# Fuzzy C-Medoids Clustering Based on Interval Type-2 Intuitionistic Fuzzy Sets

Nguyen Anh Cuong
*Faculty of Information Technology*
*Le Quy Don Technical University*
Hanoi, Vietnam
nguyenanhcuong@lqdtu.edu.vn

Dinh Sinh Mai
*Institute of Techniques for Special Engineering*
*Le Quy Don Technical University*
Hanoi, Vietnam
maidinhsinh@lqdtu.edu.vn

Do Viet Duc
*Faculty of Information Technology*
*Le Quy Don Technical University*
Hanoi, Vietnam
dovietduc@gmail.com

Dang Trong Hop
*Faculty of Information Technology*
*Hanoi University of Industry*
Hanoi, Vietnam
hopdt@haui.edu.vn

Long Thanh Ngo
*Faculty of Information Technology*
*Le Quy Don Technical University*
Hanoi, Vietnam
ngotlong@mta.edu.vn

Pham The Long
*Faculty of Information Technology*
*Le Quy Don Technical University*
Hanoi, Vietnam
longpt@mta.edu.vn

***Abstract*—For clustering problems, each data sample has the potential to belong to many different clusters depending on the similarity. However, besides the degree of similarity and non-similarity, there is a degree of hesitation in determining whether or not a data sample belongs to a defined cluster. Besides the fuzzy c-means algorithm (FCM), another popular algorithm is fuzzy C-medoids clustering (FCMdd). FCMdd chooses several existing objects as the cluster centroids, while FCM considers the samples' weighted average to be the cluster centroid. This subtle difference causes the FCMdd is more resistant to interference than FCM. Since noise samples will more easily affect the center of centroids of the FCM, it is easier to create clustering results with great accuracy. In this paper, the interval type-2 intuitionistic fuzzy c-medoids clustering algorithm (IT2IFCMdd) is proposed by extending the fuzzy c-medoids clustering based on interval type-2 intuitionistic fuzzy sets. With this combination, the proposed algorithm can take advantage of both the fuzzy c-medoids clustering (FCMdd) method and the interval type-2 intuitionistic fuzzy sets applied to the clustering problem. Experiments performed on data sets commonly used in machine learning show that the proposed method gives better clustering results in most experimental cases.**

***Index Terms*—Fuzzy clustering, c-Medoids clustering, Interval type-2 fuzzy set, Intuitionistic fuzzy set.**

## I. Introduction

Clustering is an unsupervised learning method widely used in various fields like data mining, information retrieval, computer vision, bio-informatics, so on. Clustering algorithms aim to organize a set of objects into clusters such that items in a given cluster have a high degree of similarity, while items in different clusters have a high degree of dissimilarity [4], [5]. The most common clustering techniques are hierarchy and partitioning [12]: the methods of hierarchy yield a complete hierarchy, that is, a nested series of partitions of the first data. While the partitioning method seeks to take a single partition of the input data into a fixed number of clusters, often by optimizing an objective function. That is after clustering is completed, the total distance from all points in the cluster to the cluster's centroid must be the minimum [16].

In the hard clustering approach, each object of the data set must be assigned exactly one cluster. After the fuzzy set theory came into being, clustering methods based fuzzy sets allowing one data point to belong to more than one cluster. It provides a fuzzy partition based on the idea of the partial membership of each sample in a given cluster. Fuzzy clustering is considered to be good method for capturing the uncertainty of real data [11]. Other most common fuzzy clustering algorithms and applications have been introduced in [6], [10], [20].

To define a data sample will belong to a certain cluster, the membership functions are used. The value of the membership function specified is in the range [0,1]. One point can be divided into many clusters but the total value of membership functions of a given object on all clusters is always 1.

There are many ways to determine the center of centroids of a cluster, of which 2 are commonly used include : Based on the average of the data samples [17] (computing by average distance) such as K-means (KM), fuzzy C-means (FCM) [4] and by medoid (representative points) such as k-Medoids (KMdd), fuzzy c-Medoids (FCMdd) [8], [15], [25]. In the second approach, each cluster will select a point representing its cluster, called the medoid point, which acts as the centroid where total distance of all objects in the cluster to its centroid is the minimum.

The intuitionistic fuzzy set (IFS) was introduced by Atanassov in [1] as an extension of the fuzzy set theory [18]. In 2020, he was summarized and built up the basic documents and expanded the details of the interval type-2 intuitionistic fuzzy set (IT2IFS) [2], which mentioned the membership degree, the degree of non-membership and the degree of hesitation. In the last decade, many scholars have devoted themselves to the study of IFS, which are widely used in a variety of fields and achieve valuable results.

978-1-6654-0435-8/21/$31.00 ©2021 IEEE

Charia et al. [7] proposed an edge detection method based on IFS on the fuzzy c-mean clustering algorithm (IFCM) which applied to detect medical images that combine local information with the performance segment high. Ansari et al. [3] proposed a new divergence and entropy measures for IFS on edge detection. Hua Zhao et al. [24] introduced an intuitionistic fuzzy clustering algorithm based on the Boole matrix and association measure and gave a specific example of the implementation method. Sahil et al. [19] introduced the intuitionistic fuzzy metric space with properties. Dzung et al. [9] proposed a new method by combining the advantages of intuitionistic fuzzy sets and interval type-2 fuzzy clustering algorithms to overcome the drawbacks of fuzzy clustering.

Besides the fuzzy c-means algorithm (FCM), another popular algorithm is fuzzy C-medoids clustering (FCMdd). Both the FCM and FCMdd algorithms try to minimize the target function and give the partition array $U$ and the set of $V$ cluster centers. The main difference between FCM and FCMdd is only in the mechanism of forming cluster centroids. FCMdd chooses a number of existing objects as the cluster center, while FCM considers the weighted average of the objects to be the cluster centroid. This subtle difference causes the two algorithms to have different performance characteristics: the FCMdd is more resistant to interference than FCM, and since noise objects will more easily affect the center of centroids of the FCM it is easier to create clustering results with great accuracy. From these reasons, the paper proposes the IT2IFCMdd algorithm to take advantage of the advantages of IFS, IT2FS, and FCMdd.

The remainder of the paper is organized as follows: Sect.II introduces related knowledge: summarizes some basic concept about IFS, FCMdd, and IT2IFS; Sect.III proposed method; Sect.IV shows some experimental results and discussion. Sect.V some conclusions.

## II. Background

### A. Intuitionistic Fuzzy Sets

An intuitionistic fuzzy set $\tilde{A}$ is represented by: $\tilde{A} = \{x, \mu_{\tilde{A}}(x), v_{\tilde{A}}(x)) : x \in X\}$

Where: $X$ is an ordinary finite non-empty set, $\mu_{\tilde{A}} : X \to [0;1], v_{\tilde{A}} : X \to [0;1]$ must satisfy $\mu_{\tilde{A}}(x) + v_{\tilde{A}}(x) \leq 1$ for $\forall x \in X$, with the element $x$ in set $\tilde{A}$, $\mu_{\tilde{A}}(x)$: its degree of membership and $v_{\tilde{A}}(x)$: its degree of non-membership.

IFSs(x) can be considered as the set of all the intuitionistic fuzzy sets in $X$. For each IFS $\tilde{A}$ in $X$, the value $\pi_{\tilde{A}}(x) = 1 - \mu_{\tilde{A}}(x) - v_{\tilde{A}}(x)$ is called the uncertain degree of $x$ to $\tilde{A}$. One noteworthy issue with an IFS $\tilde{A}$, IF $\mu_{\tilde{A}}(x) = 0$, THEN $v_{\tilde{A}}(x) + \pi_{\tilde{A}}(x) = 1$, and IF $\mu_{\tilde{A}}(x) = 1$ THEN $v_{\tilde{A}}(x) = 0$ and $\pi_{\tilde{A}}(x) = 0$.

### B. The fuzzy c-medoids algorithm

The fuzzy c-medoids algorithm (*FCMdd*) [15] was proposed by R. Krishnapuram at al, let $X = \{x_1, x_2, \ldots, x_N\}$ with $N$ objects. Each object is represented by a property vector. Let $D_{ij} = d(x_i, x_j)$ describe the difference between the $x_i$ object and the $x_j$ object. Let $Z = \{Z_1, Z_2, \ldots, Z_K\}, z_i \in X$ describe a subset of $X$ with $K$ parts, i.e., $Z$ is a set $K$-subset of $X$. Let $X^K$ represent the set of all $K$-subsets $Z$ of $X$. *FCMdd* minimum target function:

$$J_m(U, v) = \sum_{k=1}^{N} \sum_{i=1}^{K} (u_{ik})^m (d_{ik})^2 \quad (1)$$

which minimizes each sub-cluster of $Z$ in $X^K$. In Eq. (1), $u_{ij}$ denotes the fuzzy degree [4] or the ability [12], [13], [14] membership of $x_j$ in the cluster:

$$u_{\mathrm{ij}} = \frac{(1/d(x_j, v_i))^{1/(m-1)}}{\sum_{k=1}^{c} (1/d(x_j, v_k))^{1/(m-1)}} \quad (2)$$

**Algorithm 1: FCMdd**
**Step 1: Begin**
1.1 Fixed number of clusters $K$; Set $iter$ = 0;
1.2 Initialize $Z = \{z_1, z_2, \ldots, z_K\}$ medoids from $X^K$;
**Step 2: REPEAT**
2.1 Compute fuzzy member matrix values $u_{ij}$ for $i = 1, 2, \ldots, K$ ; $j = 1, 2, \ldots, N$, using Eq.(2);
2.2 Store current medoids: $Z^{old} = Z$;
2.3 Compute new medoids: $z_i$ for $i = 1, 2, \ldots, K$:
2.3.1 $q = \underset{1 \leq k \leq N}{\arg\min} \sum_{j=1}^{N} u_{ij}^m d(x_k, x_j)$
2.3.2 $z_i = x_q$
3.3 $iter = iter + 1$;
**UNTIL** $\left(Z^{old} = Z \text{ or } iter = MAX\_ITER\right)$.

### C. Interval type-2 intuitionistic fuzzy sets

$\tilde{A}^*$ represents a type-2 intuitionistic fuzzy set in $X$, and $\mu_{\tilde{A}^*}(x, u_1)$ is membership grade $x \in X$ with $u_1 \in J_x^1 \subseteq [0,1]$, also $v_{\tilde{A}^*}(x, u_2)$ is non-membership grade $x \in X$ with $u_2 \in J_x^2 \subseteq [0,1]$.

The elements of domain of $(x, u_1), (x, u_2)$ are called primary membership and primary non-membership of $x \in \tilde{A}^*$, respectively, memberships of primary memberships $\mu_{\tilde{A}^*}(x, u_1)$ and non-memberships of primary memberships $v_{\tilde{A}^*}(x, u_1)$.

$\mu_{\tilde{A}^*}(x, u_2)$,$v_{\tilde{A}^*}(x, u_2)$ are called secondary memberships and secondary non-memberships, respectively, of $x \in \tilde{A}^*$, with $u_1 \in J_x^1 \subseteq [0,1]$, $u_2 \in J_x^2 \subseteq [0,1]$ which are intuitionistic fuzzy sets.

When a type-2 intuitionistic fuzzy sets satisfies: the secondary membership function $\mu_{\tilde{A}^*}(x, u_1) = 1$ and $\mu'_{\tilde{A}^*}(x, u_2) = 1 \ (\forall u_1, u_2 \in J_x)$ , then it is called an interval type-2 intuitionistic fuzzy sets and it defined as follows:

**Definition 1**: $\tilde{A}^*$ is denoted of a type-2 intuitionistic fuzzy set (IT2IFS), which is characterized by two type-2 intuitionistic membership functions, $\mu_{\tilde{A}}(x, u_1), \mu'_{\tilde{A}}(x, u_2)$ and two type-2 intuitionistic non-membership function $v_{\tilde{A}^*}(x, u_1), v'_{\tilde{A}^*}(x, u_2)$ where $x \in X$ and $u_1 \in J_x^1 \subseteq [0,1], u_2 \in J_x^2 \subseteq [0,1]$ i.e.

$$\tilde{A}^* = \left\{ \begin{array}{l} \left(\left(x,u_1\right), \mu_{\tilde{A}^*}\left(x,u_1\right), v_{\tilde{A}^*}\left(x,u_1\right)\right), \\ \left(\left(x,u_2\right), \mu'_{\tilde{A}^*}\left(x,u_2\right), v'_{\tilde{A}^*}\left(x,u_2\right)\right) \\ \left|\forall x \in X, \forall u_1 \in J_x^1 \subseteq [0,1]\right., \\ \forall u_2 \in J_x^2 \subseteq [0,1] \end{array} \right\} \quad (3)$$

where

$$\begin{array}{l} 0 \le \mu_{\tilde{A}^*}\left(x,u_1\right), \mu'_{\tilde{A}^*}\left(x,u_2\right), v_{\tilde{A}^*}\left(x,u_1\right), v'_{\tilde{A}^*}\left(x,u_2\right) \le 1 \\ 0 \le v_{\tilde{A}^*}\left(x,u_1\right) + \mu_{\tilde{A}^*}\left(x,u_1\right) \le 1 \\ 0 \le v'_{\tilde{A}^*}\left(x,u_2\right) + \mu'_{\tilde{A}^*}\left(x,u_2\right) \le 1 \end{array} \quad (4)$$

Type-2 intuitionistic fuzzy sets are called an IT2IFS, if the secondary membership function $\mu_{\tilde{A}}\left(x,u_1\right) = 1$ and $\mu'_{\tilde{A}^*}\left(x,u_2\right) = 1$ $\left(\forall u_1 \in J_x^1, u_2 \in J_x^2\right)$ i.e. an interval type-2 intuitinistic fuzzy set is defined as follows:

**Definition 2**: An IT2IFS $\tilde{A}^*$ is characterized by membership bounding functions $\bar{\mu}_{\tilde{A}^*}(x), \underline{\mu}_{\tilde{A}^*}(x)$ and non-membership bounding functions $\bar{v}_{\tilde{A}^*}(x), \underline{v}_{\tilde{A}^*}(x)$ where $x \in X$ in which

$$0 \le \bar{\mu}_{\tilde{A}}(x) + \underline{v}_{\tilde{A}}(x) \le 1 \quad (5)$$

$$0 \le \underline{\mu}_{\tilde{A}}(x) + \bar{v}_{\tilde{A}}(x) \le 1 \quad (6)$$

Through describing foot of uncertainty (FOU), an Interval type 2 intuitionistic fuzzy set can be:

$$\tilde{A} = \left\{ \begin{array}{l} x, \forall x \in X, \bar{\mu}_{\tilde{A}}(x), \underline{\mu}_{\tilde{A}}(x), \bar{v}_{\tilde{A}}(x), \underline{v}_{\tilde{A}}(x), \\ \forall \bar{\mu}_{\tilde{A}}(x), \underline{\mu}_{\tilde{A}}(x), \bar{v}_{\tilde{A}}(x), \underline{v}_{\tilde{A}}(x) \in [0,1] \end{array} \right\} \quad (7)$$

## III. Fuzzy C-Medoids Clustering Based on Interval Type-2 Intuitionistic Fuzzy Sets

For interval type-2 fuzzy c-means clustering, we have two objective function:

$$\begin{cases} J_{m_1}(U,Z) = \sum_{j=1}^{N} \sum_{i=1}^{K} \left(u_{ij}\right)^{m_1} d_{ij}^2 \\ J_{m_2}(U,Z) = \sum_{j=1}^{N} \sum_{i=1}^{K} \left(u_{ij}\right)^{m_2} d_{ij}^2 \end{cases} \quad (8)$$

Where $d_{ij} = \|x_j - z_i\|$ is the Euclidean distance between $x_j$ and medoid $z_i$, $K$ is the cluster number, $N$ is the number of members. $m_1$ and $m_2$ are optional constants. The upper and lower margins ($\bar{u}_{ij}$ and $\underset{-ij}{u}$) of membership are determined as follows:

$$\bar{u}_{ij} = \begin{cases} \frac{1}{\sum_{k=1}^{K} \left(d_{ij}/d_{kj}\right)^{2/(m_1-1)}} & if \frac{1}{\sum_{k=1}^{K} \left(d_{ij}/d_{kj}\right)} < \frac{1}{K} \\ \frac{1}{\sum_{k=1}^{K} \left(d_{ij}/d_{kj}\right)^{2/(m_2-1)}} & if \frac{1}{\sum_{k=1}^{K} \left(d_{ij}/d_{kj}\right)} \ge \frac{1}{K} \end{cases} \quad (9)$$

And

$$\underset{-ij}{u} = \begin{cases} \frac{1}{\sum_{k=1}^{K} \left(d_{ij}/d_{kj}\right)^{2/(m_1-1)}} & if \frac{1}{\sum_{k=1}^{K} \left(d_{ij}/d_{kj}\right)} \ge \frac{1}{K} \\ \frac{1}{\sum_{k=1}^{K} \left(d_{ij}/d_{kj}\right)^{2/(m_2-1)}} & if \frac{1}{\sum_{j=1}^{K} \left(d_{ij}/d_{kj}\right)} < \frac{1}{K} \end{cases} \quad (10)$$

where $i = \overline{1,K},\ j = \overline{1,N}$. Type 2 fuzzy sets and intuitionistic fuzzy sets have been applied to handle uncertainty, but both their process of uncertainty is different. If the handling of the uncertainty of type 2 fuzzy sets is based on an uncertain choice of related functions, then that of the intuitionistic fuzzy sets is based on the identification of membership functions, non-membership functions with the hesitance assessment function.

The aggregate membership functions for interval type-2 intuitionistic fuzzy sets are defined as follow:

$$\bar{\tilde{u}}_{ij}^* = \bar{u}_{ij} + \frac{\underline{\pi}_{\mathrm{ij}}}{2}; \underline{\tilde{u}}_{ij}^* = \underline{u}_{\mathrm{ij}} + \frac{\bar{\pi}_{ij}}{2} \quad (11)$$

To build the *IT2IFSs*, the choice of the interval membership functions $\underline{\tilde{u}}_{ij}^*$ and $\bar{\tilde{u}}_{ij}^*$ is conditioned by the interval hesitance degree $\underline{\pi}_j$ and $\bar{\pi}_j$ of the $j^{th}$ data with the clusters as follows:

$$\begin{array}{l} \underline{\pi}_j = \wedge \left(1 - \bar{\tilde{u}}_{1j}^*, 1 - \bar{\tilde{u}}_{2j}^*, \ldots, 1 - \bar{\tilde{u}}_{Kj}^*\right) \\ \bar{\pi}_j = \vee \left(1 - \underline{\pi}^*{}_{1j}, 1 - \underline{\pi}^*{}_{2j}, \ldots, 1 - \underline{\pi}^*{}_{Kj}\right) \end{array} \quad (12)$$

in which: Each $\left[\underline{\tilde{u}}_{ij}^*, \bar{\tilde{u}}_{ij}^*\right], (1 \le i \le K)$ is the membership functions of the $j^{th}$ data in the cluster $i$ and $K$ is the number of clusters. Since each data $j^{th}$ has membership interval as the upper $\bar{\tilde{u}}_{ij}^*$ and the lower $\underline{\tilde{u}}_{ij}^*$ in the $i^{th}$ cluster. So when it is as a centroid of the cluster (medoid) could be represented by the interval between $u^L$, and $u^R$:

$$u_{ij} = \left(u_i^R\left(x_j\right) + u_i^L\left(x_j\right)\right)/2,\ i = 1,\ldots,K, j = 1,\ldots,N \quad (13)$$

$u^L$, and $u^R$ are calculated according to Eq. (14) and Eq. (15) respectively.

where:

$$u_i^L = \frac{1}{M} \sum_{l=1}^{M} u_{il},\ u_{il} = \begin{cases} \bar{u}_i\left(x_j\right) & \text{if } x_{il} \text{ uses } \bar{u}_i\left(x_j\right) \text{ for } z_i^L \\ \underline{u}_i\left(x_j\right) & otherwise \end{cases} \quad (14)$$

$$u_i^R = \frac{1}{M} \sum_{l=1}^{M} u_{il},\ u_{il} = \begin{cases} \bar{u}_i\left(x_j\right) & \text{if } x_{il} \text{ uses } \bar{u}_i\left(x_j\right) \text{ for } z_i^R \\ \underline{u}_i\left(x_j\right) & otherwise \end{cases} \quad (15)$$

The method of determining $u_{il}$ is the same as described in Eq. (5).

Similar to the argument for calculating the objective function of the intuitionistic fuzzy cluster, the first term of the objective function in this cluster is:

$$J_1 = \sum_{i=1}^{K} \sum_{j=1}^{N} \tilde{u}_{ij}^{*m_1} d\left(x_j, z_j^*\right)^2 + \sum_{i=1}^{K} \sum_{j=1}^{N} \tilde{u}_{ij}^{*m_2} d\left(x_j, z_j^*\right)^2 \quad (16)$$

A second term of the objective function is added via the entropy of intuitionistic fuzzy (IFE):

$$J_2 = \frac{1}{N} \sum_{j=1}^{N} \pi_j \quad (17)$$

From Eqs. (16) and (17), the final objective function contains two terms that need to be minimum as follows:

$$\tilde{J} = J_1 + J_2 \quad (18)$$

The objective function consists of 2 parts that need to be optimized. $J_1$ to be optimized as shown in [9] by *FCM* and $J_2$ is the reducer, via Entropy [23].

The given parameters of the problem are the number of clusters $K\,(1 < K < N)$, the fuzzy coefficient $m\,(1 < m < +\infty)$ and the error threshold $\varepsilon$. Fuzzy C-medoids Clustering

Algorithm using Interval Type-2 Intuitionistic Fuzzy Sets (IT2IFCMdd) can be briefly described as follows:

**Algorithm 2: IT2IFCMdd**

**Input**: Database $X$, cluster number $K$

Fixed number of clusters $K$; Set $iter$ = 0; Select $S$ (the number of samples for which it has the highest membership value per a cluster)

Using **Procedure 1** to initialize medoids $Z = \{z_1, z_2, \ldots, z_K\}$ from $X^K$

**REPEAT**

*Step 1*. $iter = iter + 1$;

*Step 2*. Store current medoids: $Z^{old} = Z$;

*Step 3*. Using **Procedure 2** to get $\tilde{u}^*_{ij}$ and $\pi_j$

*Step 4*. Take a list of $S$ with top $\tilde{u}^*_{ij}$ scores in each cluster, then consider the top scores $P_i = X_{(p)i}, i = 1, 2, \ldots, K$ with the number of $S \times K$ medoids.

*Step 5*. Calculate and update $K$ medoids $Z = z_i, i = 1, .., K$:

$$\begin{array}{l} q = \arg\min_{x_k \in X_{(p)i}} \tilde{J} \\ z_i = x_q; Z \to update\{z_i\} \end{array} \tag{19}$$

**UNTIL** $\left\|Z^{old} - Z\right\|_f < \varepsilon || iter = MAX\_ITER$

Calculate the target function $\tilde{J}$ by Eq.(18)

Output $Z$, $\tilde{J}$, $\tilde{u}^*_{ij}$, $\pi_j$

The procedures to initialize medoids for clusters and compute $\tilde{u}^*_{ij}$ and $\pi_j$ values are performed as follows:

**Procedure 1. Initialize set of medoids for $K$ clusters**

*Step 1*. Random initialization $z_i$.

*Step 2*. Calculate the distance matrix $D$: $z_i$ to the $x_j$.

*Step 3*. Select from $x_j$ to get $z_{i+1}$ so that the distance $d(z_i, z_{j+1})$ is the second largest or largest.

*Step 4*. Calculate $Z = unique(Z)$ to make sure the elements of $Z$ are not the same

*Step 5*. Go back to Step 3 to get the desired amount $K$ of potential medoids.

**Procedure 2. Calculating $\tilde{u}^*_{ij}$ and $\pi_j$**

*Step 1*. Calculate $\bar{u}_{ij}$ and $\underline{u}_{ij}$.

*Step 2*. Calculate $u^L$, $u^R$ using Eq. (14) and Eq. (15), then $u_{ij}$ using Eq. (13)

*Step 3*. Calculate $\pi_j$:

$$\pi_j = \wedge(1 - u_{1j}, 1 - u_{2j}, \ldots, 1 - u_{Kj}) \tag{20}$$

*Step 4*. Calculate $\left[\underline{\tilde{u}}^*_{ij}, \bar{\tilde{u}}^*_{ij}\right]$

$$\bar{\tilde{u}}^*_{ij} = \underline{u}_{ij} + \frac{\pi_j}{2}; \underline{\tilde{u}}^*_{ij} = \bar{u}_{ij} - \frac{\pi_j}{2}. \tag{21}$$

*Step 5*. Calculate $\left[\underline{\pi}_j, \bar{\pi}_j\right]$ using Eq. (12)

*Step 6*. Compute $\tilde{u}^*_{ij}$ and $\pi_j$:

$$\tilde{u}^*_{ij} = (\bar{\tilde{u}}^*_{ij} + \underline{\tilde{u}}^*_{ij})/2; \pi_j = (\underline{\pi}_j + \bar{\pi}_j)/2 \tag{22}$$

The complexity of the FCMdd algorithm is $O(nCT_{max})$, where n is the number of input objects, $C$ is number of clusters. The computational complexity of IT2IFCMdd algorithm is $O(n^2)$.

## IV. Experiment

The proposed algorithm is implemented with $m = 2, m_1 = 1.5, m_2 = 3.5$, the number of iterations $L = 500$ and the error $\epsilon < 0.00001$. To evaluate the effectiveness of the proposed algorithm, the experimental paper on data clustering on a number of data sets was taken from the UCI machine learning library (https://archive.ics.uci.edu/ml/index.php) including Contraceptive Method Choice (Data1), Dermatology (Data2), Hayes-Roth (Data3), Lymphography (Data4), Zoo (Data5), and Balance Scale (Data6). Details of the experimental data sets are shown in Table I.

TABLE I
Experimental data

| **Dataset** | Number of Instances | Number of Attributes | Date Donated |
|---|---|---|---|
| Data1 | 1473 | 9 | 1997-07-07 |
| Data2 | 366 | 33 | 1998-01-01 |
| Data3 | 160 | 5 | 1989-03-01 |
| Data4 | 148 | 18 | 1988-11-01 |
| Data5 | 101 | 17 | 1990-05-15 |
| Data6 | 625 | 4 | 1994-04-22 |

The results were measured on the basis of several validity indexes to assess the performance of the algorithms. Indicators used include the Beni's index (XB), the Classification Entropy index (CE), the Bezdeks partition coefficient (PC) [21], [22]. The accuracy of clustering results by IT2IFCMdd algorithm is compared with FCMdd, IFCM and FCM algorithms. The better algorithm has smaller XB, CE and larger PC and the best results are marked bold. Furthermore, the paper also evaluates accuracy by calculating the correct classification ratio (Acc).

The detailed values of the indicators according to the IT2IFCMdd, FCMdd, IFCM and FCM algorithms are shown in Table II.

Table II shows that the proposed algorithm IT2IFCMdd gives better results than the three algorithms FCMdd, IFCM and FCM on most indicators. Only the CE index in the Hayes-Roth dataset gives the best results for the FCMdd algorithm. These results indicate that the IT2IFCMdd algorithm is able to significantly improve clustering results compared to the FCMdd, IFCM and FCM algorithms.

IT2IFCMdd algorithm gives the highest correct classification rate on all 6 data sets. The highest value reaches 97.048% on the Zoo dataset. While the lowest rate is 94.786% on Dermatology data set. The FCM algorithm gives the lowest accuracy of most indicators including the correct classification ratio.

## V. Conclusion

This paper presented the interval type-2 intuitionistic fuzzy c-medoid clustering algorithm by expanding fuzzy c-medoid clustering based on interval type-2 intuitionistic fuzzy set. The results showed that IT2IFCMdd algorithm can give better clustering results than the algorithms FCM, IFCM and FCMdd.

In the future, the paper will be developed based on the optimization algorithm to optimize the objective function. This

TABLE II
THE VALUE OF THE INDEXES ON THE ALGORITHMS

| Dataset | Indexes | FCM | IFCM | FCMdd | IT2IFCMdd |
|---|---|---|---|---|---|
| Data1 | PC | 0.3301 | 0.3298 | 0.3652 | **0.3921** |
| | CE | 0.4872 | 0.3716 | 0.2896 | **0.2875** |
| | XB | 4.8102 | 3.8261 | 3.8036 | **2.5663** |
| | Acc | 89.874% | 92.785% | 94.723% | **95.351%** |
| Data2 | PC | 0.1888 | 0.2075 | 0.2218 | **0.2277** |
| | CE | 0.5601 | 0.5623 | 0.4879 | **0.3861** |
| | XB | 3.4832 | 3.2751 | 2.5482 | **1.0027** |
| | Acc | 92.765% | 93.675% | 94.082% | **94.786%** |
| Data3 | PC | 0.2874 | 0.3872 | 0.3809 | **0.3862** |
| | CE | 0.6189 | 0.6805 | **0.5892** | 0.5982 |
| | XB | 2.3764 | 2.0993 | 2.1973 | **1.0118** |
| | Acc | 91.673% | 94.037% | 93.879% | **95.872%** |
| Data4 | PC | 0.2182 | 0.2287 | 0.2765 | **0.2833** |
| | CE | 0.7582 | 0.5983 | 0.5673 | **0.4885** |
| | XB | 1.1983 | 1.0845 | 0.8734 | **0.4371** |
| | Acc | 92.514% | 92.883% | 94.764% | **96.261%** |
| Data5 | PC | 0.5873 | 0.5629 | 0.6168 | **0.6585** |
| | CE | 0.9813 | 0.7209 | 0.6571 | **0.5826** |
| | XB | 0.3721 | 0.3673 | 0.2789 | **0.2400** |
| | Acc | 93.889% | 95.187% | 94.762% | **97.048%** |
| Data6 | PC | 0.4459 | 0.5462 | 0.6251 | **0.7271** |
| | CE | 0.8995 | 0.8752 | 0.6898 | **0.6824** |
| | XB | 3.0338 | 2.4875 | 1.6342 | **0.3728** |
| | Acc | 94.982% | 95.217% | 95.739% | **96.872%** |

is to improve the accuracy of clustering results as well as speed up computation when applied on large data sets.

## REFERENCES


[1] K. Atanassov (1986). Intuitionistic fuzzy sets, Fuzzy Sets and Systems, vol.20, pp. 87-96.

[2] K. T. Atanassov (2020). Interval-Valued Intuitionistic Fuzzy Sets, Studies in Fuzziness and Soft Computing, Springer Nature Switzerland AG 2020, Vol.388.

[3] M. D. Ansari, A. R. Mishra, and F. T. Ansari (2018). New divergence and entropy measures for intuitionistic fuzzy sets on edge detection, International Journal of Fuzzy Systems, vol.20, no.2, pp. 474-487.

[4] J. C. Bezdek (1981). Pattern Recognition with Fuzzy Objective Function Algorithms, New York: Plenum.

[5] D.S, Mai, L.T, Ngo, T.H, Le (2018). Advanced Semi-supervised Possibilistic Fuzzy C-means Clustering using Spatial-Spectral distance for Land-cover Classification. IEEE International Conference on Systems, Man, and Cybernetics, pp. 4375-4380.

[6] D.S, Mai, L.T, Ngo, T.H, Le (2018). Satellite Image Classification Based Spatial-Spectral Fuzzy Clustering Algorithm. ACIIDS 2018, Lecture Notes in Computer Science, vol 10752.

[7] T. Chaira (2011). A novel intuitionistic fuzzy C means clustering algorithm and its application to medical images, Applied Soft Computing, vol.11, no.2, pp. 1711-1717.

[8] P. D'Urso, J.M. Leski (2016). Fuzzy c-ordered medoids clustering for interval-valued data, Pattern Recognition, Volume 58, October 2016, Pages 49-67.

[9] Dzung D.N., Long T.N and Long T.P (2013). Interval Type-2 Fuzzy C-means Clustering using Intuitionistic Fuzzy Sets, Third World Congress on Information and Communication Technologies (WICT), IEEE, pp. 299-304.

[10] D.S, Mai, T.H, Dang, L.T, Ngo (2020). Optimization of interval type-2 fuzzy system using the PSO technique for predictive problems, Journal of Information and Telecommunication, pp.1-17.

[11] L. Hu and K. C. C. Chan (2016). Fuzzy Clustering in a Complex Network Based on Content Relevance and Link Structures, IEEE Transactions on Fuzzy Systems, vol.24, no.2, pp. 456-470.

[12] R. J. Hathaway and J. C. Bezdek (1993). Switching regression models and fuzzy clustering, IEEE Trans. Fuzzy Syst., vol.1, pp. 195–204.

[13] R. Krishnapuram and J. M. Keller (1993). A possibilistic approach to clustering, IEEE Trans. Fuzzy Syst., vol.1, pp. 98–110.

[14] R. Krishnapuram and J. M. Keller (1996). The possibilistic c-means algorithm: Insights and recommendations, IEEE Trans. Fuzzy Syst., vol.4, pp. 385–393.

[15] R. Krishnapuram, Anupam Joshi, Olfa Nasraoui, Liyu Yi (2001). Low-Complexity Fuzzy Relational Clustering Algorithms for Web Mining, IEEE Transactions On Fuzzy Systems, Vol.9, No 4.

[16] D. Molina, J. Poyatos, J.D. Ser, S. García, A. Hussain, F. Herrera, (2020). Comprehensive Taxonomies of Nature- and Bio-inspired Optimization: Inspiration versus Algorithmic Behavior, Critical Analysis and Recommendations. Cognitive Computation volume 12, pp.897–939.

[17] Pandey, Pooja , Virk, Ishpreet Singh. (2016). Comparison between Standard K-Mean Clustering and Improved K-Mean Clustering. International Journal of Computer Applications. vol.146. pp.39-42.

[18] Pinto, P. M.,Couto, P., Bustince, H., Barrenechea, E., Pagola, M. and Fernandez, J. (2013). Image segmentation using Atanassovs intuitionistic fuzzy sets, Expert Systems with Applications, vol.40, pp.15-26.

[19] Sahil, Sahil , Kumar, Tanuj. (2017). ST-intuitionistic fuzzy metric space with propertie". AIP Conference Proceedings. pp.18-60.

[20] D.S, Mai, L.T, Ngo, T.H, Le, H. Hagras (2021). A hybrid interval type-2 semi-supervised possibilistic fuzzy c-means clustering and particle swarm optimization for satellite image analysis, Information Sciences, Vol. 548, 2021, pp.398-422.

[21] W. Wang,Y. Zhang (2007). On fuzzy cluster validity indices, Fuzzy Sets and Systems vol.158, pp.2095-2117.

[22] Yongli Liu, Xiaoyang Zhang, Jingli Chen, and Hao Chao (2019), A Validity Index for Fuzzy Clustering Based on Bipartite Modularity Hindawi, Journal of Electrical and Computer Engineering. Vol. 2019, Article ID 2719617, 9 pages.

[23] Q. Zhang, S. Jiang (2010). Relationships between entropy and similarity measure of interval-valued intuitionistic fuzzy sets, Int. J. Intell. Syst. vol.25(no.11), pp.1121–1140.

[24] Zhao, Hua and Xu, Zeshui and Wang, Zhong (2013). Intuitionistic fuzzy clustering algorithm based on Boole matrix and association measure. International Journal of Information Technology, Decision Making.

[25] Velmurugan (2010). Computational complexity between k-Means and k-Medoids Clustering Algorithms for Normal and Uniform Distributions of Data Points. Journal of Computer Science. Vol.6. pp.363-368.