}(J(1) - J(\mu(x))), \tag{4}$$

where $J(.)$ is an increasing function and $J : [0,1] \rightarrow [0,1]$.

If we suppose the membership value in Yager class as function $\mu : [0,1] \rightarrow [0,1]$, where $\mu(x) = x$ and $J : [0,1] \rightarrow [0,1]$ such that $J(x) = x^{\alpha}$. The non-membership value can be deduced with Yager's intuitionistic fuzzy complement given as:

$$K(x) = (1 - x^{\alpha})^{\frac{1}{\alpha}}, \alpha > 0 \text{ where } K(1) = 0, K(0) = 1. \tag{5}$$

So, the AIFS A can be obtained as:

$$A = \left\{ < x, \mu_A(x), (1 - \mu_A(x)^{\alpha})^{\frac{1}{\alpha}} > | x \in X \right\}. \tag{6}$$

The hesitancy, $\pi_A(x)$ is computed as:

$$\pi_A(x) = 1 - \mu_A(x) - (1 - \mu_A(x)^{\alpha})^{\frac{1}{\alpha}}. \tag{7}$$

Table 1 Mathematical symbols

Symbols	Details
α	Tuning parameter in Yager'complement function
β	Tuning parameter in membership function
D	Number of attributes
m	Fuzziness index
c	Number of clusters
P	Number of data points
J_m	Objective function
u_{li}	Membership degree of ith data point in lth cluster
$\tilde{D}$	Distance function
x_i	ith data point
s_l	Centroid of lth cluster

3 Proposed generalized intuitionistic fuzzy *c*-means algorithm (G-IFCM)

On the basis of motivation of the paper discussed in Sect. 1, a new clustering algorithm, named as Generalized Intuitionistic fuzzy *c*-Means algorithm (G-IFCM) is proposed using an adaptive technique of intuitionistic fuzzification of any real-world dataset in AIFS-based distance measure. Before proposing the clustering algorithm, we first introduce the techniques of intuitionistic fuzzification of the datasets. Further, a procedure to implement the G-IFCM algorithm is also presented in the section step by step with a flowchart given in Fig. 1.

Let us take $X = \{x_1, x_2, \ldots, x_P\}$ as a set of data items, where each data-item x_i has D dimensions. The intuitionistic version of X as $\tilde{X}$ can be obtained with the proposed methods as:

The given dataset X is first scaled to the range of interval [0, 1] using the normalization equation given below:

$$N_d = a + \frac{x_i - (x_{\min})^d}{(x_{\max})^d - (x_{\min})^d}(b - a), \tag{8}$$

where $(x_{\min})^d$ is the minimum value of the dth dimension of the dataset X, and $(x_{\max})^d$ is the maximum value of the

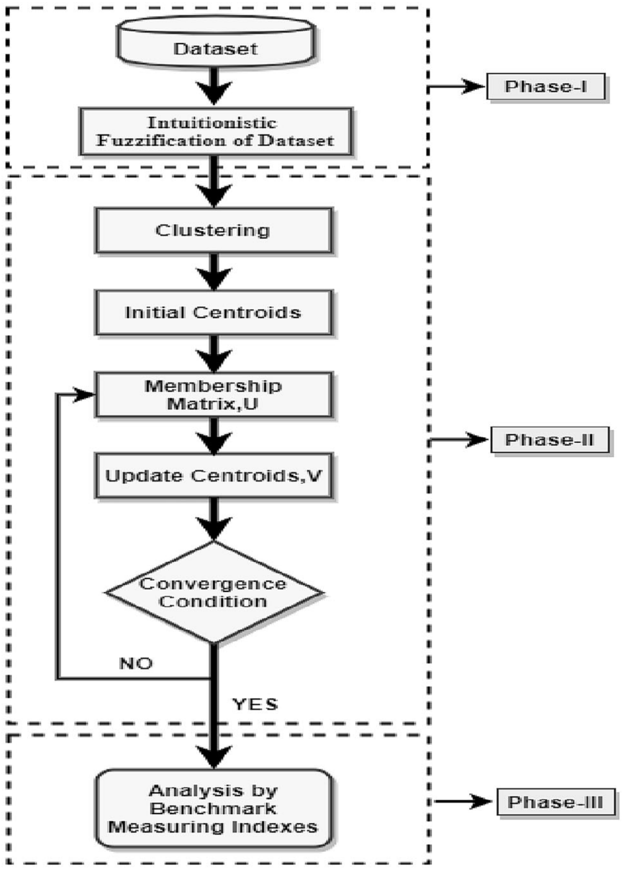


Fig. 1 Proposed architecture of G-IFCM

dth dimension of the dataset X; a and b are the parameters used for the new dataset values being mapped. Here, $a = 0$ and $b = 1$. The matrix $N = \{N_d | N_d = (n_{jd})_{j=1}^{P}, 1 \leq d \leq D\}$ is the obtained normalization matrix. This is the first step of the intuitionistic fuzzification. Once the dataset is normalized to [0, 1], distance of each data-item x_i is deduced with data-item x_j. As each data-item is considered as a row vector, we use `norm` function. It assigns a scalar value to each pair (x_i, x_j) where $1 \leq i, j \leq P$. The distance matrix obtained for the normalized dataset is termed as $dis = \{dis_{ij} = norm(x_i, x_j)\}$. Both the matrices, that is, the normalized dataset N and the distance matrix dis are multiplied together to obtain a new matrix, defined as:

$$Prod = \{p_{id} | p_{id} = dis_{ij} * N_{jd}, 1 \leq i, j \leq P, 1 \leq d \leq D\}. \tag{9}$$

Further the reciprocation of the matrix, $Prod$ is computed to scale down each p_{id} to scalar value in (0, 1). It is termed as new matrix, $M_{id} = \{\mu_{id} = \frac{1}{p_{id}}; 1 \leq i \leq P, 1 \leq d \leq D\}$. The reciprocation of $Prod$ matrix is given below:

$$M = \frac{1}{Prod}. \tag{10}$$

The matrix M gives the membership value matrix to each data-item x_i. For the allocation of non-membership value v_{id}, intuitionistic fuzzy generator is used. And further, the Yager's intuitionistic fuzzy complement calculates the non-membership value which is given as:

$$v_{id} = (1 - \mu_{id}^{\alpha})^{\frac{1}{\alpha}}. \tag{11}$$

Therefore the non-membership value matrix N is obtained such $N = \{v_{id}, 1 \leq i \leq P, 1 \leq d \leq D\}$. The hesitancy value is deduced using the formula given as:

$$\pi_{id} = 1 - \mu_{id} - (1 - \mu_{id}^{\alpha})^{\frac{1}{\alpha}}. \tag{12}$$

Similarly, the hesitancy value matrix is defined as $H = \{\pi_{id},\ 1 \leq i \leq P, 1 \leq d \leq D\}$. Hence, the AIFS data are represented as $\hat{x}_i = (\mu_i, v_i, \pi_i)$, where each dimension is itself a s AIFS. The procedure to compute AIFS dataset, $\tilde{X}$ from the given dataset, X is given in Algorithm I in a simplied form.

Example 1: In this example, we use a 5×5 dataset with positive values taken from interval [1, 10] in Eq. (13). Each row depicts a student grading regarding 5 subjects on the marking scale from 1 to 10. We consider each student as a data-item with 5 dimensions. To transform the dataset to AIFS dataset, we follow the steps of Algorithm I as below:

$$A = \begin{bmatrix} 6 & 4 & 1 & 10 & 9 \\ 7 & 2 & 4 & 6 & 8 \\ 6 & 7 & 10 & 9 & 2 \\ 7 & 3 & 8 & 1 & 4 \\ 7 & 4 & 8 & 5 & 10 \end{bmatrix} \tag{13}$$

1. Normalization of the dataset: The dataset A is normalized with step 1 of Algorithm I as:

$$N = \begin{bmatrix} 0.00 & 0.40 & 0.00 & 1.00 & 0.88 \\ 1.00 & 0.00 & 0.33 & 0.56 & 0.75 \\ 0.00 & 1.00 & 1.00 & 0.89 & 0.00 \\ 1.00 & 0.20 & 0.78 & 0.00 & 0.25 \\ 1.00 & 0.40 & 0.78 & 0.44 & 1.00 \end{bmatrix} \tag{14}$$

2. The distance matrix, dis: Using the step 2 of the algorithm, we obtain the matrix as:

$$dis = \begin{bmatrix} 0.00 & 1.22 & 1.46 & 1.74 & 1.39 \\ 1.22 & 0.00 & 1.77 & 0.89 & 0.66 \\ 1.46 & 1.77 & 0.00 & 1.59 & 1.61 \\ 1.74 & 0.89 & 1.59 & 0.00 & 0.89 \\ 1.39 & 0.66 & 1.61 & 0.89 & 0.00 \end{bmatrix} \tag{15}$$

3. The product matrix, *Prod*: Using the step 3 of the algorithm, the *Prod* obtained is:

$$Prod = \begin{bmatrix} 4.35 & 2.37 & 4.30 & 2.59 & 2.74 \\ 1.55 & 2.69 & 2.97 & 3.08 & 1.95 \\ 4.97 & 1.55 & 3.08 & 3.16 & 4.62 \\ 1.79 & 2.65 & 2.59 & 4.05 & 3.09 \\ 1.55 & 2.35 & 2.53 & 3.19 & 1.93 \end{bmatrix} \tag{16}$$

4. Membership value matrix, M: The membership value matrix $M = \{[\mu_i d]\}$Using step 4, we obtain the matrix M as:

$$M = \begin{bmatrix} 0.230 & 0.423 & 0.232 & 0.386 & 0.365 \\ 0.645 & 0.371 & 0.337 & 0.325 & 0.514 \\ 0.201 & 0.645 & 0.324 & 0.316 & 0.217 \\ 0.560 & 0.377 & 0.386 & 0.247 & 0.324 \\ 0.644 & 0.426 & 0.395 & 0.314 & 0.518 \end{bmatrix} \tag{17}$$

5. Non-membership value matrix, N and Hesitancy matrix, H: Using step 5, we obtain the matrix N and H as:

$$N = \begin{bmatrix} 0.199 & 0.080 & 0.197 & 0.096 & 0.106 \\ 0.022 & 0.103 & 0.122 & 0.129 & 0.050 \\ 0.228 & 0.022 & 0.129 & 0.134 & 0.212 \\ 0.038 & 0.100 & 0.096 & 0.184 & 0.129 \\ 0.022 & 0.079 & 0.092 & 0.135 & 0.049 \end{bmatrix} \tag{18}$$

$$H = \begin{bmatrix} 0.571 & 0.497 & 0.571 & 0.518 & 0.529 \\ 0.333 & 0.526 & 0.542 & 0.547 & 0.437 \\ 0.571 & 0.333 & 0.547 & 0.550 & 0.571 \\ 0.402 & 0.522 & 0.518 & 0.569 & 0.547 \\ 0.334 & 0.495 & 0.513 & 0.551 & 0.434 \end{bmatrix} \tag{19}$$

The matrix triplet (M, N, H) is the required AIFS dataset, $\tilde{X}$. We may observe that each data point of dataset carries some useful information. Once the original dataset is transformed to another form, there are chances of losing some useful information of the dataset. During the fuzzification process as given by Algorithm I, some information might get lost which can affect the performance of the clustering algorithm. Though Algorithm I seems to be one way to intuitionisticly fuzzify any given dataset, we may opt for one more way as a better options. To overcome this issue, a factor called $\aleph_i$ can be incorporated during the intuitionistic fuzzification process such that it allocates some value to each data point x_i so as to retain information given by x_i to be utilized efficiently. To find an appropriate $\aleph_i$ for each ith data-item, the criteria to calculate may differ from dataset to dataset. The intend is to maintain the true nature of the given dataset with least changes during the fuzzification processing. In the paper, we use the following way to compute the factor $\aleph_i$ as: First, the given dataset X is called in non-normalized form. Then the distance matrix *dis* is derived which is followed up with the computation of matrix *Prod*. The reciprocation of the matrix *Prod* is computed and is named as factor matrix $\aleph = \{\aleph_{id}, 1 \le i \le P, 1 \le d \le D\}$. Each factor $\aleph_i$ is then summed to the dataset, X such that

$$sum_{id} = (x_{id} + \aleph_{id}). \tag{20}$$

The matrix *sum* is then normalized and defined as $\tilde{sum} = \{(x_{id} + \aleph_{id})' | x_{id} \in X\}$. We define the membership value as:

$$\mu_{id} = (\tilde{sum})^{\beta}, \quad \beta > 0. \tag{21}$$

Here, the tuning parameter β optimizes the membership value of the data item. The value chosen for parameter β is always considered positive. For the allocation of non-membership values to each data-item, generalized intuitionistic fuzzy generator proposed by Kaushal et al. (2018) is used. Yager's intuitionistic fuzzy complement calculates the non-membership value which is given as:

$$\nu_{id} = (1 - \mu_{id}^{\alpha\beta})^{\frac{1}{\alpha}}. \tag{22}$$

The parameter $\alpha, \alpha \in (0, 1]$ tunes the non-membership value of the dataset. The value chosen for parameter α is always considered positive. The hesitancy value is deduced using the formula (23) given as:

$$\pi_{id} = 1 - \mu_{id} - (1 - \mu_{id}^{\alpha\beta})^{\frac{1}{\alpha}}. \tag{23}$$

Hence, the AIFS data are represented as $\hat{x}_{id} = (\mu_{id}, \nu_{id}, \pi_{id})$.

We call the dataset A here to perform its intuitionistic fuzzification to an AIFS dataset, $\tilde{A}$. Here the factor $\aleph_{id}$ is considered from the matrix N given in Eq. (14). The tuning parameter $\beta = 1$ and the parameter considered is $\alpha = 0.45$ in Eqs. (22) and (23). We use Algorithm II in the similar fashion and obtained the matrices M', N' and H' as:

$$M' = \begin{bmatrix} 0.020 & 0.354 & 0.000 & 1.000 & 0.787 \\ 1.000 & 0.000 & 0.401 & 0.556 & 0.805 \\ 0.000 & 1.000 & 1.000 & 0.842 & 0.000 \\ 0.941 & 0.162 & 0.854 & 0.000 & 0.275 \\ 0.999 & 0.357 & 0.862 & 0.449 & 1.000 \end{bmatrix} \quad (24)$$

$$N' = \begin{bmatrix} 0.657 & 0.112 & 1.000 & 0.000 & 0.006 \\ 0.000 & 1.000 & 0.089 & 0.039 & 0.005 \\ 1.000 & 0.000 & 0.000 & 0.003 & 1.000 \\ 0.000 & 0.275 & 0.003 & 1.000 & 0.162 \\ 0.000 & 0.111 & 0.002 & 0.070 & 0.000 \end{bmatrix} \quad (25)$$

$$H' = \begin{bmatrix} 0.323 & 0.534 & 0.000 & 0.000 & 0.207 \\ 0.000 & 0.000 & 0.510 & 0.405 & 0.190 \\ 0.000 & 0.000 & 0.000 & 0.155 & 0.000 \\ 0.059 & 0.563 & 0.144 & 0.000 & 0.563 \\ 0.001 & 0.533 & 0.136 & 0.481 & 0.000 \end{bmatrix} \quad (26)$$

If we compare AIFS dataset (M, N, H) with (M', N', H') given in Eqs. (17), (18), (19) and Eqs. (24), (25), (26), respectively, we may observe that the weightage to non-membership value matrix is more in the former case whereas in the latter case, more weightage to membership value has been given. The main idea to deal with uncertainty in the dataset through an AIFS model depends on the three components of it. Sometimes, membership component plays an important role and sometimes the other (or both). It depends on the given dataset. We consider both the Method I and Method II to generate AIFS dataset from the given crisp dataset.

Therefore, on the basis of intuitionistic fuzzification of the dataset X, we may now define the proposed clustering algorithm called generalized intuitionistic fuzzy c-means (G-IFCM) with an adaptive distance measure for intuitionistic fuzzy datasets. The set $\tilde{X} = \{\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_P\}$ with $\hat{x}_i = (\mu_i, \nu_i, \pi_i)$ is an AIFS representation of ith data-point which possesses D features and every feature is represented as AIFS. The criterion function can be defined as follows:

$$J_m = \sum_{l=1}^{C} \sum_{i=1}^{P} u_{li}^m \tilde{D}(\hat{x}_i, \hat{s}_l),$$

where

$$\tilde{D}(\hat{x}, \hat{s}) = \frac{1}{2P} \sum_{d=1}^{D} \Big((\mu_{id} - \tilde{\mu}_{ld})^2 + (\nu_{id} - \tilde{\nu}_{ld})^2 + (\pi_{id} - \tilde{\pi}_{ld})^2 \Big) \quad (27)$$

$$\text{such that} \sum_{l=1}^{c} u_{il} = 1, \quad 1 \le i \le P \quad (28a)$$

$$u_{il} \in [0, 1] \quad 1 \le i \le P, 1 \le l \le c \quad (28b)$$

$$\sum_{i=1}^{P} u_{il} > 0, \quad 1 \le l \le c, \quad (28c)$$

where $\hat{S} = \{\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_c; \hat{s}_l = (\tilde{\mu}_l, \tilde{\nu}_l, \tilde{\pi}_l), 1 \le l \le c\}$ be the set of centroids of these c clusters, u_{il} represents membership degree of ith sample $\hat{x}_i$ in lth cluster $\hat{s}_l$ and $M = (u_{il})_{P \times c}$ represents a membership matrix of order $P \times c$.

To solve the membership matrix U and the centroids $(\tilde{\mu}_l, \tilde{\nu}_l, \tilde{\pi}_l)$, the Lagrange's condition of undetermined multipliers for G-IFCM is used. Thus, the membership value μ_{si} for ith data-item in lth centroid is calculated as follows:

$$u_{li} = \frac{(\tilde{D}(\hat{x}_i, \hat{s}_l))^{\frac{1}{m-1}}}{\sum_{k=1}^{c} (\tilde{D}(\hat{x}_i, \hat{s}_k))^{\frac{1}{m-1}}}. \quad (29)$$

The centroids $(\tilde{\mu}_l, \tilde{\nu}_l, \tilde{\pi}_l)$ are updated using the formula given below:

$$\tilde{\mu}_l = \frac{\sum_{i=1}^{P} u_{il}^m \mu_{\hat{x}_i}}{\sum_{i=1}^{P} u_{il}^m}, \tilde{\nu}_l = \frac{\sum_{i=1}^{P} u_{il}^m \nu_{\hat{x}_i}}{\sum_{i=1}^{P} u_{il}^m}, \tilde{\pi}_l = \frac{\sum_{i=1}^{P} u_{il}^m \pi_{\hat{x}_i}}{\sum_{i=1}^{P} u_{il}^m}. \quad (30)$$

Therefore, we implement the proposed G-IFCM in the following Sect. 3.1.

Algorithm I: Intuitionistic fuzzification Method I

Input: Dataset (X), Intuitionistic fuzzy parameter (α)

Output: Intuitionistic fuzzy data point, (μ_{id}, ν_{id}, π_{id}) corresponding to each dimension x_{id} of ith data-item.

1. Normalize the dataset, X using the Eqn. (8) to obtain matrix N.
2. Deduce the distance, dis_{ij} between each data-item x_i with data-item x_j using Euclidean distance measure.
3. Deduce the matrix $Prod$ using Eqn. (9) to find the product of the distance matrix, dis with normalized feature dataset, N.
4. Reciprocate the product using the formula given in Eqn. (10) to obtain μ_{id}.
5. Using Eqn. (11) and Eqn. (12), compute the non-membership value ν_{id} and hesitancy π_{id} for the dimension x_{id} of ith data-item.

End procedure

Algorithm II: Intuitionistic fuzzification Method II

Input: Dataset (X), Intuitionistic fuzzy parameter (α), Tuning parameter for membership function (β)

Output: Intuitionistic fuzzy data point, (μ_{id}, ν_{id}, π_{id}) corresponding to each dimension x_{id} of ith data-item.

1. Deduce the distance, dis_{ij} between each data-item, x_i with data-item x_j using Euclidean distance.
2. Deduce the matrix Sum using Eqn. (20) with factor $\aleph$.
3. Using Eqn. (21), compute the membership value μ_{id} for each dth dimension of the data-item x_i.
4. Using Eqn. (22) and Eqn. (23), compute the non-membership value ν_{id} and hesitancy π_{id} for the dth dimension of data-item x_i.

End procedure

Algorithm III: Generalized-IFCM clustering algorithm

Input: Dataset (X), Number of centroids (c), Fuzzy factor (m), Intuitionistic fuzzy parameter (α), Tuning parameter for membership function (β), Tolerance level (ε)

Output: Fuzzy partition U, centroids (μ_l, ν_l, π_l), $1 \leq l \leq c$

1. Data fuzzification using Algorithm I (or Algorithm II).
2. Initialize centroid $\hat{S}$, U at $t = 0$
3. **Repeat**
4. Update $(U = u_{il})^{t+1}$ by calculating the fuzzy partition using Eqn. (29)
5. Update centroid (μ_l, ν_l, π_l) using Eqn. (30)
6. **Until** $\sum_{i=1}^{c} \frac{d_2(S_i(t), S_i(t+1))}{c} < \varepsilon$ is satisfied.
7. **Return** U^{t+1}, μ_l^{t+1}, ν_l^{t+1} and π_l^{t+1}.

End procedure

3.1 Procedure to implement G-IFCM algorithm

The proposed algorithm G-IFCM can be divided into five steps initiating from the intuitionistic fuzzification of the dataset. It can be implemented to cluster the dataset X with the help of following steps as follows

Step 1.: *Intuitionistic fuzzification of dataset, X*: Using the above proposed technique for intuitionistic fuzzification of any crisp dataset, we call may call Algorithm I (or Algorithm II) to generate $\tilde{X}$.

Step 2.: *Initialization of centroid*: Any IFCM algorithm is initialized with a prior number of clusters. Here, the initialization of cluster centroids, $\hat{s}_l$, where $1 \leq l \leq c$ are done with randomization using the `rand` function.

Step 3.: *Computation of Membership matrix, U*: A membership matrix, U depicts the membership values of each data points possessed to belong each cluster. Higher the value of the membership grade of a data point, more visible is the associativity of it in the particular cluster. Here, membership matrix, $U = [u_{kl}]$ has a dimension of $n \times l$, where n is the number of data points and l is the number of clusters. The mathematical formula used to calculate the membership matrix is given by Eq. (29).

Step 4.: *Updation of cluster centroids, $\hat{S}$*: The centroids of the clusters, $\hat{s}_l$ are obtained in the form of intuitionistic fuzzy numbers as (μ_l, ν_l, π_l) from the membership matrix obtained in Step 3. The cluster centroids use the qualitative information obtained from the membership matrix (29). It has also utilized the hesitant component in the centroid formulation. The centroid $\hat{S}$ can be computed using formula given by Eq. (30).

Step 5.: *Convergence criterion*: The convergence of the IFCM algorithm is decided by the error which is a positive number, ϵ equal to 10^{-6}. The convergence is met if the termination criterion $\sum_{l=1}^{C} \frac{d_2(\hat{S}_l(t), \hat{S}_l(t+1))}{c} < \epsilon$ is reached else the algorithm is repeated from Step 3.

4 Experimental analysis

In the section, the performance of proposed clustering algorithm, generalized IFCM (G-IFCM) has been compared with few other clustering algorithms namely, modified IFCM (mIFCM) Kumar and Harish (2018), novel-IFCM Chaira (2011), and FCM using various benchmark measuring indexes. The clustering algorithms FCM and novel-IFCM use the given dataset in its original form whereas in G-IFCM and mIFCM, the dataset is first converted to AIFS dataset to deal with uncertainty and vagueness present in the dataset itself. In mIFCM clustering algorithm (Kumar and Harish (2018); Kumar et al. (2019a)), the dataset is transformed to AIFS dataset using simple normalized technique. Here G-IFCM calls two proposed techniques of the paper which reflects an adaptive way to tranform the dataset into an AIFS dataset. In the experimentation, we use the intuitionistic fuzzification techniques (Method I and Method II) in mIFCM to improvise the clustering results over the UCI machine learning datasets.

4.1 UCI machine learning datasets

The experiments are performed on eight machine learning datasets including Iris, Haberman, *E. coli*, Breast Cancer, Two-Moon, Heart, Wine, and Zoo dataset. The datasets vary in their sizes and number of attributes. The clustering experiments performed on datasets are taken from the UCI machine learning repository. The link for the repository is (link:https://archive.ics.uci.edu/ml/datasets.php). The description of these datasets is given in the Table 2. The computation is conducted using MATLAB version 8.1 running on a PC with 3.40 GHz frequency and RAM 16 GB.

4.2 Benchmark measuring Indexes

The benchmark measuring indexes used in the evaluation of the clustering performance are defined as follows:

1. *Clustering Accuracy (CA)*: The clustering accuracy is defined as number of correctly classified data points

over the total number of data-points. Clustering Accuracy is calculated as:

$$\mathrm{CA} = \frac{\text{No. of correctly classified samples}}{\text{Total no. of samples in dataset}}. \tag{31}$$

It is an important criterion to measure the performance of the clustering method over a labeled dataset.

2. *Partition coefficient (PC)*: The cluster validation measure known as Partition Coefficient is defined as summation of square of membership value of *i*th data point in *j*th cluster. It is defined as follows:

$$\mathrm{PC} = \frac{1}{P}\sum_{i=1}^{P}\sum_{j=1}^{c} u_{ij}^2. \tag{32}$$

It represents the trade-off between the clusters.

3. *Partition index (SC)*: Partition index is defined as the ratio of compactness of a cluster to its separation with other clusters. The mathematical formulation of the index is given as follows:

$$\mathrm{SC} = \frac{\sum_{i=1}^{P}\sum_{j=1}^{C} u_{ij} d^2(x_k, s_i)}{\sum_{k=1}^{P} u_{ik} \sum_{t=1}^{C} d^2(s_i, s_t)}. \tag{33}$$

A lower value of SC represents the better partition of the dataset.

4. *Xie–Beni index (XB)*: Xie–Beni index computes the intra-cluster deviation and inter cluster distance of their centers for the partitioned dataset. The index is defined as:

$$\mathrm{XB} = \frac{\sum_{i=1}^{P}\sum_{j=1}^{c} u_{ij}^2 d^2(x_i, s_j)}{n(\min_{i \neq t} d^2(s_i, s_t))}. \tag{34}$$

A lower value of XB produces the most desirable partition of the dataset.

5. *Dunn separation index, (DI)*: Dunn Index asks the clusters to be "compact and separate" relative to the metric used. Let $A_1, A_2, \ldots, A_c$ be c-partition of X and let U be the partition matrix. Dunn index can be defined as:

$$\mathrm{DI} = \min_{1 \le i \le C}\left\{\min_{i+1 \le j \le C-1}\left\{\frac{dis(u_i, u_j)}{\max_{1 \le c \le k} dia(u_k)}\right\}\right\} \tag{35}$$

where $dis(u_i, u_j) = \min_{A_i \in u_i, A_j \in u_j} d(A_i, A_j),$ (36)

$$dia(u_k) = \min_{A_i, A_j \in u_k} d(A_i, A_j). \tag{37}$$

A lower value of DI closer to zero produces the most desirable partition of the dataset.

Table 2 Summary of UCI datasets

Dataset	Instances	Features	Classes
Two moon	200	2	2
Iris	150	4	3
Zoo	101	17	7
Wine	178	13	3
Haberman	306	3	2
E. coli	327	7	5
Heart	270	13	2
Breast cancer	569	30	2

4.3 Analysis of G-IFCM algorithm

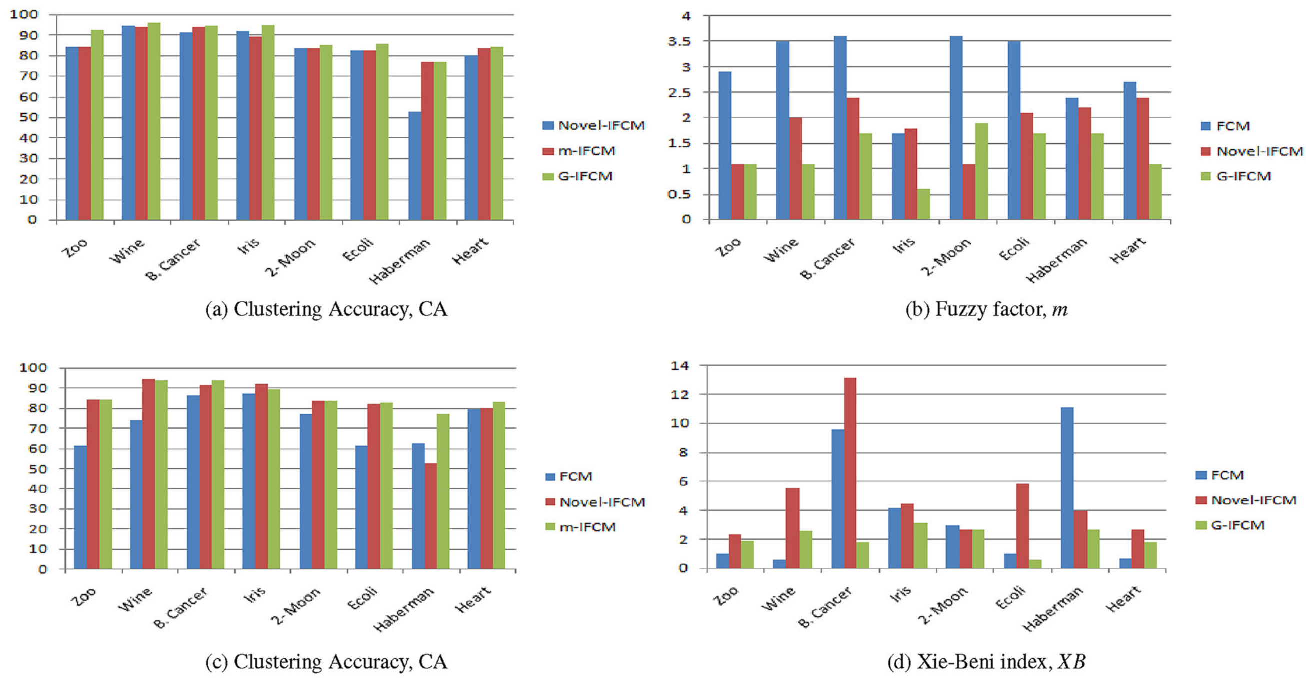
The experiment is a three-phase procedure. In Phase-I, the dataset goes through the fuzzification process where it is converted to AIFS using the Algorithm I (Method I) and Algorithm II (Method II). In Method I, parameter, namely α is tuned in the process; whereas in Method II, two parameters, namely α and β are tuned to get better AIFS dataset. The range of the parameter α is closed interval [0, 1] whereas the range of parameter β is considered as closed interval, [0.5, 10]. To each value of α and β, the fuzzification takes place. In our experiment, we have selected 30 values of α as $\{0.05, 0.10, 0.15, \ldots, 1\}$ and 10 values of β as $\{0.5, 1.5, 2.5, 3.5, 4.5, 5.5, \ldots, 9.5\}$. Once the dataset is fuzzified to AIFS using the pair of (α, β) from the given domain, the experiment enters Phase-II which calls the intuitionistic fuzzified dataset for clustering using Algorithm III. The clustering is performed on the UCI machine learning datasets using the proposed clustering algorithm G-IFCM. The performance is improved with initialization of the algorithm by selecting optimal fuzzy factor, m from the given range [1, 4]. The desirable convergence of the clustering algorithm leads to the Phase-III of the experiment. It covers the analysis of performance of the clustering algorithm using benchmark measuring indexes, known as partition coefficient (PC), partition index (SC), Xie–Beni index (XB), and clustering accuracy (CA).

The results obtained in the clustering algorithms are summarized in Table 3. We have tested the algorithms with proposed technique on the Eight UCI datasets. The Method I and Method II (see Algorithm I and II) are incorporated in proposed Generalized IFCM, namely G-IFCM. We have also incorporated the AIFS technique in mIFCM clustering algorithm. It can be clearly seen in Table 3 that the clustering results obtained by G-IFCM (Method II) gives better accuracies with comparison to all other algorithms including FCM and novel-IFCM. The results for mIFCM are also improvised and better than FCM and novel-IFCM with incorporation of proposed intuitionistic fuzzification

Table 3 Comparison of G-IFCM with other clustering algorithms using intuitionistic fuzzification techniques

Dataset	Methods	Algorithms	(m, α, β)	Accuracy	PC	SC	XB	DI
	–	FCM	$m = 2.9$	61.3861	0.2135	0.4724	0.9970	0.5248
	–	Novel-IFCM	(1.3, 0.05, –)	84.1584	–	–	2.2995	0.2587
	Method I	mIFCM	(1.1, 0.10, –)	87.1287	–	–	17.391	0.2897
Zoo		G-IFCM	(1.2, 0.05, 3.5)	85.1485	0.9454	0.0251	3.1447	0.0045
	Method II	mIFCM	(1.4, 0.90, 1.0)	84.1584	–	–	0.9087	0.1365
		G-IFCM	(**1.1**, 0.15, 3.5)	**92.079**2	**0.9907**	**0.3658**	**1.8506**	**0.0048**
	–	FCM	$m = 3.5$	73.5955	0.3343	1.8114	0.6095	0.7547
	–	Novel-IFCM	(2, 0.05, –)	94.3820	–	–	5.5901	0.4569
	Method I	mIFCM	(1.7, 0.20, –)	77.5281	–	–	37.325	0.9847
Wine		G-IFCM	(1.1, 0.35, 1.5)	94.3820	0.9654	0.1236	4.0179	0.0023
	Method II	mIFCM	(2.1, 0.95, 1.5)	93.8202	–	–	3.7793	0.2365
		G-IFCM	(**1.1**, 0.45, 2.5)	**96.0674**	**0.9668**	**0.1254**	**2.5902**	**0.1345**
	–	FCM	$m = 3.6$	85.9402	0.6936	0.0006	9.6285	0.3587
	–	Novel-IFCM	(2.4, 0.45, –)	91.2127	–	–	13.111	0.8521
	Method I	mIFCM	(1.6, 0.10, –)	**90.3339**	–	–	20.825	0.7412
Breast Cancer		G-IFCM	(2.4, 0.25, 8.5)	89.2794	0.5000	0.1254	1.9739	0.0654
	Method II	mIFCM	(1.9, 0.35, 1.5)	94.0246	–	–	1.6468	0.0134
		G-IFCM	(**1.7**, 0.35, 1.5)	**94.5674**	**0.5776**	**0.1212**	**1.8100**	**0.0014**
	–	FCM	$m = 1.7$	87.3333	0.8373	0.5239	4.1911	0.6589
	–	Novel-IFCM	(1.8, 0.05, –)	92.0000	–	–	4.4964	0.9585
	Method I	mIFCM	(1.8, 0.05, –)	89.3333	–	–	27.381	0.7897
Iris		G-IFCM	(3.9, 0.25, 8.5)	74.0000	0.3333	0.2587	2.5049	0.0045
	Method II	mIFCM	(1.8, 0.35, 1.5)	89.3333	–	–	7.2306	0.4587
		G-IFCM	(**1.6**, 0.25, 1.5)	**95.0000**	**0.7142**	**0.3578**	**3.1391**	**0.0404**
	–	FCM	$m = 3.6$	77.0000	0.5004	3.5921	2.9776	0.7938
	–	Novel-IFCM	(1.1, 0.05, –)	83.5000	–	–	2.6948	0.9989
	Method I	mIFCM	(2.4, 0.90, –)	83.5000	–	–	20.488	0.1345
Two-Moon		G-IFCM	(2.1, 0.10, 8.5)	85.0000	0.8000	0.2000	1.1101	0.0245
	Method II	mIFCM	(1.1, 0.65, 1.5)	83.5000	–	–	21.407	0.6541
		G-IFCM	(**1.9**, 0.65, 0.5)	**85.0000**	**0.8000**	**0.2012**	**2.7051**	**0.0069**
	–	FCM	$m = 3.5$	61.4679	0.2212	2.6949	1.0098	0.6473
	–	Novel-IFCM	(2.1, 0.05, –)	82.2630	–	–	5.8392	0.5468
	Method I	mIFCM	(2.0, 0.20, –)	64.2202	–	–	12.438	0.4251
Ecoli		G-IFCM	(2.6, 0.05, 0.5)	64.5260	0.8000	0.1354	41.891	0.0134
	Method II	mIFCM	(1.1, 1.00, 6.5)	82.5688	–	–	5.1040	0.0045
		G-IFCM	(**1.7**, 0.65, 1.5)	**85.5648**	**0.2806**	**0.0258**	**0.6001**	**0.0036**
	–	FCM	$m = 2.4$	62.4183	0.5000	5.3642	11.103	0.9987
	–	Novel-IFCM	(2.2, 0.30, –)	52.6144	–	–	4.0010	0.4615
	Method I	mIFCM	(1.1, 0.05, –)	75.4902	–	–	20.454	0.9987
Haberman		G-IFCM	(3, 0.25, 3.5)	73.5294	0.8000	0.5241	13.373	0.6413
	Method II	mIFCM	(1.9, 0.70, 1.5)	76.7974	–	–	5.7991	0.5545
		G-IFCM	(**1.7**, 0.90, 3.5)	**77.0000**	**0.7000**	**0.5201**	**2.7028**	**0.3125**
	–	FCM	$m = 2.7$	79.6296	0.5000	1.9513	0.6701	0.5785
	–	Novel-IFCM	(2.4, 0.85, –)	80.0000	–	–	2.6577	0.4651
	Method I	mIFCM	(2.4, 0.55, –)	**84.4444**	–	–	3.6730	0.6452
Heart		G-IFCM	(2.8, 0.15, 7.5)	78.8889	0.8000	0.2001	3.2590	0.0134
	Method II	mIFCM	(1.3, 0.15, 9.5)	83.3333	–	–	0.7831	0.4251
		G-IFCM	(**1.1**, 0.15, 8.5)	**84.0741**	**0.9168**	**0.3002**	**1.8283**	**0.0005**

The optimal values of the tuning parameters, cluster indices with respective clustering accuracy results obtained are in bold



(a) Clustering Accuracy, CA

(b) Fuzzy factor, m

(c) Clustering Accuracy, CA

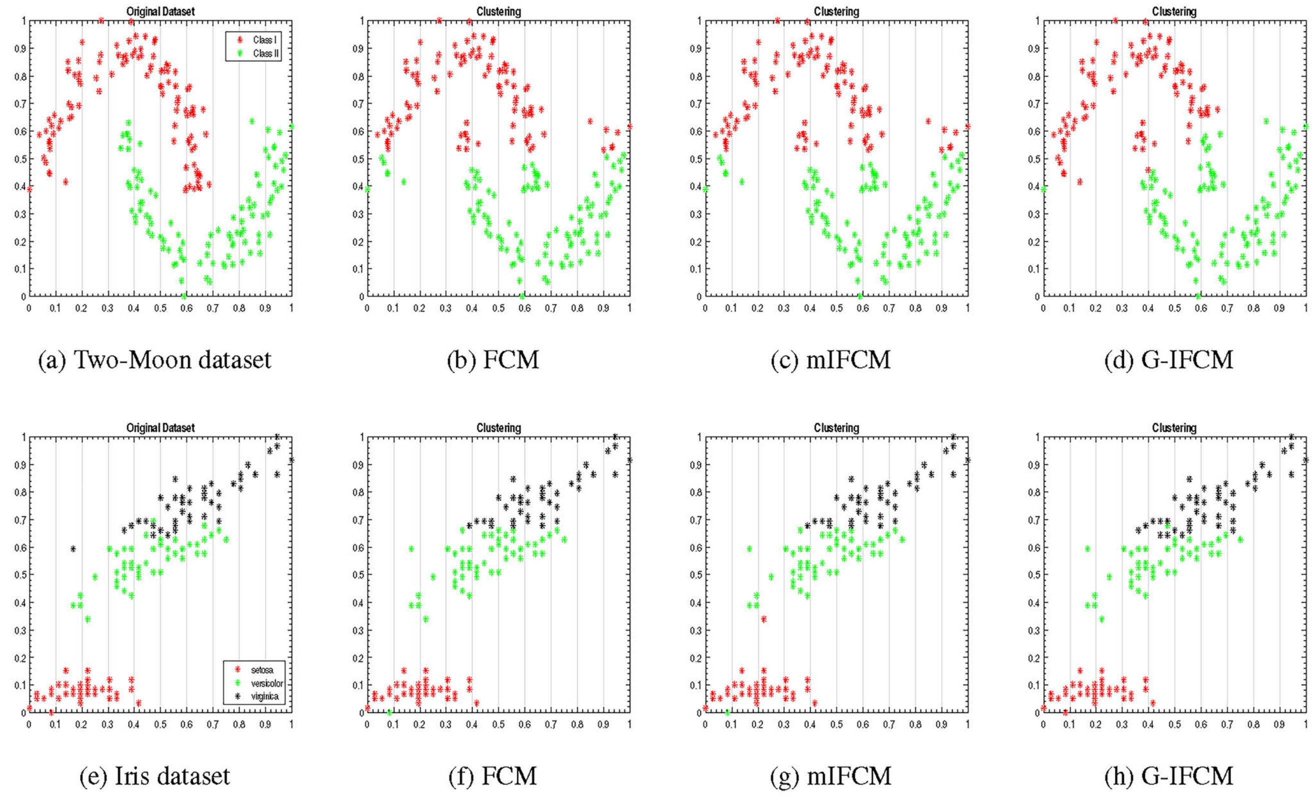
(d) Xie-Beni index, XB

Fig. 2 Clustering performance comparison of G-IFCM with FCM, mIFCM and novel-IFCM algorithm

techniques. Therefore, out of the four algorithms where we incorporated Method I and Method II, the performance of G-IFCM (Method II) is graphically compared with FCM and novel-IFCM in Fig 2. Along with it, the performance of mIFCM has also been compared with FCM and novel-IFCM in Fig 2c. Further, a 2D graphical comparison of the proposed algorithm G-IFCM has also been executed with FCM and mIFCM over Iris dataset and Two-Moon dataset in Fig. 3. It can be observed in Fig. 3d, the clustering of Two-Moon dataset with G-IFCM is more refined as compared with FCM and mIFCM. Similarly, the clustering 2D graph obtained in Fig. 3h shows better performance of G-IFCM over other two algorithms for the Iris dataset.

The clustering accuracies (CA) of the FCM, novel-IFCM, G-IFCM, and mIFCM are elaborated in Table 3. As the labels for each cluster may change in multiple runs of the algorithm, CA is computed by first matching the obtained labels of the data-item in the concerned clusters with the true labels of the data-items in the concerned clusters. It can be clearly seen that CA obtained by G-IFCM (Method II) is higher than the accuracies obtained with FCM and novel-IFCM due to incorporation of intuitionistic data fuzzification technique and optimal selection of the parameters m, α and β. In Fig. 2a and c, we can clearly see, the performance of G-IFCM and mIFCM rules out FCM and novel-IFCM. The adaptive technique of converting dataset to AIFS dataset handles the uncertainty well with optimal selection of the three parameters in Method II. The bars in Fig. 2b showcase the search space of m used in all the four clustering algorithms. The search space obtained for m in FCM obtained is [1, 3.7] and for novel-IFCM, it is [1, 2.4]. The proposed techniques in G-IFCM (Method II) has tremendously reduced the search space of m to closed interval [1, 2]. Once the search space of fuzzy factor, m reduces, the number of running iterations reduces tremendously. Therefore, the running computation of G-IFCM algorithm gets minimized.

The efficacy of the clustering algorithms are also measured with four other indexes, namely PC, SC, XB, and DI. The indexes DI, PC, and SC defines the better compactness and well separated clusters without affecting the trade-off between the clusters for G-IFCM in comparison to FCM and novel-IFCM in Table 3. As stated in the above definitions in Sect. 4.2, the lower values of SC and DI whereas higher values of PC denote good clustering performance of the algorithm.Most of the PC values of G-IFCM lie in the range of interval [0.5, 1] which shows good performance of G-IFCM than other FCM and mIFCM. If the clustering algorithm is non-normalized, then partition coefficient, (PC) is not an appropriate performance measuring index (Refer Wu 2008). In the non-normalized clustering algorithms, Novel-IFCM and mIFCM, the terminating matrix column sum need is not always equal to one. Therefore, the final matrix with membership value of ith data-item in jth cluster for all i, j does not end to a membership matrix, U. Similarly, SC is also a membership dependent performance measuring index (Refer to Xie and Beni (1991)), it may not be a suitable index for novel-IFCM and mIFCM clustering algorithm. To compare the performance of proposed G-IFCM with the non-normalized algorithms, we



(a) Two-Moon dataset (b) FCM (c) mIFCM (d) G-IFCM

(e) Iris dataset (f) FCM (g) mIFCM (h) G-IFCM

Fig. 3 Comparison of clustering performance of G-IFCM with FCM and mIFCM over Two-Moon and Iris dataset

choose two other measuring indexes called Dunn index (DI) and Xie–Beni Index (XB). The Xie–Beni index calculates the ratio of intra cluster distance with inter cluster distance and lower value of it depicts the better performance of the clustering algorithm. The values of index XB for G-IFCM (Method II) lies in the interval [0.5, 3.2] unlike FCM, novel-IFCM and mIFCM as shown in Fig. 2d. Similarly, the DI value for G-IFCM over most of the dataset is below 0.5 unlike other three clustering algorithms and therefore,the clusters produced by it are compact and separate. The intuitionistic data fuzzification technique incorporated in G-IFCM has improved the clustering results based on some measuring indexes. Therefore, the Generalized IFCM might hold the property to cluster more datasets with uncertainty.

5 Conclusion

In the paper, a new version of intuitionistic fuzzy c-means algorithm namely, generalized intuitionistic fuzzy c-means clustering algorithm (G-IFCM) has been proposed to clusters the ambiguous datasets using an adaptive AIFS distance measure. First, two adaptive techniques of intuitionistic fuzzification of a crisp dataset are introduced, namely Method I and Method II that generate AIFS datasets from the given dataset. The Method II uses a factor, denoted as $\aleph$, in the membership value of each data-item such that it maintains the true nature of the dataset which tends to change during any fuzzification processing. Both the methods are proposed with the idea to perform clustering under AIFS environment wherever uncertainty is present. The performance of the proposed algorithm is checked over UCI machine learning datasets and is also compared with the other well known fuzzy clustering algorithms, namely FCM, novel-IFCM and mIFCM. The study shows that the proposed technique in G-IFCM clusters the datasets efficiently and minimizes the search space of fuzzy factor, m to [1, 2] which reduces the expense of computation. It also improves the performance of mIFCM over machine learning datasets with incorporation of proposed generation technique of AIFS in the paper. Along with it, the values of various benchmark indexes (PC, SC, XB, and DI) show the better compact and separate clusters obtained by G-IFCM. In our future work, we will extend our work to study feature values of data-items in the clusters that possess outliers and noises using AIFS-based techniques.

Acknowledgements The research work of the first author is funded by University Grants Commission (UGC), New Delhi, India under the grant number 19/06/2016(i)EU-V.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Anderson D, Luke RH, Keller JM, Skubic M, Rantz M, Aud M (2009) Linguistic summarization of video for fall detection using voxel person and fuzzy logic. Comput Vis Image Underst 113(1):80–89

Atanassov KT (1986) Intuitionistic fuzzy sets. Fuzzy Sets Syst 20(1):87–96

Aziz AM (2009) Effects of fuzzy membership function shapes on clustering performance in multisensor-multitarget data fusion systems. In: 2009 IEEE international conference on fuzzy systems. IEEE, pp 1839–1844

Barua A, Mudunuri LS, Kosheleva O (2014) Why trapezoidal and triangular membership functions work so well: towards a theoretical explanation. J Uncertain Syst 8:1–5

Bede B (2019) Fuzzy systems with sigmoid-based membership functions as interpretable neural networks. In: International fuzzy systems association world congress. Springer, pp 157–166

Bezdek JC, Ehrlich R, Full W (1984) FCM: the fuzzy c-means clustering algorithm. Comput Geosci 10(2–3):191–203

Bustince H, Kacprzyk J, Mohedano V (2000) Intuitionistic fuzzy generators application to intuitionistic fuzzy complementation. Fuzzy Sets Syst 114(3):485–504

Castro JR, Castillo O, Melin P (2007) An interval type-2 fuzzy logic toolbox for control applications. In: 2007 IEEE international fuzzy systems conference. IEEE, pp 1–6

Chaira T (2011) A novel intuitionistic fuzzy c means clustering algorithm and its application to medical images. Appl Soft Comput 11(2):1711–1717

Chakraborty S, Chatterjee S, Das A, Mali K (2020) Penalized fuzzy c-means enabled hybrid region growing in segmenting medical images. In: Hybrid machine intelligence for medical image analysis. Springer, pp 41–65

Chen N, Zs Xu, Mm Xia (2014) Hierarchical hesitant fuzzy k-means clustering algorithm. Appl Math A J Chin Univ 29(1):1–17

Chen SM (1996) A fuzzy reasoning approach for rule-based systems based on fuzzy logics. IEEE Trans Syst Man Cybern Part B (Cybern) 26(5):769–778

Chen SM, Chang CH (2016) Fuzzy multiattribute decision making based on transformation techniques of intuitionistic fuzzy values and intuitionistic fuzzy geometric averaging operators. Inf Sci 352:133–149

Chen SM, Huang CM (2003) Generating weighted fuzzy rules from relational database systems for estimating null values using genetic algorithms. IEEE Trans Fuzzy Syst 11(4):495–506

Chen SM, Jong WT (1997) Fuzzy query translation for relational database systems. IEEE Trans Syst Man Cybern Part B (Cybern) 27(4):714–721

Chen SM, Cheng SH, Lan TC (2016) Multicriteria decision making based on the topsis method and similarity measures between intuitionistic fuzzy values. Inf Sci 367:279–295

Hellmann M (2001) Fuzzy logic introduction. Université de Rennes 1, pp 1–9

Huang CW, Lin KP, Wu MC, Hung KC, Liu GS, Jen CH (2015) Intuitionisti fuzzy c-means clustering algorithm with neighborhood attration in segmenting medial image. Soft Comput 19(2):459–470

Iakovidis DK, Pelekis N, Kotsifakos E, Kopanakis I (2008) Intuitionistic fuzzy clustering with applications in computer vision. In: International conference on advanced concepts for intelligent vision systems. Springer, pp 764–774

Ji Z, Sun Q, Xia Y, Chen Q, Xia D, Feng D (2012) Generalized rough fuzzy c-means algorithm for brain MR image segmentation. Comput Methods Programs Biomed 108(2):644–655

Kaushal M, Solanki R, Lohani QD, Muhuri PK (2018) A novel intuitionistic fuzzy set generator with application to clustering. In: 2018 IEEE international conference on fuzzy systems (FUZZ-IEEE). IEEE, pp 1–8

Kumar D, Verma H, Mehra A, Agrawal R (2019a) A modified intuitionistic fuzzy c-means clustering approach to segment human brain MRI image. Multimed Tools Appl 78(10):12663–12687

Kumar G, Kulshrestha N, Srivastava S (2019b) Analysis of different fuzzy c-means membership functions for the design of cad tool of early breast cancer detection. Available at SSRN 3350301, pp 335–341

Kumar SA, Harish B (2018) A modified intuitionistic fuzzy clustering algorithm for medical image segmentation. J Intell Syst 27(4):593–607

Kunwar V, Sabitha AS, Choudhury T, Aggarwal A (2019) Chronic kidney disease using fuzzy c-means clustering analysis. Int J Bus Anal (IJBAN) 6(3):43–64

Li G, Peng C, Fei M, Tian Y (2020) Local stability conditions for TS fuzzy time-delay systems using a homogeneous polynomial approach. Fuzzy Sets Syst 385:111–126

Li X, Mehran K (2019) Model-based control and stability analysis of discrete-time polynomial fuzzy systems with time delay and positivity constraints. IEEE Trans Fuzzy Syst 27(11):2090–2100

Lin KP (2013) A novel evolutionary kernel intuitionistic fuzzy *c*-means clustering algorithm. IEEE Trans Fuzzy Syst 22(5):1074–1087

Logambigai R, Kannan A (2016) Fuzzy logic based unequal clustering for wireless sensor networks. Wirel Netw 22(3):945–957

Mandal SN, Choudhury JP, Chaudhuri SB (2012) In search of suitable fuzzy membership function in prediction of time series data. Int J Comput Sci Issues 9(3):293–302

Melin P, Castillo O (2014) A review on type-2 fuzzy logic applications in clustering, classification and pattern recognition. Appl Soft Comput 21:568–577

Narimani M, Lam HK (2010) SOS-based stability analysis of polynomial fuzzy-model-based control systems via polynomial membership functions. IEEE Trans Fuzzy Syst 18(5):862–871

Nguyen DD, Ngo LT, Pham LT (2013) Interval type-2 fuzzy c-means clustering using intuitionistic fuzzy sets. In: 2013 third world congress on information and communication technologies (WICT 2013). IEEE, pp 299–304

Pedrycz W (1994) Why triangular membership functions? Fuzzy Sets Syst 64(1):21–30

Priya RV (2019) Emotion recognition from geometric fuzzy membership functions. Multimed Tools Appl 78(13):17847–17878

Qiu C, Xiao J, Yu L, Han L, Iqbal MN (2013) A modified interval type-2 fuzzy c-means algorithm with application in MR image segmentation. Pattern Recognit Lett 34(12):1329–1338

Shyi-Ming C, Jyh-Sheng K, Jin-Fu C (1990) Knowledge representation using fuzzy petri nets. IEEE Trans Knowl Data Eng 2(3):311–319. https://doi.org/10.1109/69.60794

Singh R, Vashishath M, Kumar S (2019) Ant colony optimization technique for edge detection using fuzzy triangular membership function. Int J Syst Assur Eng Manag 10(1):91–96

Verma H, Agrawal R, Sharan A (2016) An improved intuitionistic fuzzy c-means clustering algorithm incorporating local information for brain image segmentation. Appl Soft Comput 46:543–557

Wu D (2012) Twelve considerations in choosing between gaussian and trapezoidal membership functions in interval type-2 fuzzy logic controllers. In: 2012 IEEE international conference on fuzzy systems. IEEE, pp 1–8
Wu KL (2008) An analysis of robustness of partition coefficient index. In: 2008 IEEE international conference on fuzzy systems (IEEE world congress on computational intelligence). IEEE, pp 372–376
Xie XL, Beni G (1991) A validity measure for fuzzy clustering. IEEE Trans Pattern Anal Mach Intell 13(8):841–847
Xu Z, Wu J (2010) Intuitionistic fuzzy c-means clustering algorithms. J Syst Eng Electron 21(4):580–590
Yager RR (1980) On the measure of fuzziness and negation. II. Lattices. Inf Control 44(3):236–260
Zadeh LA (1965) Fuzzy sets. Inf Control 8(3):338–353
Zadeh LA (1996) Soft computing and fuzzy logic. In: Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi a Zadeh. World Scientific, pp 796–804
Zeng S, Chen SM, Kuo LW (2019) Multiattribute decision making based on novel score function of intuitionistic fuzzy values and modified VIKOR method. Inf Sci 488:76–92
Zhao F, Chen Y, Liu H, Fan J (2019) Alternate PSO-based adaptive interval type-2 intuitionistic fuzzy c-means clustering algorithm for color image segmentation. IEEE Access 7:64028–64039
Zhou X, Zhao R, Yu F, Tian H (2016) Intuitionistic fuzzy entropy clustering algorithm for infrared image segmentation. J Intell Fuzzy Syst 30(3):1831–1840