Clustering Fuzzy Relational Data base on fuzzy Cardinality

Narci Clara

Abstract

Our purpose is to provide a set-theoretical frame to clustering fuzzy relational data basically based on the cardinality of the fuzzy subsets and their complementaries that represent the objects under study, without applying any crisp property when the various elements that compose the process are fuzzified.
From this perspective we define a family of fuzzy similarity indexes which includes a set of fuzzy indexes introduced by Tolias et al, and we analyze under which conditions a fuzzy proximity relation is defined. Following an original idea due to S. Miyamoto we evaluate the similarity between objects and between clusters by means the same mathematical procedure. Joining these concepts and methods we establish an algorithm to clustering fuzzy relational data. Finally, we present an example to make clear the whole process.

Keywords: Clustering algorithm, Fuzzy relation, Fuzzy clustering, Generalized fuzzy similarity index.

1. Introduction

Clustering analysis is one of the most important applied techniques for pattern recognition. The basic idea is to group together objects closely related. There are two kinds of general ways to make a clustering process: methods with objective function models using object data and methods with relational clustering. The first ones are very developed and some of them with a great success as the fuzzy c-means families [1-3], [7], [12], [20] or hybrid clustering models [17]. For the second ones we can distinguish also several types: methods that rely on optimization of an objective function of the relational data and methods that use decompositions of relation matrices, transitive closures or mathematical algorithms. In all approaches is essential to recognize how similar are objects. Similarity between objects is a datum in some environments done by a square matrix but in the context of Fuzzy Features Contrast Model (FFCM) the data is a fuzzy relation between objects and features (in general a rectangular matrix)and the similarity matrix is obtained from it.

Basically, from an algebraic point of view, and so using matrices of relational data, a fuzzy clustering process consists in three steps. The first one is to define a fuzzy relation between objects and features, after that we have to evaluate the similarity between objects which can be thought as a fuzzy proximity relation and, finally, we determine the partitions by means an algorithm [16] or by its transitive closure for some *t*-norm. All these methods permit to represent the clusters in a dendogram.

In this case of using *t*-norms, if the selected *t*-norm is the minimum we obtain directly the dendogram [11], on the contrary we need to apply some iterative method [21].

If we prefer to apply any hierarchical method based on an algorithm then two clusters are usually grouped to constitute a new cluster when their similarity attaints the maximum value. The process continues defining the similarity between this new element an the others. Many methods deal with this objective, the most usual is the single linkage, complemented by the complete linkage and the average linkage.

In the following subsections we revise and summarize these concepts focusing in those aspects that we believe that are much relevant.

1.1. A set-theoretical model based on the Fuzzy Features Contrast Model

Features Contrast Model (FCM), a crisp version of the Fuzzy Features Contrast Model, asserts that similarity should be described as a comparison of features that describe the objects under consideration, and expresses similarity between objects as a function of their common and distinctive features [18].

Let $X = \{A_1, A_2, ..., A_n\}$ be a set of objects and $Y = \{P_1, P_2, ..., P_m\}$ a set of features. Function

Corresponding Author: Narci Clara
Departament d'Informµatica i Matemµatica Aplicada Universitat de Girona Campus de Montilivi, 17071 Girona, Catalonia (Spain)
narcis.clara@udg.edu
November 22, 2007

$v : X \rightarrow \wp(Y)$ defined as $P_j \in v(A_i)$ if and only if A_i verifies P_j serves to relate objects with features. FCM defines a crisp binary relation determined by a matrix $A = (a_{ij})$ with n rows and m column where $a_{ij} = 1$ if $P_j \in v(A_i)$. It is equivalent to assign to each element A_i a value $\mu A_i(P_k) \in \{0,1\}$ depending on if A_i verifies or not feature P_k, synthetically

$$A_i \rightarrow (\mu A_i(P_1), \mu A_i(P_2), ..., \mu A_i(P_m)) \quad (1)$$

The model FCM has a great disadvantage because it is only intended for binary values. In the fuzzy context the binary condition is equivalent to that the membership function of the element A_i in the feature P_j takes the values 1 or 0. Therefore $|v(A_i)| = \sum_{j=1}^{n} \mu A_i(P_j)$, namely the number of feature that verifies A_i.

Fuzzy Features Contrast Model (FFCM) is an extension of FCM when the components of the vector in (1) belong to the interval [0, 1] so they are not crisp but fuzzy. We represent each object A_i as a fuzzy subset $\tilde{A}_i$ of the set of features Y namely

$$A_i \rightarrow \tilde{A}_i = (\mu\tilde{A}_i(P_1), \mu\tilde{A}_i(P_2), ..., \mu\tilde{A}_i(P_m)) \quad (2)$$

In these conditions $card(\tilde{A}_i) = |\tilde{A}_i| = \sum_{j=1}^{n} \mu\tilde{A}_i(P)_j$ [6]. All information can be put in a matrix $n \times m$ representing a fuzzy relation between objects and features. Each row is the set of true predicates for an object depending on the features. Therefore two objects A and B are represented by two fuzzy subsets $\tilde{A}$ and $\tilde{B}$. The membership values are obtained from the inner structure of the data or by experts. Following the same accepted hypothesis that for FCM, FFCM asserts that similarity between A and B has to be a function of $\tilde{A} \cap \tilde{B}, \tilde{A} - \tilde{B}$ and $\tilde{B} - \tilde{A}$.

1.2. Fuzzy similarity measures

A very common approach in engineering and applied sciences is to measure the similarity between objects associating them to a *n*-dimensional vector and calculating their similarity by means of a decreasing function $s = f(d)$ (usually $s = 1 - d$) of their normalized distance d defined by the same properties that in classical geometry (Ex: Ecludian, L^p, Hamming, Minkowski, Mahalanobis, etc.) so following the first assumption of FFCM but the second one because an object is only thought as a point in a metric space. Very strong properties are fulfilled what can be appropriate for several applied problems but, on the other hand, when the human judgement has a principal role in the description of the objects can lead to undesirable properties because do not take in account the deep fuzzy character of the data. For instance, two objects with the maximum uncertainty are judged as completely similar because the distance is 0 thus the similarity is 1 which could be not adequate in some scientific and technological domains. Measures associated with distance have been very studied and applied [10], in particular, the Euclidean [8] or Minkowski [9] distances.

To deal with the fuzzy character of the data sometimes a softer definition is usually taken in account. A function $s : \tilde{P}(E) \times \tilde{P}(E) \rightarrow [0,1]$ is a fuzzy similarity measure if and only if verifies that for any pair of fuzzy subsets $\tilde{A}, \tilde{B}$ then $0 \leq s(\tilde{A}, \tilde{B}) \leq 1$ and $s(\tilde{A}, \tilde{B}) = s(\tilde{B}, \tilde{A})$ (symmetry). If moreover $s(\tilde{A}, \tilde{A}) = 1$ for any $\tilde{A}$ (reflexivity), it can be interpreted as a *fuzzy proximity relation* in X [11], [13]. These kind of fuzzy relations are fundamental to carry out a fuzzy clustering process [21].

For binary relational data and, in order to analyse how similar are two objects, similarity and dissimilarity crisp parameters are very appropiated [16]:

$$\begin{aligned} a &= |v(A_i) \cap v(A_j)| \quad b = |v(A_i) \cap v(A_j)^c| \\ c &= |v(A_i)^c \cap v(A_j)| \quad d = |v(A_i)^c \cap v(A_j)^c| \end{aligned} \quad (3)$$

Parameters a and d evaluate common features and $b$and c distinctive features. Many similarity indexes are defined from similarity and dissimilarity parameters for crisp sets, the most usuals take the form

$$s_m(A_i, A_j) = \frac{\alpha_1 a + \alpha_2 d}{\alpha_3 a + \alpha_4 d + \alpha_5 c + \alpha_6 d} \quad (4)$$

Where $a_i \in \mathbf{N}$, for instance, s_m is called simple matching coefficient if $\forall i \alpha_i = 1$; Jackard coefficient if $\alpha_1 = \alpha_3 = \alpha_4 = \alpha_5 = 1$ and $\alpha_2 = \alpha_6 = 0$; Rao's coefficient is defined by $\alpha_i = 1, i \neq 2$ and $\alpha_2 = 0$ and so on.

Notice that these indexes verify some crisp properties. It is trivial that $a + b + c + d = m$, moreover b and c are the sum of the absolute values of the differences between the components of the respective binary vectors [16]. At this point of the line of argument we can generalize indexes to vectors defined in (2). For instance, from the previous observations we obtain the fuzzy simple matching coefficient s_{fm} as following

$$s_{fm}(\tilde{A},\tilde{B}) = 1 - \frac{1}{m}\sum_{k=1}^{m} \left|\mu_{\tilde{A}}(P_k) - \mu_{\tilde{B}}(P_k)\right| \quad \textbf{(5)}$$

It is trivial that $s_{fm}(\tilde{A},\tilde{B}) = 1 - d_H(\tilde{A},\tilde{B})$ where d_H is the normalized distance of Hamming. Note that s_{fm} is reflexive and symmetric thus define a fuzzy proximity relation and in particular is a fuzzy similarity measure. Other generalizations can be obtained for the rest of the indexes.

1.3. Transitive closures

Following concepts and results are forward applied to fuzzy clustering. Let $\tilde{R}$ be a fuzzy relation in a set of reference X defined as a fuzzy subset of $X \times X$ [6], then $\tilde{R}$ is transitive if and only if

$$\forall a,b \in A \quad \mu_{\tilde{R}}(a,b) \geq \max_{c\in A} \min(\mu_{\tilde{R}}(a,c), \mu_{\tilde{R}}(c,d))$$

A more general concept that includes the preceding one is *t*-transitivity, which is a generalization of the transitivity for any *t*-norm. $\tilde{R}$ is *t*-transitive if and only if

$$\forall a,b \in A \quad \mu_{\tilde{R}}(a,b) \geq \max_{c\in A} t(\mu_{\tilde{R}}(a,c), \mu_{\tilde{R}}(c,d))$$

A proximity fuzzy relation $\tilde{R}$ is a *t*-equivalence fuzzy relation if verifies the *t*-transitive condition. In some scientific literature is called a similarity relation but in this paper we save this nomenclature to determine how similar are two objects. The concept of closure *t*-transitive $\tilde{R}^*$ of a proximity relation $\tilde{R}$ is defined as the *smaller* *t*-equivalence relation that includes $\tilde{R}$ namely

$$\tilde{R}^* = \min\left\{\tilde{S} : \tilde{R} \subset \tilde{S} \text{ and } \tilde{S} \textit{ is } t - \text{equivalent}\right\} \quad (6)$$

When the *t*-norm is not explicitly mentioned means that we use the *t*-norm of the minimum. Designating by A the matrix representation of $\tilde{R}$, we calculate its *t*-transitive closure by [4], [11]

$$A^* = \sum_{i=1}^{n} A^i \quad (7)$$

where $A^2 = (a_{ij}^2)$ and $(a_{ij}^2) = \max_{k=1\ldots n} t(a_{ik}, a_{ki})$. We define A^i in a recurrent form for $i > 2$.

If $\tilde{R}$ is a proximity relation therefore $A^{i-1} \leq A^i$ and applying (7)

$$A^* = A^n \quad (8)$$

If n is not finite then $A^* = \lim_{n\to\infty} A^n$. Other methods most performing to calculate the transitive closure, reducing its computing time from order $O(n^5)$ to order $O(n^2)$, have been published [14].

1.4. Hierarchical methods for fuzzy clustering relational data

Hierarchical methods consist in not considering fix beforehand how may clusters there are and are divided in agglomerative and divisive methods. In this paper we only refer to agglomerative methods in which we start from the most finer partition composed by all the singletons until we merge all the objects in an unique cluster. A cluster has to be understood as a group of similar objects with the condition that we do not have information about the classes of grouping a priori of the data processing. Objects are grouped only in function of their elements. It seems very logical that at each step we group elements with the maximum similarity (minimum distance). When two objects have been grouped our set of reference changes and its cardinal diminishes in one unit. In order to continue the procedure we have to define the similarity between this new element and the others. The form in which is carried out this process determines all the clustering process because different definitions for similarity between clusters induce different algorithms. If Δ and Γ are two clusters, the most known is given by the single linkage method

$$s(\Delta,\Gamma) = \max_{x\in\nabla, y\in\Gamma} s(x,y) \quad (9)$$

Equivalently, if we group Δ and Γ so $\sum = \Delta \bigcup \Gamma$ therefore

$$\forall \mathbf{\Phi}, \mathbf{\Phi} \neq \Delta,\Gamma \;\; s(\sum,\mathbf{\Phi}) = \max(s(\Delta,\mathbf{\Phi}), s(\Gamma,\mathbf{\Phi})) \quad (10)$$

Others, as complete linkage or average linkage s change the maximum for the minimum or an average respectively.

Another approach is based on the theory of fuzzy relations. We calculate the transitive closure with the *t*-norm of the minimum and for each α level we obtain a partition. A very important theorem proves that this partition coincides with the results of the simple linkage clustering process and the connected components of a fuzzy graph [16]. All that makes advisable to use this method for a great set of applications. In fact what happens is that the condition of transitivity for the similarity measure s $s(x,z) \geq \max_{y\in A}(\min(s(x,y), s(y,z)))$ is equivalent to the ultrametric property $\forall x,z \in A$ $d(x,z) \leq \min_{y\in A} \max(d(x,y), d(y,z))$

We can easily check this relevant result simply remembering that $s(x,y)=1-d(x,y)$.

Moreover, we know that the *t*-norm of the minimum is the greatest *t*-norm and it is easy to prove that if t_1 and t_2 are *t*-norms therefore if $t_1 \le t_2$ the a fuzzy relation t_1-transitive is also t_1-transitive.

All that show that our best option is to find the transitive closure for the *t*-norm of the minimum (also called max-min composition). Unfortunately, we obtain undesirable results in some applications what means that for all α the partition is formed only by all the set of objects or by *n* clusters all them constituted by only one element. Obviously, these partitions are trivial and without applied interest. When that happens, and taken in account that the ultrametric property derives the triangular inequality, it seems logical, if necessary, to soften the *t*-norm but only until the triangular inequality was verified, what happens with the *t*-norm of the bounded product $t_{bp}(x,y)=\max(0,x+y-1)$. This property arises from the fact that the triangular inequality $\forall x,y,z \in A \quad d(x,z) \le d(x,y)+d(y,z)$ is equivalent to $1-s(x,z) \le 1-s(x,y)+1-s(y,z)$ so $s(x,z) \ge s(x,y)+s(y,z)-1$ and $s(x,z) \ge 0$, therefore for any pair of elements $x,z \in A$

$$s(x,z) \ge \max_{y \in A}\{\max\{0, s(x,y)+s(y,z)-1\}\}=$$

$$=\max_{y \in A}\{t_{bp}(s(x,y),s(y,z))\}$$ thus s is t_{bp}-transitive.

All that makes clear a new more general strategy consisting in calculating the closure *t*-transitive for some *t*-norm greater that the bounded product, and by means an iterative process obtaining a partition. Following this procedure Miin-Shen Yang et al have implemented and algorithm that in certain domains improve the single linkage method but loosing uniqueness what is a great inconvenient [21]. The algorithm proposed at the end of the paper wants to be another alternative to these kind of methods.

2. A Homogeneous Fuzzy Set Theoretical Frame

The aim of this section is to provide an homogeneous structure for fuzzy clustering especially concerning the way in updating the similarity between clusters. Our fonamental assumption is to relate all the different parts of the process with the fuzzy cardinality. Similarity is modelled by means a set of similarity indexes which are a generalized set of the generalized Tverski indexes [18]. We define the fuzzy structure of the clusters from the fuzzy structure of the objects that compose them, what allow us to determine similarities between objects and between clusters with the same mathematical methodology. In fact, we in tend to put together in a new way known results of relational fuzzy clustering with a new family of similarity indexes.

2.1. Generalized fuzzy similarity indexes

We wish to generalize crisp indexes defined in (4) without applying any crisp property (as is the case of (5)), and proving that reliable properties are fulfilled focusing principally in the simple matching coefficient. Following definitions and properties are a generalization of some concepts due to Tolias et al which define a set of fuzzy similarity indexes Generalized Tversky index [18, 19] - that represents a fuzzified restraint set of crisp indexes. This set contains as a particular case Jackard's coefficient but simple matching coefficient. What we propose has a similar structure but including the simple matching coefficient.

FFCM assumes that similarity between two objects depends on their common and distinctive features. We think that this affirmation can be interpreted in a general way in the sense that it is also relevant in which degree two objects share the negation of a specific feature. Assuming this point of view similarity between *A* and *B* is a function of $\tilde{A} \cap \tilde{B}$, $\tilde{A}-\tilde{B}=\tilde{A}\cap\tilde{B}^c$, $\tilde{B}-\tilde{A}=\tilde{B}\cap\tilde{A}^c$ and $\tilde{A}^c \cap \tilde{B}^c$. From now on *t* and *n* means a *t*-norm and a negation respectively [15]. Making a fuzzy generalization of (3) we introduce the fuzzy similarity and dissimilarity parameters in the following form:

$$\begin{aligned}
a &= \left|\tilde{A}_i \cap \tilde{A}_j\right| = \sum_{k=1}^{m} t(\mu_{\tilde{A}_i}(P_k), \mu_{\tilde{A}_j}(P_k)) \\
b &= \left|\tilde{A}_i \cap \tilde{A}_j^{\,c}\right| = \sum_{k=1}^{m} t(\mu_{\tilde{A}_i}(P_k), n(\mu_{\tilde{A}_j}(P_k))) \\
c &= \left|\tilde{A}_i^{\,c} \cap \tilde{A}_j\right| = \sum_{k=1}^{m} t(n(\mu_{\tilde{A}_i}(P_k)), \mu_{\tilde{A}_j}(P_k)) \\
d &= \left|\tilde{A}_i^{\,c} \cap \tilde{A}_j^{\,c}\right| = \sum_{k=1}^{m} t(n(\mu_{\tilde{A}_i}(P_k)), n(\mu_{\tilde{A}_j}(P_k)))
\end{aligned} \tag{11}$$

Equivalently to (3) we determine a family of generalized fuzzy similarity indexes defined by

$$s_g(\tilde{A},\tilde{B}) = \frac{\left|\tilde{A}^c \cap \tilde{B}^c\right| + \left|\tilde{A}^c \cap \tilde{B}^c\right|}{\left|\tilde{A} \cap \tilde{B}\right| + \left|\tilde{A}^c \cap \tilde{B}^c\right| + \lambda\left|\tilde{A}-\tilde{B}\right| + \mu\left|\tilde{B}-\tilde{A}\right|} \tag{12}$$

This family includes a generalization of the simple matching coefficient for ¸ $\lambda=\mu=1$ and verifies the expected assumption of FFCM:

Exists $f : \tilde{P}(Y)\times\tilde{P}(Y)\times\tilde{P}(Y)\times\tilde{P}(Y)\to R$ that for any elements $\tilde{A},\tilde{B},\tilde{C},\tilde{D}$ belonging to $\tilde{P}(E)$ therefore $s_g(\tilde{A},\tilde{B}) = f(\tilde{A}\cap\tilde{B},\tilde{A}^c\cap\tilde{B}^c,\tilde{A}-\tilde{B},\tilde{B}-\tilde{A})$ Moreover monotony is also fulfilled in the sense that if $\tilde{A}\cap\tilde{C}\subset\tilde{A}\subset\tilde{B},\tilde{A}^c\cap\tilde{C}^c\subset\tilde{A}^c\cap\tilde{B}^c$, $\tilde{A}-\tilde{B}\subset\tilde{A}-\tilde{C}$ and $\tilde{B}-\tilde{A}\subset\tilde{C}-\tilde{A}$ therefore $s_g(\tilde{A},\tilde{B})\geq s_g(\tilde{A},\tilde{C})$ [5].

It is clear that s_g is a fuzzy similarity measure. A rellevant property issues from the fact that s_g is a proximity relation if the associated t-norm verifies the noncontradiction principle [15] namely $s_g(\tilde{A},\tilde{A}) = 1$ if and only if $\left|\tilde{A}-\tilde{A}\right| = \left|\tilde{A}\cap\tilde{A}^c\right| = 0$ what means that for all $x\in[0,1]$ $t(x,1-x)=0$ thus the *t*-norm have to verify the noncontradiction principle. Moreover, under prototypical conditions $s_g((1,1,...,1),\tilde{B})$ does not depend on the *t*-norm, and $s_g((1,1,...,1),(1,1,...,1)) = 1$ what seems to be very plausible properties. Under maximum uncertainty conditio $s_g((0.5,0.5,...,0.5),\tilde{B}) = 0.5$. This last property solves the question about the inconsistency, in some domains, in considering absolutely similar two objects with the maximum uncertainty condition.

In order to show the behaviour of the previous similarity measures we will draw two similarity surfaces supposing that $Y=\{P_1,P_2\}$ and $\mu_{\tilde{A}}(P_k)=\mu_{\tilde{A}}(P_k)(x)$ for $k=1,2$. In our case these two fuzzy subsets are two triangular fuzzy numbers: $T_1=(12,16,20)$ and $T_2=(1,3,5)$ represented in Figures 1 and 2. On the other hand, $\mu_{\tilde{B}}(P_k)$ are two reference values assigned, for instance, by experts. Therefore $s(\tilde{A},\tilde{B})=s(\tilde{A},\tilde{B})(x,y)$ with $(x,y)\in D_1\times D_2$ where D_1 and D_2 are the domains of $\mu_{\tilde{A}}(P_1)$ and $\mu_{\tilde{A}}(P_2)$ respectively. Figures 3 and 4 show the behaviour of the similarity measure s_g for $\lambda=\mu=1$ (we will call it s_{gsm} forward) under prototypical conditions and for another one as for instance: $\mu_{\tilde{B}}(P_1)=0.85$ and $\mu_{\tilde{B}}(P_2)=0.31$.

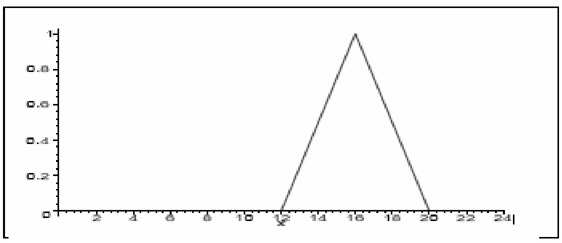

Figure 1: 1 Membership function of $\mathbf{T}_1$.

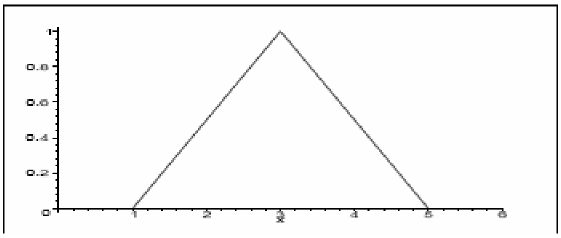

Figure 2: Membership function of $\mathbf{T_2}$.

2.2. Clustering method based on the fuzzy cardinality

This method has been conceived from some observations due to S. Miyamoto. From a methodological point of view it seems logical to use similars procedures in the diferent steps of the clustering process seeking a most homogenous theoretical frame.

If we analyze the deep structure of the different math ematical objects that take part in all the process we discover some analogies and differences. Let $\tilde{\wp}(Y)$ be the power set of fuzzy subsets of *Y* therefore we can define a map between objects and fuzzy subsets of features.

$$X\to\tilde{\wp}(Y) \qquad\qquad Y\overset{\mu_{\tilde{A}_i}}{\to}[0,1]$$
$$\text{so that}$$
$$A_i\to\tilde{A}_i \qquad\qquad P_j\to\mu_{\tilde{A}_i}(P_j))$$

When we form a new cluster we make the union of two precedent clusters. If order to attaint the homogeneity in the procedure it is necessary to define the membership function of a cluster $\tilde{\Delta}\in\tilde{P}(Y)$ with out confusing it with the union of two fuzzy subsets, namely if $\Delta=\{A,B\}$ then $\tilde{A}\cup\tilde{B}\neq\tilde{\Delta}=\{A,B\}^{\sim}$.

Let $\Delta=\{A_i\}_{i\in I^*}\in\wp(X)$ be a cluster, $I^*\subset I$ where $I=\{1,2,...,n\}$. We define the following map

$$\wp(X)\to\hat{\wp}(Y) \quad \text{so that} \quad \Delta\to\tilde{\Delta}$$

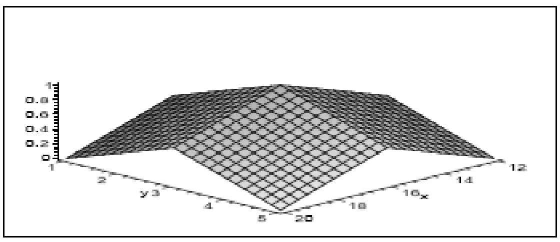

Figure 3: $\mathbf{s_{gsm}}$ under prototypical conditions.

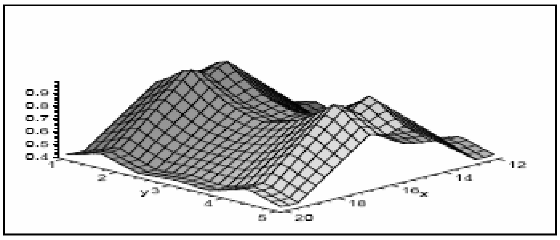

Figure 4: $\mathbf{s}_{gsm}$ under condition **(0:85; 0:31)**.

$$Y \xrightarrow{\mu_{\tilde{\Delta}}} [1,0]$$

$$P_j \rightarrow \mu_{\tilde{\Delta}}(p_j) = \frac{\sum_{i \in I^*} \mu_{\tilde{A}_i}(P_j)}{|\Delta|} \quad (13)$$

If Γ and Λ are two clusters represented by fuzzy subsets $\tilde{\Gamma} = \{\tilde{g}, \tilde{g} \in \Gamma\}$ and $\tilde{\Lambda} = \{\tilde{h}, \tilde{h} \in \Lambda\}$, we define

$$s(\Gamma, \Lambda) = \\ = \frac{|\Gamma \cap \Lambda| + |\Gamma^c \cap \Lambda^c|}{|\Gamma \cap \Lambda| + |\Gamma^c \cap \Lambda^c| + \lambda|\Gamma \cap \Lambda^c| + \mu|\Gamma^c \cap \Lambda|} \quad (14)$$

Unfortunately, do not exist a formula to update the similarity measures from one step to another. It is necessary to calculate another time the membership function of the new cluster determining the similarity measure between this new element and the others. After these considerations we have to establish an algorithm to find the partition. We present a procedure that has a similar structure to the classical algorithms of hierarchical clustering; obviously, other algorithms can be introduced but all the ideas lead us to an algorithm similar to this one. We do not use a programming language, is a description of the necessary steps to achieve the hierarchical clustering. The iterative process of the classical methods is modified by the following form

Input: A set of objects X with $|X| = n$, a set of features Y with $|Y| = m$ and a fuzzy relation $\tilde{S}$ between X and Y represented by a matrix $A \in M_{n \times m}([0,1])$.

Output: A directed tree that describes the process of generating clusters called dendogram.

Algorithm.

- For $i = 1,...,n$ do $\Delta_i = \{A_i\}$, $Z^{(0)} = \{\Delta_1,..., \Delta_n\}$ and $I = \{1,...,n\}$
- For $i, j \in I, i \neq j$ do $s(\Delta_i, \Delta_j) = s_g(A_i, A_j)$
- $l = n$, $k = 0$ where l enumerate clusters and k partitions.
- $l = l + 1$, $k = k + 1$
- Find $\alpha = \max_{G,H \in Z^{(k)}; G \neq H} s(G,H) = s(\Delta_i, \Delta_j)$. (If there is more than a pair of elements that verify this condition we select between them randomly).
- $\Delta_l = \Delta_i \bigcup \Delta_j$
- Calculating the membership function of the new cluster Δ_l

$$\mu_{\tilde{\Delta}_l}(P_j) = \frac{\sum_{A_{i \in \Delta_l}} \mu_{\tilde{A}_i}(P_j)}{|\Delta l|}$$

$$I = I - \{i, j\} \cup \{l\}, \quad Z^{(k)} = \bigcup_{i \in I} \{\Delta_i\}$$

- Updating similarities
  For $i \in I, i \neq l$ do $s(\Delta_i, \Delta_j) = s_g(\Delta_i, \Delta_j)$
- Repeat the process until $|I| = 1 (l = 2n - 1,$ $k = n - 1 or |Z^{(k)}| = 1)$

end of the algorithm.

The fuzzy relation is changed from a step to another step so that the maximum values of the similarity(or minimums of the distance) refer to different matrices and, in consequence, we do not have an order of the similarity levels as is the case for instance in the simple linkage method. We can order only the steps.

2.3. Comparison with other methods

The transitive closure by the minimum is the more appropriated method because of its powerful theoretical properties but sometimes do not give suitable clusters. In this case we test other methods. Apart from the method proposed in this article a very general procedure is given in [21]. We will refer to them as A and B methods respectively. These methodologies share and differ in the following points:

- In A is calculated the membership function of the new cluster at each step of the process so is necessary to define the inner fuzzy structure of the data in contrast with B where the similarity matrix can be given by experts.
- The complexity of the calculus lies in different factors. In A consists in calculating the fuzzy structure of the new cluster and its similarity with the other ones. On the other hand, in B is necessary to calculate the max-*t* transitive closure for an specific *t*-norm which can be expensive in time if the dimension of the matrix is quite big.

- B is based on that the $\max - t_{bp}$ composition is more meaningful and effective than the max-min composition because keeps the closet values from the original proximity relation so rely on the depth algebraic structure of the fuzzy relation. In contrast, A defines the fuzzy structure of the clusters by means of aggregating the fuzzy information of its elements.
- Both algorithms lose uniqueness but in A is quite unlikely that two clusters share exactly the same similarity value because of the changing of membership function at each step of the algorithm and the complexity of the calculus of the similarity.

3. Example

With the aim to clarify all the process by means an application we have chosen a complete geometrical example. Like this we have the great advantage of

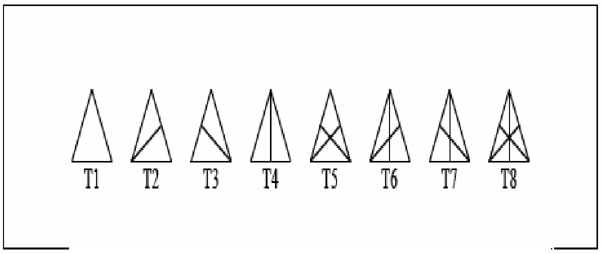

Figure 5: Equilateral triangles

checking the results in a unmistakable way. On the other hand, the fuzzy relation between objects and features becomes crisp. Calculations have been reduced to the minimum expression because they can be easily checked. Our objects are equilateral triangles with all the possible combinations of bisectors therefore $X = \{$ T1, T2, T3, T4, T5, T6, T7, T8 $\}$
One triangle has no bisectors, three have one bisector, three have two bisectors and one has three bisectors.
Our objective is to group objects more similar in the sense that they contain the same bisectors (not the same number of bisectors). Let b_i be the bisectors with positive, negative and vertical slope respectively. Let Y = $\{P_1, P_2, P_3\}$ be the set of features in which P_i expresses if b_i is contained or not. The whole set of the equilateral triangles is represented at Figure 5. The fuzzy relation $\tilde{S}$ between X and Y is determined by the matrix

$$S = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \end{array} \begin{array}{c} \begin{matrix} T1 & T2 & T3 & T4 & T5 & T6 & T7 & T8 \end{matrix} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix} \end{array}$$

As similarity index we have chosen a member of the family of generalized fuzzy indexes for $\lambda = \mu = 1$ That index is a fuzzy generalization of the simple matching coefficient and we call it s_{gsm}. Therefore

$$s_{gsm}(\tilde{A}, \tilde{B}) = \\ = \frac{|\tilde{A} \cap \tilde{B}| + |\tilde{A}^c \cap \tilde{B}^c|}{|\tilde{A} \cap \tilde{B}| + |\tilde{A}^c \cap \tilde{B}^c| + |\tilde{A} - \tilde{B}| + |\tilde{B} - \tilde{A}|} \quad (15)$$

In order to ensure the reflexive hypothesis cardinals are calculated by the *t*-norm of the bounded product which verifies the non-contradiction principle. Let $R = (r_{ij})$ be the matrix of similarities then, for instance,

$$|\tilde{T}_4 \cap \tilde{T}_7| = \sum_{i=1}^{3} \max(0, \mu_{\tilde{T}_4}(P_i) + \mu_{\tilde{T}_7}(P_i) - 1) = 1$$

and $|\tilde{T}_4 \cap \tilde{T}_7^c| = 0, |\tilde{T}_4^c \cap \tilde{T}_7| = 1 = |\tilde{T}_4^c \cap \tilde{T}_7^c|$.

Computing $s_{gsm}(\tilde{T}_4, \tilde{T}_7) = r_{47} = 2/3$. In a similar way we would determine the reminder elements of the fuzzy relation $\tilde{R}$ with associated matrix *A*, composed by all the similarities between the equilateral triangles, finally we obtain

$A =$

	T1	T2	T3	T4	T5	T 6	T7	T8
T1	1	0.67	0.67	0.67	0.33	0.33	0.33	0
T2	0.67	1	0.33	0.33	0.67	0.67	0	0.33
T3	0.67	0.33	1	0.33	0.67	0	0.67	0.33
T4	0.67	0.33	0.33	1	0	0.67	0.67	0.33
T5	0.33	0.67	0.67	0	1	0.33	0.33	0.67
T6	0.33	0.67	0	0.67	0.33	1	0.33	0.67
T7	0.33	0	0.67	0.67	0.33	0.33	1	0.67
T8	0	0.33	0.33	0.33	0.67	0.67	0.67	1

We will apply three methods in order to notice the differences. First, we will calculate the transitive closure by means the *t*-norm of the minimum that will lead to the same partition that the single linkage method for each α -level. Second, we will determine the transitive closure by means the *t*-norm of the bounded product and we will apply the method described in [21]. Finally, we will do the clustering process by the method exposed in subsection. 2.2.

3.1. Transitive closure by the *t*-norm of the minimum

As $0.33 = r_{51} < 0.67 = \max_{j=1...5} \min\{r_{5j}, r_{j1}\}$, results that $\tilde{R}$ is not min-transitive. We have to find the transitive closure with the powers' method: $A \neq A^2$, $A^2 \neq A^3$ but $A^3 \neq A^4$ so $A^* = A^4$

$$A^* = \begin{pmatrix} 1 & 0.67 & 0.67 & 0.67 & 0.67 & 0.67 & 0.67 & 0.67 \\ 0.67 & 1 & 0.67 & 0.67 & 0.67 & 0.67 & 0.67 & 0.67 \\ 0.67 & 0.67 & 1 & 0.67 & 0.67 & 0.67 & 0.67 & 0.67 \\ 0.67 & 0.67 & 0.67 & 1 & 0.67 & 0.67 & 0.67 & 0.67 \\ 0.67 & 0.67 & 0.67 & 0.67 & 1 & 0.67 & 0.67 & 0.67 \\ 0.67 & 0.67 & 0.67 & 0.67 & 0.67 & 1 & 0.67 & 0.67 \\ 0.67 & 0.67 & 0.67 & 0.67 & 0.67 & 0.67 & 1 & 0.67 \\ 0.67 & 0.67 & 0.67 & 0.67 & 0.67 & 0.67 & 0.67 & 1 \end{pmatrix}$$

We obtain the different partitions in function of the α -levels. If $0 < \alpha \leq 0.67$, only one cluster is formed: the whole set *X*. If $0.67 < \alpha \leq 1$, seven clusters are formed:

$\{T1\}, \{T2\}, \{T3\}, \{T4\}, \{T5\}, \{T6\}, \{T7\}$, and $\{T8\}$.It do not seem that for any α the results are acceptable because we group all the triangles in an unique object or we separe all the triangles in singletons. At this point is when seems logical to look for another kind of method.

3.2. Transitive closure by the *t*-norm of the bounded product

Now we check another possibility. We will apply a method proposed by M. S. Yang and H. M. Shih [21] using the *t*-transitive closure with the *t*-norm of the bounded product which is the smallest *t*-norm that verify the triangular inequality. We obtain $A = A^2$ SO $\tilde{R}$ is t_{bp} -transitive. This method depends on a selected α -level, we have chosen α = 0:55 and applying the algorithm, which do not verify uniqueness, we obtain the following results which are showed in the same nomenclature that in the original paper

$A_{0.55} = A^{(2)} =$

$$\begin{array}{c|cccccccc} & T1 & T2 & T3 & T4 & T5 & T6 & T7 & T8 \\ \hline T1 & 0 & & & & & & & \\ T2 & 0.67 & 0 & & & & & & \\ T3 & 0.67 & 0 & 0 & & & & & \\ T4 & 0.67 & 0 & 0 & 0 & & & & \\ T5 & 0 & 0.67 & 0.67 & 0 & 0 & & & \\ T6 & 0 & 0.67 & 0 & 0.67 & 0 & 0 & & \\ T7 & 0 & 0 & 0.67 & 0.67 & 0 & 0 & 0 & \\ T8 & 0 & 0 & 0 & 0 & 0.67 & 0.67 & 0.67 & 0 \end{array}$$

The maximum value is 0:67. Randomly we select $a_{21}^{(2)} = 0.67 \neq 0$; so $C = \{1,2\}$ is our first set of indexes, and $I - C = \{3,4,5,6,7,8\}$. As $a_{31}^{(2)} + a_{32}^{(2)}$, $a_{41}^{(2)} + a_{42}^{(2)}$, $a_{51}^{(2)} + a_{52}^{(2)}$, $a_{61}^{(2)} + a_{62}^{(2)}$, $a_{71}^{(2)} + a_{72}^{(2)}$, and $a_{81}^{(2)} + a_{82}^{(2)}$, have some element equal to 0 therefore our first cluster is $\Delta_1 = \{T1, T2\}$. Following the same procedure we find $\Delta_2 = \{T3, T5\}$, $\Delta 3 = \{T4, T6\}$ and $\Delta 4 = \{T7, T8\}$. We achieve a partition in which the elements with more differences have not been grouped. These results are better than those of the transitive closure (or single linkage method).

3.3. Method exposed at subsection 2.2.

At the first step for all $i = 1,...,8$ $\Delta_i = \{\mathbf{Ti}\}$ so $Z^{(0)}$ is made up of all the singletons. As some objects attain the maximum value (0.67) we choose between them randomly, for instance $s_{78} = 0.67$ and we put them together so $\Delta_9 = \{T7, T8\}$. At this moment

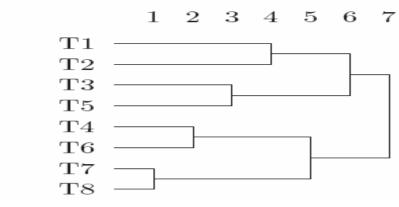

Figure 6: Dendogram of the example

we have to calculate the similarity between this new clusters and the others by means (15). The membership function of $\tilde{\Delta}_1$ is defined by the values $\mu_{\tilde{\Delta}_1}(P_1) = 1/2$, $\mu_{\tilde{\Delta}_1}(P_2) = 1$ and $\mu_{\tilde{\Delta}_1}(P_3) = 1$.

Following the same methodology we obtain the new matrix of similarities which is one dimension less that the preceding one

$$S_1 = \begin{array}{c} \\ T1 \\ T2 \\ T3 \\ T4 \\ T5 \\ T6 \\ \Delta_1 \end{array} \begin{array}{c} \begin{array}{ccccccc} \mathrm{T1} & \mathrm{T2} & \mathrm{T3} & \mathrm{T4} & \mathrm{T5} & \mathrm{T6} & \Delta_1 \end{array} \\ \begin{pmatrix} 1 & 0.67 & 0.67 & 0.67 & 0.33 & 0.33 & 0.17 \\ 0.67 & 1 & 0.33 & 0.33 & 0.67 & 0.67 & 0.17 \\ 0.67 & 0.33 & 1 & 0.33 & 0.67 & 0 & 0.5 \\ 0.67 & 0.33 & 0.33 & 1 & 0 & 0.67 & 0.5 \\ 0.33 & 0.67 & 0.67 & 0 & 1 & 0.33 & 0.5 \\ 0.33 & 0.67 & 0 & 0.67 & 0.33 & 1 & 0.5 \\ 0.17 & 0.17 & 0.5 & 0.5 & 0.5 & 0.5 & 1 \end{pmatrix} \end{array}$$

The maximum value is 0:67. We build $\Delta_{10} = \{\mathrm{T4}, \mathrm{T6}\}$. For the next steps we find $\Delta_{11} = \{\mathrm{T3}, \mathrm{T5}\}$, $\Delta_{11} = \{\mathrm{T3}, \mathrm{T5}\}$ $\Delta_{13} = \Delta_5 \cup \Delta_2 = \{\mathrm{T4}, \mathrm{T6}, \mathrm{T7}, \mathrm{T8}\}$, $\Delta_{14} = \Delta_3 \cup \Delta_4 = \{\mathrm{T1}, \mathrm{T2}, \mathrm{T3}, \mathrm{T5}\}$, and finally, $\Delta_{15} = X$. For instance, the partition at the third step is .

We represent all these results in a dendogram in Figure 6. In the vertical axe there are represented the objects that we want to group. It the horizontal axe there is represented the different levels at which the objects are merged. As we have formerly commented we do not have an increasing set of levels because the reference set have changed from one step to another step. Notice that the results are quite similar to the method presented by M. S. Yang and H. M.Shih [21].

4. Conclusions

We consider that further researches should focus in designing a great sample of experiments which could give statistical results about the convenience of selecting one of the different procedures related to the different proposed conditions in the sample.

References

[1] J. C. Bezdek, *Fuzzy mathematics in pattern classification*, Ph. D. Thesis, Cornell U, New York,1973.

[2] J. C. Bezdek, “Numerical taxonomy with fuzzy set,” Journal Mathematics Biology, vol. 1, no. 1, pp. 57-71, 1974.

[3] J. C. Bezdek, J. Keller, R. Krisnapuram, N. R. Pal, *Fuzzy Models and Algorithms for pattern recognition and Image Processing*, Kluwer Academic Publishers, Boston, 1999.

[4] D. Boixader, J. Jacas, J. Recasens, “Transitive closure and betweenness relations,” *Fuzzy Sets and Systems*, vol. 120, pp. 415-422, 2001.

[5] N. Clara, *Generalized fuzzy similarity indexes*, Neural Nets, Lecture Notes in Computer Science, vol. 3931, pp. 163-170, Springer, Berlin, 2006.

[6] D. Dubois, and H. Prade, *Fuzzy Sets and Systems, Theory and Applications*, Academic Press, San Diego, 1980.

[7] J. C. Dunn, “A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters,” *J. Cyberns*, vol. 3, no. 3, pp. 32-57, 1974.

[8] A. Flores-Sintes, J. M. Cadenas, and F. Martin, F. “Detecting homogeneous groups in clustering using the Euclidean distance,” *Fuzzy Sets and Systems*, vol. 120, pp. 213-225, 2001.

[9] P. J. F. Groenen, and K. Jajuga, “Fuzzy clustering with squared Minkowski distances,” *Fuzzy Sets and Systems*, vol. 120, pp. 227-237, 2001.

[10] H. Kamimura, and M. Kurano, “Clustering by a fuzzy metric,” *Fuzzy Sets and Systems*, vol. 120, pp. 249-254, 2001.

[11] G. J. Klir, and T. A. Folger, *Fuzzy Sets, Uncerttainty, and Information*, Prentice Hall, New Jersey, 1988.

[12] R. Krishnapuram, and J. M. Keller, “A possibilistic approach to clustering,” *IEEE Trans, Fuzzy Systems*, vol. 1, no. 2, pp. 98-110, 1993.

[13] S. Kundu, “Similarity relations, fuzzy linear orders, and fuzzy partial orders,” *Fuzzy Sets and Systems*, vol. 109, pp. 419-428, 2000.

[14] H-S. Lee, “An optimal algorithm for computing the max-min transitive closure of a fuzzy similarity matrix,” *Fuzzy Sets and Systems*, vol. 123, pp. 129-136, 2001.

[15] R. Lowen, *Fuzzy Set Theory*, Kluwer Academic Publishers, Dordrecht, 1996.

[16] S. Miyamoto, *Fuzzy Sets on Information Retrieval and Cluster Analysis*, Kluwer Academic Publishers, Dordrecht, 1990.

[17] P. K. Simpson, “Fuzzy min-max neural networks - Part 2: Clustering,” *IEEE Trans, Fuzzy Systems*, vol. 1, no. 1, pp. 32-45, 1993.

[18] Y. A. Tolias, S. M. Panas, and L. H. Tsoukalas, “Generalized fuzzy indices for similarity matching,” *Fuzzy Sets and Systems*, vol. 120, pp. 255-270, 2001.

[19] A. Tversky, “Features of similarity,” *Pshicology Review*, vol. 34, pp. 286-377, 1977.

[20] M. S. Yang, “A survey of fuzzy clustering,” *Math. Comput. Modelling*, vol. 18, no. 11, pp. 1-16, 1993.

[21] M. S. Yang, and H. M. Shih, “Cluster analysis based on fuzzy relations,” *Fuzzy Sets and Systems*, vol. 120, pp. 197-212, 2001.