On intuitionistic fuzzy clustering for its application to privacy

Vicenç Torra, Sadaaki Miyamoto, Yasunori Endo and Josep Domingo-Ferrer

***Abstract*— Motivated by our research on specific information loss measures (in privacy preserving data mining) and our need to compare fuzzy clusters, we proposed in a recent paper a definition for intuitionistic fuzzy partitions. We showed how to define them in the framework of fuzzy clustering. That is, we introduced a method to define intuitionistic fuzzy partitions from the results of fuzzy clustering.**

In this paper we further study such intuitionistic fuzzy partitions and we extend our previous results with other types of fuzzy clustering algorithms.

I. Introduction

CLUSTERING and fuzzy clustering [8], [13] are well established areas, and clustering algorithms are standard tools in unsupervised machine learning and statistical learning [7]. Their goal is to divide the data into meaningful clusters. That is, similar objects should be put in the same clusters and dissimilar objects should be left into different clusters.

Different methods have been developed in fuzzy clustering based on different assumptions on the data and on the different properties that the resulting clusters should satisfy. Among the differences, we might underline the case of unstructured and structured clusters. That is, some methods as c-means and fuzzy c-means construct a set of *independent* clusters (no formal relationships exist between such clusters) and others, as single linkage, construct a set of clusters with some clear relationships between them (e.g. clusters define a dendrogram).

In the case of unstructured clusters, clusters usually define a partition of the domain. Differences on the methods correspond to the properties of such partitions. For example, the c-means algorithm builds a crisp partition and the fuzzy c-means method defines a fuzzy partition.

[19] introduced a new approach where the partition corresponds to an intuitionistic fuzzy partition. Then, [19] defined a method, based on the fuzzy c-means algorithm, to construct an intuitionistic fuzzy partition from the results of several clustering methods. In such an approach, the intuitionistic fuzzy partition permits us to cope with the uncertainty present in the different executions of the same clustering algorithm.

In this paper we further study this aproach. We discuss how to extend our previous definition for other fuzzy clustering algorithms. In particular, we consider the entropy based fuzzy c means and the fuzzy c means with tolerance.

The structure of the paper is as follows. In Section II we present a motivation of our work that is focused on the need to compare the results of different fuzzy clustering algorithms in the field of data privacy. Then, in Section III, we review a few concepts related to fuzzy clustering and intuitionistic fuzzy partitions. Then, in Section IV we present our approach for intuitionistic fuzzy clustering and our new results. The paper finishes with some conclusions and some lines for future research.

II. Motivation

Our interest in intuitionistic fuzzy clustering is rooted in an application in privacy preserving data mining. In that setting, we have been faced with the problem of comparing the results of two clustering methods for the same set of individuals.

In fact, our final goal is to define specific information loss measures to evaluate the degree of distortion of a particular data protection method. That is, given a perturbative protection method that introduces *noise* to the data we need to know in what extent the noise *destroys* the original structure of the data.

To evaluate the extent in which the data is distorted, several measures have been defined. One alternative is to use specific information loss measures. That is, measures that evaluate the extent of the perturbation for a specific use. In our case, the use we are considering is clustering. Therefore, we need to compare the clustering results obtained with the original data and the ones obtained with the perturbated data. Naturally, when no perturbation or only a small perturbation is applied, the clustering results of both sets should be almost the same. In contrast, when a large perturbation is applied, the results should be rather divergent. We have faced this problem in [10], [20].

In the case of clustering algorithms that build crisp clusters (and that such crisp clusters define a partition of the data set), there exist a few indices and distances that can be applied. This is the case, for example, of the Rand, adjusted Rand, Jaccard and Wallace indexes and the Mantaras distance [12]. Nevertheless, there are not so established methods for comparing fuzzy sets.

In [10], [11], we used α-cuts and, then, we computed an absolute distance between the resulting memberships. In [20], we have proposed two different distances, one based on the comparison of the cluster centers, and the other comparing membership functions once a matching between cluster centers has been defined.

Vicenç Torra is with the IIIA, Institut d'Investigació en Intel·ligència Artificial - CSIC; Campus UAB s/n; 08193 Bellaterra (Catalonia, Spain) (email: vtorra@iiia.csic.es). Sadaaki Miyamoto and Yasunori Endo are with the Department of Risk Engineering; School of Systems and Information Engineering; University of Tsukuba; Ibaraki 305-8573; Japan (email: {miyamoto, endo}@risk.tsukuba.ac.jp). Josep Domingo-Ferrer is with the UNESCO Chair in Data Privacy; Rovira i Virgili University; Dept. of Computer Engineering and Maths; Av. Països Catalans 26, E-43007 Tarragona (Catalonia, Spain) (email: josep.domingo@urv.cat)

978-1-4244-1819-0/08/$25.00©2008 IEEE

Nevertheless, such approaches present an important shortcomming due to the underlying clustering algorithm. Fuzzy clustering methods are typically formalized in terms of an optimization problem. That is, an objective function has to be minimized. Nevertheless, existing algorithms do not ensure the user to obtain the global optimum, but only a local one.

This fact is not always relevant in practice, because the results of different executions might be similar, or they permit a user to see the data from different perspectives. Nevertheless, in some applications such differences are important and relevant. In our context, where data is perturbated to ensure privacy and where the distance between clusters is used as a measure of information loss [5] (i.e., as a measure of data distortion), local optima cause difficulties for the evaluation of the methods. Note that the comparison of different local optima gives no real information on the distortion applied to the original data.

In order to solve this difficulty, in [20], we considered 20 executions of each fuzzy clustering method, and then the solution with the lowest objective function was selected. In [21], 30 executions were considered to have still more robust results. All in all, such lowest values are still locally optimal and divergences between the data appear.

Intuitionistic fuzzy partitions permit us to represent in a compact manner the divergence between different executions of the fuzzy clustering algorithms. This is so because intuitionistic fuzzy sets are an adequate concept to represent the uncertainties of the membership functions. In our context, we can then use such intuitionistic fuzzy partitions to compare the results of different executions of fuzzy clustering algorithms on the two different data sets (the original one and the protected one).

III. Preliminaries

In this section we present some concepts that are needed later on. In particular we review some algorithms for fuzzy clustering and the concept of intuitionistic fuzzy set and fuzzy partition.

A. Fuzzy clustering

In this section we review a few concepts on fuzzy clustering. We begin by reviewing the concepts of fuzzy set and fuzzy partitions as all fuzzy clusters are based on such terms.

Definition 1: [22] Let X be a reference set. Then $\mu : X \to [0, 1]$ is a membership function.

Definition 2: [18] Let X be a reference set. Then, a set of membership functions $\mathcal{M} = \{\mu_1, \ldots, \mu_m\}$ is a fuzzy partition of X if for all $x \in X$ it holds that

$$\sum_{i=1}^{m} \mu_i(x) = 1$$

Several algorithms have been developed for fuzzy clustering, including fuzzy c-means, fuzzy c-means with entropy and fuzzy c-means with variable size.

We describe below some of the algorithms using a common framework. In particular, we consider a set of objects $X = \{x_1, \ldots, x_n\}$ from which we want to construct c clusters. Parameter c is assumed to be an input parameter of the clustering algorithm. Then, clustering methods construct the clusters. Fuzzy clustering defines the clusters in terms of a set of membership functions μ_{ik}, where μ_{ik} is the membership of the kth object (x_k) w.r.t. the ith cluster.

Algorithm 1 Fuzzy c-means

Step 1: Generate initial μ and V

Step 2: Solve $min_{\mu \in M} J(\mu, V)$ by computing:

$$\mu_{ik} = \Big(\sum_{j=1}^{c} \Big(\frac{||x_k - v_i||^2}{||x_k - v_j||^2} \Big)^{\frac{1}{m-1}} \Big)^{-1}$$

Step 3: Solve $min_V J(\mu, V)$ by computing:

$$v_i = \frac{\sum_{k=1}^{n} (\mu_{ik})^m x_k}{\sum_{k=1}^{n} (\mu_{ik})^m}$$

Step 4: If the solution does not converge, go to Step 2; otherwise, stop.

We start with a description of fuzzy c-means (FCM). This method was first proposed by Bezdek in [4] and it is described in most books on fuzzy sets and fuzzy clustering. See, e.g., [8], [13], [16]. We will give a brief description below.

Fuzzy c-means uses, besides of parameter c (the number of clusters as explained above), an additional parameter m. This parameter m, which should be such that $m \geq 1$, plays a central role and is used to specify the degree of fuzziness for the solution. The larger the value m, the larger the fuzziness in the clusters. With values near to 1, solutions tend to be crisp. In fact, the solutions with $m = 1$ correspond to the solutions with the k-means algorithm (a crisp clustering algorithm).

The fuzzy c-means clustering algorithm is defined in terms of a minimization problem. That is, its solution is the fuzzy partition μ that minimizes a given expression. The problem is stated below in terms of X, μ, and v_i where v_i is used to represent the cluster center, or centroid, of the i-th cluster.

$$\text{Minimize} J_{FCM}(\mu, V) = \{ \sum_{i=1}^{c} \sum_{k=1}^{n} (\mu_{ik})^m ||x_k - v_i||^2 \} \quad (1)$$

subject to the constraints $\mu_{ik} \in [0, 1]$ and $\sum_{i=1}^{c} \mu_{ik} = 1$ for all k.

A (locally) optimal solution of this problem is obtained using the iterative process described in Algorithm 1. In this algorithm, we do not discuss the case of denominators equal to zero. This is solved with adhoc definitions for μ (see the original references given above). This process interleaves two steps. One that estimates the optimal membership functions of elements to clusters (when centroids are fixed) and another that estimates the centroids for each cluster (when membership functions are fixed).

Entropy based fuzzy c-means (EFCM) is another method of fuzzy clustering. This method, proposed in [14] (see also [13]), introduces fuzziness into the solution with a term based on entropy and a parameter λ ($\lambda \geq 0$) that forces the solution to be fuzzy.

Formally, EFCM is also defined in terms of a function to be minimized. In this case, the objective function is as follows:

$$J_{EFCM}(\mu, V) = \sum_{k=1}^{n} \sum_{i=1}^{c} \{\mu_{ik}||x_k - v_i||^2 + \lambda^{-1}\mu_{ik}log\mu_{ik}\} \tag{2}$$

As in the case of the FCM, the solutions must satisfy a set of constraints. Now, the constraints are as follows $\mu_{ik} \in [0,1]$ and $\sum_{i=1}^{c} \mu_{ik} = 1$ for all k.

The role of parameter λ is similar to the role of m in the fuzzy c-means. Here, the smaller λ, the fuzzier the solutions. Instead, when λ tends to infinity, the second term becomes negligible and the algorithm yields a crisp solution.

The solution of the problem defined by Equation (2) is obtained as above using an interactive process. As above, the steps for computing the membership values (μ_{ik}) and the centers (v_i) are interleaved. However, the expressions in this case are different. The expressions for EFCM are as follows:

$$v_i = \frac{\sum_{k=1}^{n} \mu_{ik} x_k}{\sum_{k=1}^{n} \mu_{ik}} \tag{3}$$

$$\mu_{ik} = \frac{e^{-\lambda||x_k - v_i||^2}}{\sum_{j=1}^{c} e^{-\lambda||x_k - v_j||^2}} \tag{4}$$

For convenience, we will use later on the following expression for μ_{ik}, which is equivalent to Expression (4):

$$\mu_{ik} = \frac{1}{1 + \frac{\sum_{j \neq i}^{c} e^{-\lambda||x_k - v_j||^2}}{e^{-\lambda||x_k - v_i||^2}}} \tag{5}$$

A variation has been introduced for these two clustering methods, which consists of including a size variable for each cluster. This method, proposed in [15], was introduced to reduce misclassification when there are clusters of different size. Otherwise, two adjacent clusters have equal membership function (equal to 0.5) in the mid-point between the two centroids. The size of the i-th cluster is represented with the parameter α_i (the greater α_i, the greater the proportion of the elements that belong to the i-th cluster). A similar approach is given in [9].

Fuzzy c-means with tolerance is another method for fuzzy clustering. The method has parameters c and m as above and an additional term κ_k that corresponds to a boundary condition for the error of the data element x_k. This problem is solved in a similar way, as an optimization problem. In this case, the function to be optimized is the following one:

$$J_{FCM}(\mu, V) = \{\sum_{i=1}^{c} \sum_{k=1}^{n} (\mu_{ik})^m ||x_k + \epsilon_k - v_i||^2\} \tag{6}$$

subject to the constraints $\mu_{ik} \in [0,1]$ and $\sum_{i=1}^{c} \mu_{ik} = 1$ for all k, and with $||\epsilon_k||^2 \leq ||\kappa_k||^2$.

As for the previous methods, this problem is solved using an interactive algorithm that interleaves the calculations of μ and v_i. In addition, we need also to compute ϵ_k. The iterative process is described in Algorithm 2. As in the case of the fuzzy c-means, a (local) minimum of equation 6 is found.

Algorithm 2 Fuzzy c-means with tolerance

Step 1: Generate initial μ and V

Step 2: Solve $min_{\mu \in M} J(\mu, E, V)$ by computing:

$$\mu_{ik} = \Big(\sum_{j=1}^{c} \Big(\frac{||x_k + \epsilon_k - v_i||^2}{||x_k + \epsilon_k - v_j||^2}\Big)^{\frac{1}{m-1}}\Big)^{-1}$$

Step 3: Solve $min_{\epsilon \leq \kappa} J(\mu, E, V)$ by computing:

$$\epsilon_k = -\alpha_k \sum_{i=1}^{c} \mu_{ik}^m (x_k - v_i)$$

where

$$\alpha_k = min\Big(\frac{\kappa_k}{||\sum_{i=1}^{c} \mu_{ik}^m (x_k - v_i)||}, \frac{1}{\sum_{i=1}^{c} \mu_{ik}^m}\Big)$$

Step 4: Solve $min_V J(\mu, E, V)$ by computing:

$$v_i = \frac{\sum_{k=1}^{n} (\mu_{ik})^m x_k}{\sum_{k=1}^{n} (\mu_{ik})^m}$$

Step 5: If the solution does not converge, go to Step 2; otherwise, stop

B. Intuitionistic fuzzy partitions

We have explained in the introduction that different clustering methods are based on different assumptions on the structure of the clusters or on the expressions for the clusters.

Intuitionistic fuzzy sets, which are a generalization of fuzzy sets, permit us to consider the definition of clustering algorithms and methods that return intuitionistic fuzzy clusters. In particular, we consider methods that return partitions defined in terms of intuitionistic fuzzy sets instead of partitions defined in terms of fuzzy sets.

In this section we review intuitionistic fuzzy sets and intuitionistic fuzzy partitions.

Intuitionistic fuzzy sets by Atanassov (for conciseness, A-IFS) are defined below. Formally, they are defined in terms of two functions μ and γ.

Definition 3: [1], [2], [3] An A-IFS (intuitionistic fuzzy set by Atanassov) A in X is defined as

$$A := \{< x, \mu_A(x), \gamma_A(x) > | x \in X\}$$

where $\mu_A : X \to [0,1]$ and $\gamma_A : X \to [0,1]$ with

$$0 \leq \mu_A(x) + \gamma_A(x) \leq 1.$$

For each x, $\mu_A(x)$ and $\gamma_A(x)$ represent the degree of membership and degree of non-membership of the element $x \in X$ to $A \subset X$, respectively.

We will use also the notation $A =< \mu, \gamma >$ if no confusion arises about μ and γ.

Definition 4: For each A-IFS $A := \{< x, \mu_A(x), \gamma_A(x) > | x \in X\}$ as in Definition 3, we define the intuitionistic fuzzy index for $x \in X$ as

$$\nu_A(x) := 1 - \mu_A(x) - \gamma_A(x)$$

Note that when $\nu = 0$, we have a standard fuzzy set with $\gamma = 1 - \mu$.

From now on, we will use the pair $< \mu, \nu >$ as it is more convenient to our purposes. Note also that in the literature is usual to use π instead of ν.

We now review the definition of intuitionistic fuzzy partition. This definition generalizes the concept of fuzzy partition given above (Definition 2). The generalization loosens the condition that the addition of all memberships be equal to one. Note that if we consider a set of A-IFS, it is not possible that all μ add to one and at the same time that all $\mu + \nu$ add to one unless all $\nu = 0$. Our generalization will only require all μ to add to one, and non overlapping $\mu + \nu$ for $\mu + \nu = 1$. We formalize this in the next definition.

Definition 5: [19] Let X be a reference set. Then, a set of A-IFS $A = \{A_1, \ldots, A_m\}$, where $A_i =< \mu_i, \nu_i >$, is an intuitionistic fuzzy partition (I-fuzzy partition, for short) if

1) $\sum_{i=1}^{m} \mu_i(x) = 1$ for all $x \in X$,
2) for all $x \in X$, $|\{i | \mu_i(x) + \nu_i(x) = 1\}| \leq 1$ (there is at most one A-IFS such that $\mu_i(x) + \nu_i(x) = 1$ for all x)

Note that this definition generalizes fuzzy partitions because the definition is equivalent to a fuzzy partition when for all x, $\nu_i(x) = 0$. The following proposition establishes this property.

Proposition 6: [19] I-fuzzy partitions generalize fuzzy partitions.

IV. Intuitionistic fuzzy clustering

Our approach for defining intuitionistic fuzzy clustering is based on the uncertainty of fuzzy clustering methods. This approach, which was first considered in [19], is based on the fact that when different fuzzy clustering algorithms are applied to the same data and with the same parameterizations, different fuzzy partitions can be obtained.

A. IFP for the fuzzy c-means

Intuitionistic fuzzy sets permit us to represent the uncertainty of the membership functions. Due to this, they seem a natural way to represent the divergence between the results of such different executions.

Formally, we define the intuitionistic fuzzy partition from a set of fuzzy partitions where each fuzzy partition is the result of a clustering algorithm. That is, the application of r clustering algorithms to a data set X results into r fuzzy partitions. Moreover, all r fuzzy partitions contain c clusters.

In the construction we will use the following notation. Let $\mu^{i,j}$ represent the membership function obtained by the ith clustering algorithm for the jth cluster. Let $c^{i,j}$ represent the centroid obtained by the ith clustering algorithm for the jth cluster. Let OF_i be the value of the objective function obtained by the ith clustering algorithm.

Now, using these definitions, we define the I-fuzzy partition. In this case, we assume that all clustering methods applied correspond to the fuzzy c-means with the same parameter m (and also the same c).

Definition 7: [19]. Let c, m, v, μ and OF be defined as above and v, c and OF be the result of the application of the fuzzy c-means to data X with parameters c and m. Let $p \leq r$ be an integer. Then, we define the I-fuzzy partition inferred from μ and c as the one obtained from the application of the next steps.

(1) Define the centroids $cm_j := c^{r_*,j}$ and the membership functions $\mu_j := \mu^{r_*,j}$ for all $j = 1, \ldots, c$ where $r_* = arg \min_{i=1}^{r} OF_i$.

(2) Define an alternative centroid cn_j for each cluster j as the average of the p centroids (for that cluster) with minimal OF. Formally, $cn_j := (1/p) \sum_{i=1}^{p} c^{s(i),j}$ where s is a permutation of $\{1, \ldots, c\}$ such that $OF_{s(1)} \leq \cdots \leq OF_{s(c)}$.

(3) Define a radius rad_j for each cluster j as the maximum distance between cn_j and $c^{s(i),j}$ for all $i \leq p$. Formally, $rad_j = \max_{i \leq p} ||c^{s(i),j} - cn_j||$. That is, rad_j is the minimal radius that permits to encompass all p cluster centers when we build a ball with center $c^{s(i),j}$.

(4) Check whether the sets $\{x| \;\; ||x - cn_j|| \leq rad_j\}$ overlap. If they overlap, reduce p and go to Step 2.

(5) Define $b_j := \{x| \, ||x - cn_j|| \leq rad_j\}$ for all clusters $j = \{1, \ldots, c\}$, and define $d(x, y) = ||x - y||$ and $d(x, b_j) = min_{y \in b_j} ||x - y||$. It is easy to see that $d(x, b_j) = ||x - cn_j|| - rad_j$. Then,

$\mu'_j(x) = 1$ for all $x \in b_j$, and,

$\mu'_j(x) = \left(1 + \sum_{t \neq j} \left(\frac{d(x,b_j)^2}{d(x,v_t)^2}\right)^{\frac{1}{m-1}}\right)^{-1}$ for all x not in b_j.

(6) Define $\nu_j(x) = \mu'_j(x) - \mu_j(x)$.

This definition is appropriate in the sense that it permits us to build an I-fuzzy partition. We establish this fact in the next proposition.

Proposition 8: [19] When the set of optimal centroids do not coincide (i.e., when $j_1 \neq j_2$ implies that $c^{r_*,j_1} \neq c^{r_*,j_2}$), the construction above finishes and builds a I-fuzzy partition.

Proof: We need to consider the following:

1) With relation to Step 1 above, the definition of $\mu^{r_*,j}$ from $c^{r_*,j}$ and the data set X is equivalent to a μ_j directly defined from X and cm_j, if no other elements than X and the cluster centers are used for defining the membership functions. This is the case for the fuzzy c-means.
2) With relation to Step 3, note that the p centers of the best p solutions are included in the set of points located at a maximum distance rad_j from cn_j. That is, $c^{i,j} \in \{x| \; ||x - cn_j|| \leq rad_j\}$ for all i in $\{s(1), \ldots, s(p)\}$. Naturally, cm_j is in this set because $cm_j = c^{s(1),j}$. In fact, cm_j is in this set for any $p \geq 1$ as in the extreme case of $p = 1$ we have that $cn_j = c^{s(1),j}$ and $rad_j = 0$.
3) The reduction of p permits us to construct non overlapping regions b_j. As according to the assumption $j_1 \neq j_2$ implies that $c^{r_*,j_1} \neq c^{r_*,j_2}$, there is some

p for which b_j do not intersect.

4) The construction of μ'_j is such that $\mu'_j(x) \geq \mu_j(x)$ for all x, because $d(x, b_j) \leq d(x, v_j)$ and thus $\frac{d(x,b_j)}{d(x,v_t)} \leq \frac{d(x,v_j)}{d(x,v_t)}$ for all t. From this, it is also clear that $\left(\frac{d(x,b_j)}{d(x,v_t)}\right)^{2/m-1} \leq \left(\frac{d(x,v_j)}{d(x,v_t)}\right)^{2/m-1}$ and $1 + \sum_{t \neq j} \left(\frac{d(x,b_j)}{d(x,v_t)}\right)^{2/m-1} \leq 1 + \sum_{t \neq j} \left(\frac{d(x,v_j)}{d(x,v_t)}\right)^{2/m-1}$ Therefore, $\left(1 + \left(\frac{d(x,b_j)}{d(x,v_t)}\right)^{2/m-1}\right)^{-1} \geq \left(1 + \left(\frac{d(x,v_j)}{d(x,v_t)}\right)^{2/m-1}\right)^{-1}$ and, thus, $\mu'_j(x) \geq \mu_j(x)$. Note that $d(x, b_j) \leq d(x, v_j)$ by construction of b_j.

5) As $\mu'_j(x) \geq \mu_j(x)$, $\nu_j(x)$ is positive and, as $\mu'_j(x)$ is a membership function, $\mu_j(x) + \nu_j(x) \leq 1$.

■

Definition 7 depends on the parameter p. The larger p, the larger ν. When $p = 1$, we have that the I-fuzzy partition corresponds to a standard fuzzy partition. This is established in the next proposition.

Proposition 9: [19] Let $A^1 =< \mu_i^1, \nu_i^1 >$ be the I-fuzzy partition inferred from p_1 and let $A^2 =< \mu_i^2, \nu_i^2 >$ be the I-fuzzy partition inferred from p_2. Then, if $p_1 < p_2$ then $\sum_i \sum_{x \in X} \nu_i^1(x) \leq \sum_i \sum_{x \in X} \nu_i^2(x)$. Moreover, if $p_1 = 0$, then $\sum_i \sum_{x \in X} \nu_i^1(x) = 0$.

Finally, we have a result on the convergence of our definition in the long run. That is, it converges into a standard fuzzy partition. To formalize this fact, let us consider a set of fuzzy partitions Π and the addition of a new one π with OF_π; (i) if OF_π is greater than $OF_{s(p)}$, the I-fuzzy partition will not be affected; (ii) if OF_π is greater than $OF_{s(1)}$ but smaller than $OF_{s(p)}$ (i. e., $OF_{s(1)} < OF_\pi \leq OF_{s(p)}$), only ν in the I-fuzzy partition will be affected, but not μ; (iii) only π with $OF_\pi < OF_{s(1)}$ cause a modification of the memberships μ.

So, in general, given a set of fuzzy partitions, the consideration of additional ones only implies a modification of the I-fuzzy partition when their objective function is less than the ones of the p best ones. Therefore, the more executions of fuzzy clustering we do, the more fuzzy partitions we get, and the more similar is the resulting I-fuzzy partition to the optimal fuzzy partition. In this way, there is a trend towards the optimal fuzzy partition. The following proposition establishes that, in the long run, we have such an equality.

Proposition 10: [19] When all p best fuzzy partitions are the same ($c^{s(i),j} = k_j$ for all $i < p$), $\nu_i(x) = 0$ for all x and all i.

B. IFP for the entropy based fuzzy c-means

The definition above is based on the fuzzy c-means. The same is applicable for the entropy based fuzzy c-means (EFCM). In this case, we need to restrict our solution to the case of different executions with the same parameter λ. The definition for the EFCM is as follows:

Definition 11: Let c, λ, v, μ and OF be defined as above and v, c and OF be the result of the application of the entropy based fuzzy c-means to data X with parameters c and λ. Let $p \leq r$ be an integer. Then, we define the I-fuzzy partition inferred from μ and c as the one obtained from the application of the next steps.

(1) Define the centroids $cm_j := c^{r_*,j}$ and the membership functions $\mu_j := \mu^{r_*,j}$ for all $j = 1, \ldots, c$ where $r_* = arg \min_{i=1}^r OF_i$.

(2) Define an alternative centroid cn_j for each cluster j as the average of the p centroids (for that cluster) with minimal OF. Formally, $cn_j := (1/p) \sum_{i=1}^p c^{s(i),j}$ where s is a permutation of $\{1, \ldots, c\}$ such that $OF_{s(1)} \leq \cdots \leq OF_{s(c)}$.

(3) Define a radius rad_j for each cluster j as the maximum distance between cn_j and $c^{s(i),j}$ for all $i \leq p$. Formally, $rad_j = \max_{i \leq p} ||c^{s(i),j} - cn_j||$. That is, rad_j is the minimal radius that permits to encompass all p cluster centers when we build a ball with center $c^{s(i),j}$.

(4) Check whether the sets $\{x| \; ||x - cn_j|| \leq rad_j\}$ overlap. If they overlap, reduce p and go to Step 2.

(5) Define $b_j := \{x| \; ||x - cn_j|| \leq rad_j\}$ for all clusters $j = \{1, \ldots, c\}$, and define $d(x, y) = ||x - y||$ and $d(x, b_j) = min_{y \in b_j} ||x - y||$. Then,
$\mu'_j(x) = 1$ for all $x \in b_j$, and,
$\mu'_j(x) = \frac{1}{1 + \frac{\sum_{t \neq j} e^{-\lambda d(x_k, v_t)^2}}{e^{-\lambda d(x_k, b_j)^2}}}$ for all x not in b_j.

(6) Define $\nu_j(x) = \mu'_j(x) - \mu_j(x)$.

The properties of this definition are similar to the ones given above for Definition 7. In particular, the following propositions can be proved.

Proposition 12: When the set of optimal centroids do not coincide, Definition 11 finishes and builds a I-fuzzy partition.

Proof: We need to consider the following:

1) With relation to (1), the definition of $\mu^{r_*,j}$ from $c^{r_*,j}$ and the data set X is equivalent to a μ_j directly defined from X and cm_j, if no other elements than X and the cluster centers are used for defining the membership functions. This is the case for the entropy based fuzzy c-means.

2) With relation to (3), note that the p centers of the best p solutions are included in the set of points located at a maximum distance rad_j from cn_j. That is, $c^{i,j} \in \{x| \; ||x - cn_j|| \leq rad_j\}$ for all i in $\{s(1), \ldots, s(p)\}$. Naturally, cm_j is in this set because $cm_j = c^{s(1),j}$. In fact, cm_j is in this set for any $p \geq 1$ as in the extreme case of $p = 1$ we have that $cn_j = c^{s(1),j}$ and $rad_j = 0$.

3) The reduction of p permits us to construct non overlapping regions b_j. As according to the assumption $j_1 \neq j_2$ implies that $c^{r_*,j_1} \neq c^{r_*,j_2}$, there is some p for which b_j do not intersect.

4) The construction of μ'_j is such that $\mu'_j(x) \geq \mu_j(x)$ for all x, because $d(x, b_j) \leq d(x, v_j)$ and thus $e^{-\lambda d(x,b_j)} \geq e^{-\lambda d(x,v_j)}$ and $\frac{e^{-\lambda d(x,b_j)}}{\sum_{t \neq j} e^{-\lambda d(x_k,v_t)^2}} \geq \frac{e^{-\lambda d(x,v_j)}}{\sum_{t \neq j} e^{-\lambda d(x_k,v_t)^2}}$. Therefore, we have that $\frac{\sum_{t \neq j} e^{-\lambda d(x_k,v_t)^2}}{e^{-\lambda d(x,b_j)}} \leq \frac{\sum_{t \neq j} e^{-\lambda d(x_k,v_t)^2}}{e^{-\lambda d(x,v_j)}}$ and

$1 + \frac{\sum_{t \neq j} e^{-\lambda d(x_k, v_t)^2}}{e^{-\lambda d(x, b_j)}} \leq 1 + \frac{\sum_{t \neq j} e^{-\lambda d(x_k, v_t)^2}}{e^{-\lambda d(x, v_j)}}$. So, we obtain $\frac{1}{1+\frac{\sum_{t \neq j} e^{-\lambda d(x_k, v_t)^2}}{e^{-\lambda d(x, b_j)}}} \geq \frac{1}{1+\frac{\sum_{t \neq j} e^{-\lambda d(x_k, v_t)^2}}{e^{-\lambda d(x, v_j)}}}$ that means that $\mu'_j(x) \geq \mu_j(x)$.

5) As $\mu'_j(x) \geq \mu_j(x)$, $\nu_j(x)$ is positive and, as $\mu'_j(x)$ is a membership function, $\mu_j(x) + \nu_j(x) \leq 1$.

■

The following two propositions on the convergence of the partitions with respect to parameter p can also be proved for EFCM.

Proposition 13: Let $A^1 =< \mu_i^1, \nu_i^1 >$ be the I-fuzzy partition inferred from p_1 and let $A^2 =< \mu_i^2, \nu_i^2 >$ be the I-fuzzy partition inferred from p_2 according to Definition 11. Then, if $p_1 < p_2$ then $\sum_i \sum_{x \in X} \nu_i^1(x) \leq \sum_i \sum_{x \in X} \nu_i^2(x)$. Moreover, if $p_1 = 0$, then $\sum_i \sum_{x \in X} \nu_i^1(x) = 0$.

Proposition 14: When the I-fuzzy partitions are constructed according to Definition 11, when all p best fuzzy partitions are the same ($c^{s(i),j} = k_j$ for all $i < p$), $\nu_i(x) = 0$ for all x and all i.

C. IFP for the fuzzy c-means with tolerance

We now consider the case of fuzzy c-means with tolerance. The main difficulty here is that the expression for μ depends on ϵ_k. Therefore, it is not easily generalizable into a function $\mu_i(x)$ as we did for the fuzzy c-means and the fuzzy c-means with entropy.

That is, we need to transform:

$$\mu_{ik} = \Big(\sum_{j=1}^{c} \Big(\frac{||x_k + \epsilon_k - v_i||^2}{||x_k + \epsilon_k - v_j||^2} \Big)^{\frac{1}{m-1}} \Big)^{-1}$$

into a function

$$\mu_i(x) = \Big(\sum_{j=1}^{c} \Big(\frac{||x + \epsilon(x) - v_i||^2}{||x + \epsilon(x) - v_j||^2} \Big)^{\frac{1}{m-1}} \Big)^{-1}$$

where $\epsilon(x_k) = \epsilon_k$.

Our approach to define μ is to consider the set $E = \{(x_k, \epsilon_k)\}_k \cup \{(v_i, 0)\}_i$ and then assign $(x, \epsilon(x)) = \arg\min_{(x_i, \epsilon_i) \in E} d(x, x_i)$. This definition satisfies $\mu_i(x_k) = \mu_{ik}$ (as desired) although continuity might not hold because $\epsilon(x)$ is not continuous.

Definition 15: Let c, λ, v, μ and OF be defined as above and v, c and OF be the result of the application of the fuzzy c-means with tolerance κ_k to the dataset X with parameters c and λ. Let $p \leq r$ be an integer. Then, we define the I-fuzzy partition inferred from μ and c as the one obtained from the application of the next steps.

(1) Define the centroids $cm_j := c^{r_*,j}$ and the membership functions

$$\mu_j(x) = \Big(\sum_{t=1}^{c} \Big(\frac{||x + \epsilon(x) - cm_j||^2}{||x + \epsilon(x) - cm_t||^2} \Big)^{\frac{1}{m-1}} \Big)^{-1}$$

for all $j = 1, \ldots, c$ and where $r_* = arg \min_{i=1}^{r} OF_i$, and where $(x, \epsilon(x)) = \arg\min_{(x_i, \epsilon_i) \in E} d(x, x_i)$ for $E = \{(x_k, \epsilon_k^{r_*})\}_k \cup \{(cm_j, 0)\}_j$ and with $\epsilon_k^{r_*}$ the ϵ_k obtained by the method r_*.

(2) Define an alternative centroid cn_j for each cluster j as the average of the p centroids (for that cluster) with minimal OF. Formally, $cn_j := (1/p) \sum_{i=1}^{p} c^{s(i),j}$ where s is a permutation of $\{1, \ldots, c\}$ such that $OF_{s(1)} \leq \cdots \leq OF_{s(c)}$.

(3) Define a radius rad_j for each cluster j as the maximum distance between cn_j and $c^{s(i),j}$ for all $i \leq p$. Formally, $rad_j = \max_{i \leq p} ||c^{s(i),j} - cn_j||$. That is, rad_j is the minimal radius that permits to encompass all p cluster centers when we build a ball with center $c^{s(i),j}$.

(4) Check whether the sets $\{x| \ \ ||x - cn_j|| \leq rad_j\}$ overlap. If they overlap, reduce p and go to Step 2.

(5) Define $b_j := \{x| \, ||x - cn_j|| \leq rad_j\}$ for all clusters $j = \{1, \ldots, c\}$, and define $d(x, y) = ||x - y||$ and $d(x, b_j) = min_{y \in b_j} ||x - y||$. Then,

$\mu'_j(x) = 1$ for all $x \in b_j$, and,

$\mu'_j(x) = \Big(1 + \sum_{t \neq 1} \Big(\frac{d(x+\epsilon(x), b_j)^2}{d(x+\epsilon(x), v_t)^2} \Big)^{\frac{1}{m-1}} \Big)^{-1}$ for all x not in b_j.

(6) Define $\nu_j(x) = \mu'_j(x) - \mu_j(x)$.

The properties of this definition are analogous to the ones given above for the other constructions. In particular, we have a proposition analogous to Proposition 8.

Proposition 16: When the set of optimal centroids do not coincide, the construction above finishes and builds a I-fuzzy partition.

Proof: The proof of this proposition is analogous to Proposition 8. We need to define here $d(x, y)$ and $d(x, b_j)$ as above and then to consider that $d(x + \epsilon(x), b_j) \leq d(x + \epsilon(x), v_j)$. From this we get $\frac{d(x+\epsilon(x), b_j)}{d(x+\epsilon(x), v_t)} \leq \frac{d(x, v_j)}{d(x, v_t)}$ for all t. The rest of the proof is as above. ■

V. Conclusions and future work

In this paper we have studied intuitionistic fuzzy partitions and proposed a method to define them from standard fuzzy ones. Such a method can be applied to the results of several executions of a fuzzy method (e.g. fuzzy c-means or entropy based fuzzy c-means).

In order to permit the application of our method to fuzzy clusters obtained by e.g. fuzzy c-means with different parameters m, we have revised a previous definition of intuitionistic fuzzy partition.

Acknowledgments

Partial support by Generalitat de Catalunya (AGAUR, 2006BE-2 00338, 2005 SGR 00446 and 2005-SGR-00093) and by the Spanish MEC (projects ARES – CONSOLIDER INGENIO 2010 CSD2007-00004 – and e-AEGIS – TSI2007-65406-C03-01/02) is acknowledged.

REFERENCES

[1] Atanassov, K. T. (1983) Intuitionistic fuzzy sets, VII ITKR's Session, Sofia, 1983 (deposed in Central Sci.-Technical Library of Bulg. Acad. of Sci., 1697/84) (in Bulgarian).

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

[3] Atanassov, K. T. (1999) Intuitionistic Fuzzy Sets, Physica-Verlag, Heidelberg.

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

[5] Domingo-Ferrer, J., Torra, V. (2001) Disclosure Control Methods and Information Loss for Microdata, in Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies, P. Doyle, J. I. Lane, J. J. M. Theeuwes, L. M. Zayatz (Eds.), Elsevier, pp. 91-110,

[6] Hasegawa, Y., Endo, Y., Hamasuna, Y., Miyamoto, S. (2007) Fuzzy c-means for data with tolerance defined as hyper-rectangle, Proc. MDAI 2007, Lecture Notes in Artificial Intelligence 4617 237-248.

[7] Hastie, T., Tibshirani, R., Friedman, J., (2001), *The Elements of Statistical Learning*, Berlin: Springer.

[8] Höppner, F., Klawonn, F., Kruse, R., Runkler, T., (1999), Fuzzy cluster analysis, Wiley.

[9] Ichihashi, H., Honda, K., Tani, N. (2000) Gaussian mixture PDF approximation and fuzzy c-means clustering with entropy regularization, Proc. of the 4th Asian Fuzzy System Symposium, 217–221.

[10] Ladra, S., Torra, V. (2007) On the comparison of generic information loss measures and cluster-specific ones, submitted.

[11] Ladra, S., Torra, V. (2007) Fuzzy clustering in data privacy: measuring information loss for synthetic data, submitted

[12] Lopez de Mantaras, R. (1991) A Distance-Based Attribute Selection. Measure for Decision Tree Induction, Machine Learning, 6, 81-92.

[13] Miyamoto, S., (1999), Introduction to fuzzy clustering, (in Japanese), Ed. Morikita, Tokyo.

[14] Miyamoto, S., Mukaidono, M. (1997) Fuzzy c - means as a regularization and maximum entropy approach, Proc. of the 7th IFSA Conference, Vol.II 86-92.

[15] Miyamoto, S., Umayahara, K. (2000) Fuzzy c-means with variables for cluster sizes (in Japanese), Proc. of the 16th Fuzzy System Symposium, 537-538.

[16] Miyamoto, S., Umayahara, K. (2000) Methods in Hard and Fuzzy Clustering, pp 85–129 in Z.-Q. Liu, S. Miyamoto (Eds.), Soft Computing and Human-Centered Machines, Springer-Tokyo.

[17] Murata, R., Endo, Y., Haruyama, H., Miyamoto, S. (2006) On fuzzy c-means for data with tolerance, J. of advanced computational intelligence and intelligent informatics, 10:5 673-681.

[18] Ruspini, E. H. (1969) A new approach to clustering, Inform. Control. 15 22-32.

[19] Torra, V., Miyamoto, S. (2007) Intuitionistic Fuzzy Partitions for the Comparison of Fuzzy Clusters, submitted.

[20] Torra, V., Endo, Y., Miyamoto, S. (2007) On the comparison of some fuzzy clustering methods for privacy preserving data mining: towards the development of specific information loss measures, submitted.

[21] Torra, V., Endo, Y., Miyamoto, S. (2007) On the analysis of the parameters for fuzzy clustering with tolerance, draft.

[22] Zadeh, L. A. (1965) Fuzzy sets, Inform. Control. 8 338-353.