Annals of **F**uzzy **M**athematics and **I**nformatics
Volume 11, No. 1, (January 2016), pp. 97–107
ISSN: 2093–9310 (print version)
ISSN: 2287–6235 (electronic version)
http://www.afmi.or.kr

Clustering using similarity measure

Sharmistha Bhattacharya (Halder), Amarjit Chanda

Received 29 December 2014; Revised 28 February 2015; Accepted 22 July 2015

Abstract. Clustering plays an important role in data mining techniques for medical diagnosis. Clustering can be considered as the most important un-supervised learning technique. Most of the clustering methods of group data based on distance and few are based on similarity. The clustering algorithms classify gene expression data into clusters and the functionally related genes are grouped together in an efficient manner. The groupings are constructed such that the degree of relationship is strong among members of the same cluster and weak among members of different clusters. In this paper, we introduced a similarity measure of intuitionistic fuzzy sets(IFSs) and developed a clustering algorithm based on similarity measure of IFSs. Also an application of similarity measure between two IFSs in a decision making problem is illustrated and also shown that our proposed method and previously defined method are giving similar results using same example. The aim of this paper is to introduce a simple method of finding cluster(s) making use of Similarity Measure.

2010 AMS Classification: 03E72

Keywords: Fuzzy Set, Intuitionistic Fuzzy Set, Similarity Measure, Clustering Algorithm

Corresponding Author: Amarjit Chanda (amarjit.chanda90@outlook.com)

1. Introduction

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics.
Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Popular notions of

clusters include groups with small distances among the cluster members, dense areas of the data space, intervals or particular statistical distributions. Clustering can therefore be formulated as a multi-objective optimization problem. The appropriate clustering algorithm and parameter settings (including values such as the distance function to use, a density threshold or the number of expected clusters) depend on the individual data set and intended use of the results. Cluster analysis as such is not an automatic task, but an iterative process of knowledge discovery or interactive multi-objective optimization that involves trial and failure. It will often be necessary to modify data preprocessing and model parameters until the result achieves the desired properties.

Similarity measure between two fuzzy sets(interval-valued fuzzy sets, intuitionistic fuzzy sets, interval-valued intuitionistic fuzzy sets) have been defined by many authors [2, 3, 5, 8].There are several techniques for defining similarity measure in such cases. Some of them are based on distances and some others are based on matching function. There are techniques based on set-theoretic approach also. Some properties are common to these measures and some are not, which influence the choice of the measure to be used in several applications.In [4] , seven similarity measures of fuzzy soft sets are introduced, which are based on the normalized Hamming distance, the normalized Euclidean distance, the generalized normalized distance, the Type-2 generalized normalized distance, the Type-2 normalized Euclidean distance, the Hausdorff distance and the Chebyshev distance. One of the significant differences between similarity measure based on matching function S and similarity measure $S^{'}$ based on distance is that if A $\bigcap$ B $=\varnothing$, then S(A,B) = 0 but $S^{'}$ (A,B) may not be equal to zero, where A and B are two fuzzy sets. But it is easier to calculate the intermediate distance between two fuzzy sets or soft sets. We also developed a new Similarity measure denoted by $SM_p(A, B)$, where A and B are IFSs.
Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics.
Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Popular notions of clusters include groups with small distances among the cluster members, dense areas of the data space, intervals or particular statistical distributions.
We have extended these concepts of similarity measure of IFS in clustering. The aim of this paper is to introduce the clustering using Similarity Measure and show its application.

2. Preliminaries

In this section we briefly review some basic definitions which will be used in the rest of the paper.

Definition 2.1 ([14]). Let us consider a non empty collection of objects denoted by X. Then a fuzzy set (FS for short)α in X is a set of ordered pairs having the form
$\alpha = \{(x, \mu_\alpha(x) : x \epsilon X\}$
where the function μ_α : X$\longrightarrow$[0,1] is called the membership function or grade of membership (also degree of compatibility or degree of truth) of X in α. The interval M = [0,1] is called membership space.

Definition 2.2 ([1]). The intuitionistic fuzzy sets (IFS) introduced by Atanassov(1983) are an extension of the theory of fuzzy sets created by L.A Zadeh as an adequate mathematical description of imprecision and uncertainty in nature.
Let X be a non empty set. Then an intuitionistic fuzzy set (IFS) A is a set having the form
A=$\{(x, \mu_A(x), \gamma_A(x) : x \epsilon X\}$
where the functions μ_A : X$\longrightarrow$[0,1] and γ_A : X$\longrightarrow$[0,1] represents the degree of membership and the degree of non-membership respectively of each element $x \epsilon X$ and $\pi_A(x) = 1 - (\mu_A(x) + \gamma_A(x))$ is called the hesitancy degree.

Definition 2.3 ([8]). Distance Measures : Let $X = x_1, x_2, x_3, ..., x_n$ be a discrete universe of discourse. Consider that the elements $x_i (i = 1, 2, 3, , ..., n)$, with $w_i \geq 0$, i=1,2,3,...,n and $\sum_{i=1}^{n} w_i = 1$, then we have for any two IFSs A and B ,
(1) The weighted Hamming distance :
$d(A, B) = \frac{1}{2} \sum_{i=1}^{n} w_i (|\mu_A(x_i) - \mu_B(x_i)| + |\gamma_A(x_i) - \gamma_B(x_i)| + |\pi_A(x_i) - \pi_B(x_i)|)$.
(2) The normalized Hamming distance :
$d(A, B) = \frac{1}{2n} \sum_{i=1}^{n} (|\mu_A(x_i) - \mu_B(x_i)| + |\gamma_A(x_i) - \gamma_B(x_i)| + |\pi_A(x_i) - \pi_B(x_i)|)$, where $w = (\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n})$.
(3) The weighted Euclidean distance :
$d^2(A, B) = \frac{1}{2} \sum_{i=1}^{n} w_i ((\mu_A(x_i) - \mu_B(x_i))^2 + (\gamma_A(x_i) - \gamma_B(x_i))^2 + (\pi_A(x_i) - \pi_B(x_i))^2)$.
(4) The normalized Euclidean distance :
$d^2(A, B) = \frac{1}{2n} \sum_{i=1}^{n} ((\mu_A(x_i) - \mu_B(x_i))^2 + (\gamma_A(x_i) - \gamma_B(x_i))^2 + (\pi_A(x_i) - \pi_B(x_i))^2)$
where $w = (\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n})$.

Definition 2.4. Let A and B be two IFS-sets over U. Then using the Hamming distance(1) , Similarity measure of A and B is defined as follows :

$$SM(A, B) = \frac{1}{1 + d(A, B)}. \tag{5}$$

3. Clustering algorithm

For any two IFS-sets A and B over U we defined the Similarity Measure as

$$SM_p(A, B) = \frac{1}{1 + \sqrt{d(A, B)}}. \tag{6}$$

Given a collection of m IFSs $A_j (j = 1, ..., m)$. In the first stage each of the m IFSs $A_j (j = 1, ..., m)$ is considered as a unique cluster. Also we considered a ideal/expert IFSs say X. The IFSs A_j(j=1,...,m) and X are then compared among themselves by weighted Hamming/Euclidean or by normalized Hamming/Euclidean distance. Now we find the Similarity Measure(using (6)) between the m IFSs A_j(j=1,...,m) and X. Now the IFSs A_j(j=1,...,m) and X having similar(or nearly similar) Similarity Measure forms the desired cluster(s).

The main steps for this algorithm are as follows :
Step 1. Consider each of the m IFSs A_j(j=1,...,m) as a unique cluster.
Step 2. Enter the number of variable and value of the weights.
Step 3. Enter the membership, non-membership and hesitancy of the expert i.e. X.
Step 4. Enter the membership, non-membership and hesitancy of the m IFSs A_j(j=1,...,m) .
Step 5. Calculate the distance between X and m IFSs A_j(j=1,...,m) using (1)
Step 6. Calculate the Similarity Measure between X and m IFSs A_j(j=1,...,m) using (6) .

Similar algorithm will be followed for the measure defined in (5) .

Example 3.1. Suppose for the case of CANCER the symptoms are x_i , i = 1,2,...,8 where x_1 = lump, x_2 = body pain, x_3 = soreness, x_4 = headache, x_5 = loss of appetite, x_6 = weight loss, x_7 = chest pain, x_8 = fever . And we have 8 patients say A_1,A_2,A_3,A_4,A_5,A_6,A_7,A_8. Let the expert set is X , where

$X = \{(x_1, 0.63, 0.25),(x_2, 0.72, 0.19),(x_3, 0.84, 0.10),(x_4, 0.51, 0.36),(x_5, 0.39, 0.42),$
$(x_6, 0.50, 0.50),(x_7, 0.48, 0.30),(x_8, 0.79, 0.11)\}$

$A_1 = \{(x_1, 0.20, 0.50),(x_2, 0.10, 0.80),(x_3, 0.50, 0.30),(x_4, 0.90, 0.00),(x_5, 0.40, 0.35),$
$(x_6, 0.10, 0.90),(x_7, 0.30, 0.50),(x_8, 1.00, 0.00)\}$

$A_2 = \{(x_1, 0.50, 0.40),(x_2, 0.60, 0.15),(x_3, 1.00, 0.00),(x_4, 0.15, 0.65),(x_5, 0.00, 0.80),$
$(x_6, 0.70, 0.15),(x_7, 0.50, 0.30),(x_8, 0.65, 0.20)\}$

$A_3 = \{(x_1, 0.45, 0.35),(x_2, 0.60, 0.30),(x_3, 0.90, 0.00),(x_4, 0.10, 0.80),(x_5, 0.20, 0.70),$
$(x_6, 0.60, 0.20),(x_7, 0.15, 0.80),(x_8, 0.20, 0.65)\}$

$A_4 = \{(x_1, 1.00, 0.00),(x_2, 1.00, 0.00),(x_3, 0.85, 0.10),(x_4, 0.75, 0.15),(x_5, 0.20, 0.80),$
$(x_6, 0.15, 0.85),(x_7, 0.10, 0.70),(x_8, 0.30, 0.70)\}$

$A_5 = \{(x_1, 0.90, 0.00),(x_2, 0.90, 0.10),(x_3, 0.80, 0.10),(x_4, 0.70, 0.20),(x_5, 0.50, 0.15),$
$(x_6, 0.30, 0.65),(x_7, 0.15, 0.75),(x_8, 0.40, 0.30)\}$

$A_6 = \{(x_1, 0.30, 0.50),(x_2, 0.30, 0.60),(x_3, 0.40, 0.60),(x_4, 0.50, 0.50),(x_5, 0.70, 0.30),$
$(x_6, 0.10, 0.90),(x_7, 0.20, 0.70),(x_8, 0.30, 0.30)\}$

$A_7 = \{(x_1, 0.35, 0.45),(x_2, 0.20, 0.70),(x_3, 0.40, 0.10),(x_4, 0.50, 0.10),(x_5, 1.00, 0.00),$
$(x_6, 0.15, 0.25),(x_7, 0.10, 0.30),(x_8, 0.65, 0.15)\}$

$A_8 = \{(x_1, 1.00, 0.00),(x_2, 0.10, 0.20),(x_3, 0.40, 0.55),(x_4, 0.25, 0.05),(x_5, 0.15, 0.43),$
$(x_6, 0.02, 0.10),(x_7, 0.65, 0.06),(x_8, 0.02, 0.01)\}$

Now for finding the distance we consider the weight

$w = \{0.15, 0.10, 0.12, 0.15, 0.10, 0.13, 0.14, 0.11\}$

Step 1 In the first stage each IFSs $A_j(j = 1, ..., 8)$ is considered as a unique clusters i.e
$\{A_1\}, \{A_2\}, \{A_3\}, \{A_4\}, \{A_5\}, \{A_6\}, \{A_7\}, \{A_8\}$

Step 2

$d(X, A_1) = 0.3359$; $d(X, A_2) = 0.2144$
$d(X, A_3) = 0.3237$; $d(X, A_4) = 0.3251$
$d(X, A_5) = 0.2507$; $d(X, A_6) = 0.36315$
$d(X, A_7) = 0.3949$; $d(X, A_8) = 0.5247$

Step 3

$SM(X, A_1) = 0.75$; $SM(X, A_2) = 0.82$
$SM(X, A_3) = 0.76$; $SM(X, A_4) = 0.75$
$SM(X, A_5) = 0.80$; $SM(X, A_6) = 0.73$
$SM(X, A_7) = 0.72$; $SM(X, A_8) = 0.66$

Now,we take intervals [0,0.69], [0.70,0.79] and [0.80,1] which denotes cluster 1, cluster 2 and cluster 3 respectively. Now by some expert knowledge we can conclude that if $SM(X, A_i)$,i=1,2,...,n lies in cluster 1 then the patient has no possibility of having Cancer, if in cluster 2 then the patient has possibility of having Cancer, but if in cluster 3 then the patient is surely suffering from Cancer. We find only A_8 lies in the first cluster ; A_1, A_3, A_4, A_6, A_7 lies in 2nd one and A_2, A_5 lies in the 3rd cluster. Thus according to our proposed method A_2 and A_5 have more possibility of having cancer(FIGURE 1).

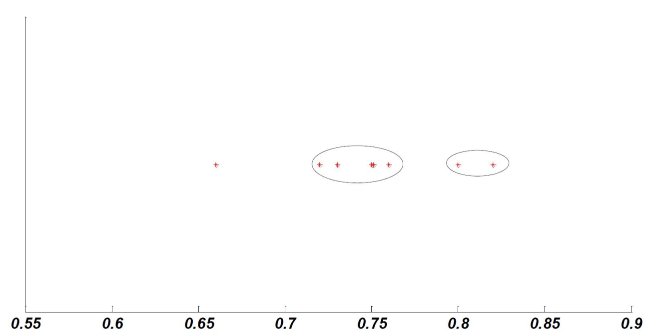

FIGURE 1.

In this process of clustering, the number of clusters can be obtained according to the problem to be solved.

Similarity Measure for our proposed definition :

$SM_p(X, A_1) = 0.63$;$SM_p(X, A_2) = 0.68$
$SM_p(X, A_3) = 0.64$;$SM_p(X, A_4) = 0.637$
$SM_p(X, A_5) = 0.67$;$SM_p(X, A_6) = 0.62$
$SM_p(X, A_7) = 0.61$;$SM_p(X, A_8) = 0.58$

Now,we take intervals [0.55,0.59], [0.60,0.65] and [0.66,1] which denotes cluster 1, cluster 2 and cluster 3 respectively. Now by some expert knowledge we can conclude that if $SM_p(X, A_i)$,i=1,2,...,n lies in cluster 1 then the patient has no possibility of having Cancer, if in cluster 2 then the patient has possibility of having Cancer, but if in cluster 3 then the patient is surely suffering from Cancer. We find only A_8 lies in the first cluster ; A_1, A_3, A_4, A_6, A_7 lies in 2nd one and A_2, A_5 lies in the 3rd cluster. Thus according to our proposed method A_2 and A_5 have more possibility of having cancer(FIGURE 2).

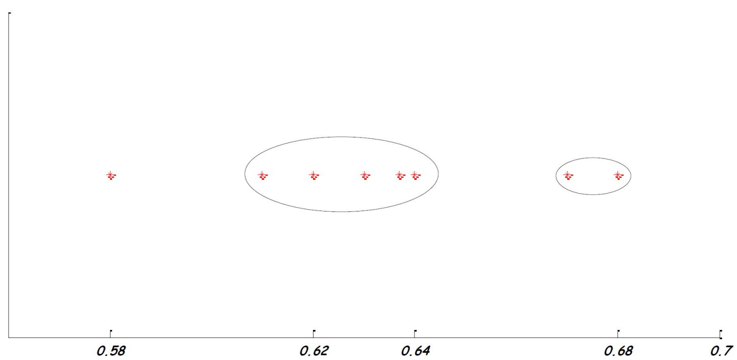

FIGURE 2.

Example 3.2. Let us take another example of CHIKUNGUNYA where the expert is X and the patients is A_1,A_2,A_3,A_4,A_5,A_6,A_7,A_8 having symptoms x_i , i = 1,2,...,8 where x_1 = fever, x_2 = joint pain, x_3 = headache, x_4 = rash, x_5 = muscle pain, x_6 = nausea, x_7 = back pain, x_8 = fatigue . Let

$X = \{(x_1, 0.50, 0.40), (x_2, 0.60, 0.15), (x_3, 1.00, 0.00), (x_4, 0.15, 0.65), (x_5, 0.00, 0.80), (x_6, 0.70, 0.15), (x_7, 0.50, 0.30), (x_8, 0.65, 0.20)\}$

$A_1 = \{(x_1, 0.20, 0.50), (x_2, 0.10, 0.80), (x_3, 0.50, 0.30), (x_4, 0.90, 0.00), (x_5, 0.40, 0.35), (x_6, 0.70, 0.15), (x_7, 0.50, 0.30), (x_8, 0.65, 0.20)\}$

$A_2 = \{(x_1, 0.45, 0.35), (x_2, 0.60, 0.30), (x_3, 0.90, 0.00), (x_4, 0.10, 0.80), (x_5, 0.20, 0.70), (x_6, 0.60, 0.20), (x_7, 0.15, 0.80), (x_8, 0.20, 0.65)\}$

$A_3 = \{(x_1, 1.00, 0.00), (x_2, 1.00, 0.00), (x_3, 0.85, 0.10), (x_4, 0.75, 0.15), (x_5, 0.20, 0.80),$

$(x_6, 0.15, 0.85), (x_7, 0.10, 0.70), (x_8, 0.30, 0.75)\}$

$A_4 = \{(x_1, 0.90, 0.00), (x_2, 0.90, 0.10), (x_3, 0.80, 0.10), (x_4, 0.70, 0.20), (x_5, 0.50, 0.15),$
$(x_6, 0.30, 0.65), (x_7, 0.15, 0.75), (x_8, 0.40, 0.30)\}$

$A_5 = \{(x_1, 0.50, 0.20), (x_2, 0.70, 0.00), (x_3, 0.10, 0.60), (x_4, 0.20, 0.40), (x_5, 0.70, 0.10),$
$(x_6, 0.60, 0.30), (x_7, 0.30, 0.50), (x_8, 0.80, 0.10)\}$

$A_6 = \{(x_1, 0.70, 0.20), (x_2, 0.40, 0.20), (x_3, 0.50, 0.20), (x_4, 0.80, 0.10), (x_5, 0.50, 0.40),$
$(x_6, 0.25, 0.05), (x_7, 0.15, 0.25), (x_8, 0.65, 0.15)\}$

$A_7 = \{(x_1, 0.90, 0.10), (x_2, 0.30, 0.60), (x_3, 0.70, 0.10), (x_4, 0.50, 0.10), (x_5, 0.10, 0.10),$
$(x_6, 0.20, 0.70), (x_7, 0.80, 0.10), (x_8, 0.10, 0.70)\}$

$A_8 = \{(x_1, 0.35, 0.45), (x_2, 0.20, 0.70), (x_3, 0.40, 0.10), (x_4, 0.50, 0.01), (x_5, 0.10, 0.00),$
$(x_6, 0.10, 0.30), (x_7, 0.65, 0.15), (x_8, 0.30, 0.30)\}$

Step 1 In the first stage each IFSs $A_j(j = 1, ..., 8)$ is considered as a unique clusters i.e
$\{A_1\}, \{A_2\}, \{A_3\}, \{A_4\}, \{A_5\}, \{A_6\}, \{A_7\}, \{A_8\}$

Step 2

$d(X, A_1) = 0.3275$; $d(X, A_2) = 0.2170$
$d(X, A_3) = 0.4450$; $d(X, A_4) = 0.4170$
$d(X, A_5) = 0.3245$; $d(X, A_6) = 0.3905$
$d(X, A_7) = 0.4675$; $d(X, A_8) = 0.4630$

Step 3

$SM(X, A_1) = 0.753$; $SM(X, A_2) = 0.822$
$SM(X, A_3) = 0.69$; $SM(X, A_4) = 0.71$
$SM(X, A_5) = 0.75$; $SM(X, A_6) = 0.72$
$SM(X, A_7) = 0.681$; $SM(X, A_8) = 0.683$

Now,we take intervals [0,0.70], [0.71,0.80] and [0.81,1] which denotes cluster 1, cluster 2 and cluster 3 respectively. Now by some expert knowledge we can conclude that if $SM(X, A_i)$,i=1,2,...,n lies in cluster 1 then the patient has no possibility of having Chikungunya, if in cluster 2 then the patient may have the possibility of having Chikungunya, but if in cluster 3 then the patient is surely suffering from Chikungunya. We find only A_3, A_7, A_8 lies in the cluster 1 ; A_1, A_4, A_5, A_6 lies in 2nd one and A_2 lies in the cluster 3. Thus A_2 have more possibility of having Chikungunya.(FIGURE 3)

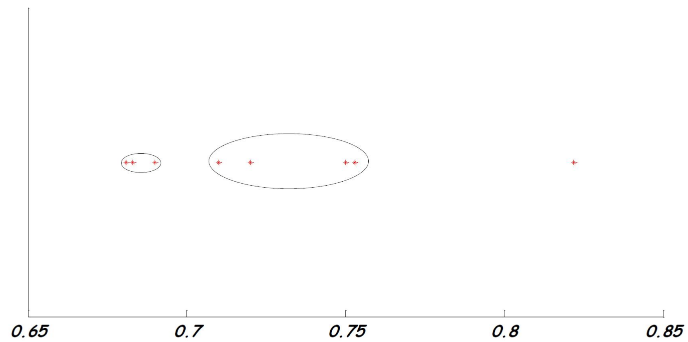

FIGURE 3.

Our proposed method

Step 1 In the first stage each IFSs $A_j(j = 1, ..., 8)$ is considered as a unique clusters i.e

$\{A_1\}, \{A_2\}, \{A_3\}, \{A_4\}, \{A_5\}, \{A_6\}, \{A_7\}, \{A_8\}$

Step 2

$\sqrt{d(X, A_1)} = 0.5723$; $\sqrt{d(X, A_2)} = 0.4658$
$\sqrt{d(X, A_3)} = 0.6671$; $\sqrt{d(X, A_4)} = 0.6458$
$\sqrt{d(X, A_5)} = 0.5696$; $\sqrt{d(X, A_6)} = 0.6249$
$\sqrt{d(X, A_7)} = 0.6837$; $\sqrt{d(X, A_8)} = 0.6804$

Step 3

$SM_p(X, A_1) = 0.636$;$SM_p(X, A_2) = 0.682$
$SM_p(X, A_3) = 0.599$;$SM_p(X, A_4) = 0.61$
$SM_p(X, A_5) = 0.637$;$SM_p(X, A_6) = 0.62$
$SM_p(X, A_7) = 0.594$;$SM_p(X, A_8) = 0.595$

Now,we take intervals [0,0.60], [0.61,0.65] and [0.66,1] which denotes cluster 1, cluster 2 and cluster 3 respectively. Now by some expert knowledge we can conclude that if $SM_p(X, A_i)$,i=1,2,...,n lies in cluster 1 then the patient has no possibility of having Chikungunya, if in cluster 2 then the patient has possibility of having Chikungunya, but if in cluster 3 then the patient is surely suffering from Chikungunya. Now, we find only A_3, A_7, A_8 lies in the cluster 1 ; A_1, A_4, A_5, A_6 lies in 2nd one and A_2 lies in the cluster 3. Thus,according to our proposed method A_2 have more possibility of having Chikungunya than others.(FIGURE 4)

4. Conclusions and comparisons

In this paper, we have defined four types of distances between two IFS-sets and proposed Similarity Measures of two IFS-sets. Then, we construct a decision making

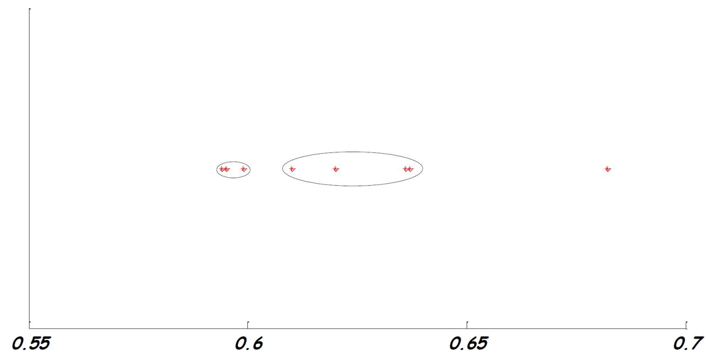

Figure 4.

method based on the similarity measures. Finally two simple examples are given to show the possibility of using this method by using Hamming distance for diagnosis of diseases having some visible symptoms. In these examples, if we use the other distances, we can obtain similar result. This method can be applied to problems that contain uncertainty such as problems in social, economic systems, pattern recognition, medical diagnosis, game theory, coding theory and so on. In the future we will investigate the activity patterns (e.g. grocery shopping paths, travel behavior paths, tourist behavior paths, skiers paths, pedestrian activity paths, web log paths, web content paths, eye tracking paths, mouse tracking paths and so on) in which there are three main issues to address: the measurement of similarity between activity patterns, the detection of natural clusters of these patterns, and how to take into account the order in which activities are executed, by using IFSs and Similarity Measure. In [11] the authors defined the concepts of association matrix and equivalent association matrix, and introduce some methods for calculating the association coefficients of IFSs. Then, they propose a clustering algorithm for IFSs. The algorithm uses the association coefficients of IFSs to construct an association matrix, and utilizes a procedure to transform it into an equivalent association matrix. In [16] the authors investigate graph theory-based clustering techniques for Atanassov's Intuitionistic fuzzy sets (A-IFSs) and interval-valued intuitionistic fuzzy sets (IVIFSs). In [15] they propose a novel hesitant fuzzy agglomerative hierarchical clustering algorithm for HFSs. The algorithm considers each of the given HFSs as a unique cluster in the first stage, and then compares each pair of the HFSs by utilising the weighted Hamming distance or the weighted Euclidean distance. The two clusters with smaller distance are jointed. The procedure is then repeated time and again until the desirable number of clusters is achieved. In [10] authors propose a technique for clustering data with intuitionistic fuzzy information. They first define two new intuitionistic fuzzy similarity measures, and then use it to construct an intuitionistic fuzzy similarity measure matrix, by which we present a spectral algorithm to cluster intuitionistic fuzzy information. In [6] the authors investigate the technique for clustering objects with intuitionistic fuzzy information. They first propose a formula to derive the intuitionistic fuzzy similarity degree between two intuitionistic fuzzy sets and develop an approach to constructing an intuitionistic fuzzy similarity matrix. In [7] they investigate the technique for clustering analysis under intuitionistic fuzzy environment. They first develop an intuitionistic fuzzy implication operator and extend the Lukasiewicz implication operator to intuitionistic fuzzy environment, and then define an intuitionistic fuzzy triangle product and an intuitionistic fuzzy square product. In [17] they develop a measure for calculating the association coefficient between Atanassov's intuitionistic fuzzy sets (A-IFSs), and show its desirable axiomatic properties. Then we present an algorithm for clustering A-IFSs. The algorithm first utilizes the association coefficient of A-IFSs to construct an association matrix, and then calculates the $\alpha-$ cutting matrix of the association matrix no matter whether it is an equivalent matrix or not. After that, the $\alpha-$cutting matrix is used to cluster A-IFSs. In [9] an algorithm is introduced for clustering IFSs, which is based on the traditional hierarchical clustering procedure, the intuitionistic fuzzy aggregation operator, and the basic distance measures between IFSs: the Hamming distance, normalized Hamming, weighted Hamming, the Euclidean distance, the normalized Euclidean distance, and the weighted Euclidean distance. Subsequently, the algorithm is extended for clustering IVIFSs. In [13] an intuitionistic fuzzy C-means algorithm to cluster IFSs is developed. In each stage of the intuitionistic fuzzy C-means method the seeds are modified, and for each IFS a membership degree to each of the clusters is estimated. In the end of the algorithm, all the given IFSs, FSs are clustered according to the estimated membership degrees. In [12] an intuitionistic fuzzy vector, the inner and outer products of intuitionistic fuzzy vectors, and study their properties. They put forward a new method of constructing intuitionistic fuzzy similarity matrix. Based on the orthogonal of intuitionistic fuzzy vectors, proposed an orthogonal algorithm for clustering intuitionistic fuzzy information. Whereas in our paper we introduced a similarity measure of intuitionistic fuzzy sets(IFSs) and developed a clustering algorithm based on similarity measure of IFSs. Also an application of similarity measure between two IFSs in a decision making problem is illustrated.

Acknowledgements. The authors are thankful to the Reviewers for their valuable suggestion to improve the paper.

References

[1] K. Atanassov, Intuitionistic fuzzy sets,Fuzzy Sets and Systems 20 (1986) 87–96.
[2] K. Atanassov and G. Gargov, Interval valued Intuitionistic fuzzy sets, Fuzzy Sets and Systems 31 (1989) 343–349.
[3] S. M. Chen, Measures of similarity between vague sets, Fuzzy Sets and Systems 74 (1995) 217–223.
[4] Qinrong Feng and Weinan Zheng, New similarity measures of fuzzy soft sets based on distance measures, Ann. Fuzzy Math. Inform. 7 (4) (2014) 669–686.
[5] Pinaki Majumdar and S. K. Samanta, Similarity Measure Of Soft Sets, New Math. Nat. Comput. 4 (1) (2008) 1–12.
[6] Z. Wang, Z. S. Xu, S. S. Liu and J. Tang, A netting clustering analysis method under intuitionistic fuzzy environment, Appl. Soft Comput. 11 (2011) 5558–5564.
[7] Z. Wang, Z. S. Xu, S. S. Liu and Z. Q. Yao, Direct clustering analysis based on intuitionistic fuzzy implication, Appl. Soft Comput. 23 (2014) 1–8.

[8] Zeshu Xu, Some Similarity Measures of intuitionistic fuzzy sets and their applications to multiple attribute decision making, Fuzzy Optim. Decis. Mak. 6 (2007) 109–121.
[9] Zeshu Xu, Intuitionistic fuzzy hierarchical clustering algorithms, Journal of Systems Engineering and Electronics 20 (1) (2009) 90–97.
[10] D. W. Xu, Z. S. Xu, S. S. Liu and H. Zhao, A spectral clustering algorithm based on intuitionistic fuzzy information, Knowledge-Based Systems 53 (2013) 20–26.
[11] Z. S. Xu, J. Chen and J. J. Wu, Clustering algorithm for intuitionistic fuzzy sets, Inform. Sci. 178 (2008) 3775–3790.
[12] Z. S. Xu, J. Tang and S. S. Liu, An orthogonal algorithm for clustering intuitionistic fuzzy information, An International Interdisciplinary Journal 14 (2011) 65–78.
[13] Z. S. Xu and J. J. Wu, Intuitionistic fuzzy c-means clustering algorithms, Journal of Systems Engineering and Electronics 21 (2010) 580–590.
[14] L.A. Zadeh, Fuzzy set, Information and Control 8 (1965) 338–353.
[15] X. L. Zhang and Z. S. Xu, Hesitant fuzzy agglomerative hierarchical clustering algorithms, Internat. J. Systems Sci. 46 (2015) 562–576.
[16] H. Zhao, Z. S. Xu, S. S. Liu and Z. Wang, Intuitionistic fuzzy MST clustering algorithms, Computers and Industrial Engineering 62 (2012) 1130–1140.
[17] H. Zhao, Z. S. Xu and Z. Wang, Intuitionistic fuzzy clustering algorithm based on Boole matrix and association measure, International Journal of Information Technology and Decision Making 12 (2013) 95–118.

Sharmistha bhattacharya Halder (halder_731@rediffmail.com)
Department of Mathematics, Tripura University-799022, Tripura, India

Amarjit Chanda (amarjit.chanda90@outlook.com)
Department of Mathematics, Tripura University-799022, Tripura, India