# Digestive Database Evidential Clustering Based on Possibility Theory

Anas DAHABIAH, John PUENTES, and Basel SOLAIMAN
TELECOM Bretagne, Département Image et Traitement de l'Information, Brest, France
INSERM, U650, Laboratoire de Traitement de l'Information Médicale, Brest, France
{anas.dahabiah, john.puentes, basel.solaiman}@telecom-bretagne.eu
http://www.telecom-bretagne.eu

*Abstract:* - A new method that aims to automatically classify a set of objects in spite of the imperfection and the uncertainty of their heterogeneously-assigned data is proposed in this paper. This method is based essentially on possibility theory to estimate the similarity among the objects, and on belief theory and multidimensional scaling methods to construct the compatible evidential class partition. This method is applied to a medical database and robust results have been obtained without knowing the key attributes of the concerned pathologies and without taking into account any a priori medical knowledge.

*Key-Words:* - Possibility Theory, Evidence Theory, Similarity, Clustering, Partition, Multi-Dimensional Scaling.

## 1 Introduction

Clustering is the unsupervised classification of patterns (observations, data items, examples or features) into groups (clusters). The clustering problem has been addressed in several contexts in many disciplines. This reflects its broad appeal and usefulness in exploratory data analysis. However, all the proposed clustering methods depend mainly on measuring the similarity and suppose that the similarity of the objects is already calculated, in spite of the fact that it is sometimes difficult to estimate it, especially in the presence of heterogeneous and imperfect data. In this paper we propose a general clustering method as follows: given a set of "$c$" classes $\Omega = \{\omega_1, \omega_2, ..., \omega_c\}$ and a set of "$n$" objects $O = \{O_1, O_2, ..., O_n\}$ where each object $O_i$ from $O$ consists of "$S$" attributes $\{x_{i1}, x_{i2}, ..., x_{iS}\}$, in such a way that these attributes could be heterogeneous (quantitative, binary, qualitative, ordinal, etc.) and could have imperfect values (imprecise and vague assignments, missing values, etc.). We aim to assign to each element of $O$ its appropriate class. For this purpose we propose the approach schematized in figure 1. This approach depends basically on possibility theory and on evidential clustering. Section 2 presents the theory of possibilities and illustrates the main steps to construct the dissimilarity matrix, whereas section 3 presents the theory of belief and summarizes the main phases to construct the evidential partition in which we aim to allocate to each object a basic belief assignment of its membership to all the possible sets of classes in such a way that the conflict degree between the masses given to any two objects reflects their dissimilarity presented in the proximity matrix. Then, the construction of the fuzzy and hard partition is described in section 4. This approach is applied to a medical database and the results presented in section 5, show the capability to distinguish the different categories of digestive pathologies, depending only on the values of the attributes, without any previous knowledge. Furthermore, this approach outperforms the prior works as we will see in section 6. Then, some discussions and conclusions will be presented in section 7. After that, a very important application of this approach in data fusion will be presented as a perspective in section 8.

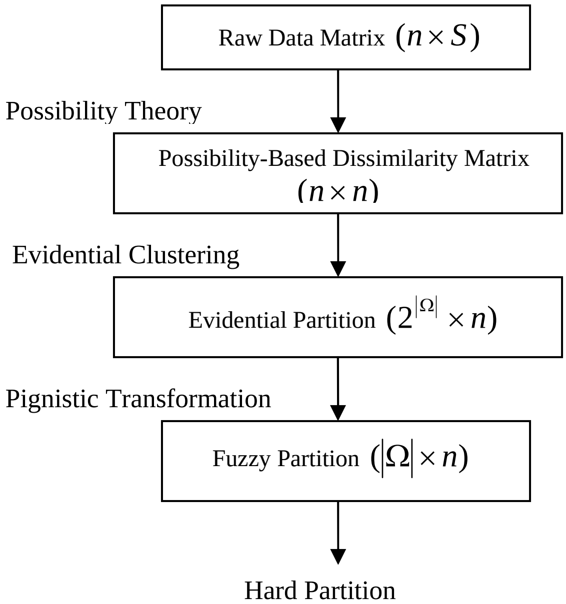


Fig.1 Proposed approach outline.

## 2 From Raw Data Matrix to Dissimilarity Matrix

Data to be analyzed are commonly presented in one of two different formats: as a raw data matrix or as a dissimilarity matrix. The raw data matrix is an $n \times S$

matrix $X \equiv (x_{ik})$, where $x_{ik}$ denotes the value of the $k^{th}$ variable observed for the $i^{th}$ object. Many data mining techniques as the evidential clustering, first require the transformation of the raw data matrix into an $n \times n$ matrix of pairwise dissimilarities (distances) $D = (\delta_{ij})$, where $\delta_{ij}$ denotes the dissimilarity between the $i^{th}$ and $j^{th}$ objects. Because of the defects and the limits of the traditional similarity measures presented briefly in the following subsection, we propose to use the possibility theory presented in subsection 2-2 to build the dissimilarity matrix *D*. The construction of the dissimilarity matrix is explained in subsection 2-3 supported with a concrete example in subsection 2-3-1.

## 2.1 Traditional Similarity Measures Limits

Traditional similarity (dissimilarity) measures (Minkowski, Canberra, Hamming, Jaccard, etc.) [1] suppose generally that the value of each attribute is precise (disregarding the existence of imprecise data), certain (disregarding the existence of uncertain values), and given (disregarding the existence of missing values) while on the contrary, real databases contain a remarkable amount of incomplete and imperfect values. Actually, the uncertainty of data is a delicate widespread problem in many domains. For instance in the medical domain, patients can not describe exactly how they feel or what has happened to them, doctors and nurses can not tell exactly what they observe, laboratories report results only with some degree of errors, physiologists don't precisely understand how the human body works, medical researchers can not precisely characterize how diseases alter the normal functioning of the body, pharmacologists don't fully understand the mechanism accounting for the effectiveness of the drugs, and no one can precisely determine one's prognosis [13]. In addition, some constraints and conditions should be considered when dealing with each similarity measure. For instance, division by zero could take place in a considerable amount of these measures, besides the need to know the nature of each variable in the records that contain heterogeneous attributes (quantitative, qualitative, ordinal, etc.) in order to choose a suitable measure. Moreover, the similarity interval should be taken into account during the aggregation and during the interpretation of the resulting value ([0,1] is the most common similarity interval usually proposed, even though some measures like the angular separation similarity belong to [-1,1]). In reality, a value of an attribute can be given in different ways. For example, if we examine the value of the attribute "age", in some patient records "age" could be assigned as {18 yeas, close to 18 years, more than 15 years, young, between 15 and 20, unknown, 18 or 19, it's quite possible to be 18 or 19 and somehow possible to be 17 or 20, defined by a probability distribution, etc.}. Similarity calculation according to the traditional measures can not be easily carried out between two heterogeneous values, for example, between a value given as 25 and another value given as close to 25, or as a probability distribution, whereas these assignments can be modeled easily in possibility theory [19-20]. For these reasons and in order to construct a general approach, we don't recommend the use of the traditional measures overburdened with a lot of conditions and constraints. Instead, we propose to use the possibility theory measures developed by Zadeh, Prade, Dubois, and Rakoto [2], [4-5], and [23] in order to build the similarity (dissimilarity) matrix among the objects of our set.

## 2.2 Possibility Theory

Possibility theory provides a method to formalize subjective uncertainties of events, that is to say a means of assessing to what extent the occurrence (the realization) of an event is possible and to what extent we are certain of its occurrence, without having however the possibility to measure the exact probability of this realization because we don't know an analogous event to be referred to, or because the uncertainty is the consequence of observation instrument reliability absence. Let's attribute to each event defined on the universe of discourse $\Omega$ (in other words to each element belonging to $\rho(\Omega)$) a coefficient ranging between 0 and 1 assessing to which degree the occurrence of an event is possible, where the value "1" means that the event is completely possible, while the value "0" means that the event is impossible. To define this coefficient, we introduce the possibility measure $\Pi$ which is a function defined over $\rho(\Omega)$, taking values in $[0,1]$, such that:

Axiom 1: $\Pi(\phi) = 0$ (1)

Axiom 2: $\Pi(\Omega) = 1$ (2)

Axiom3: $\forall A_1, A_2, ... \in \rho(\Omega)$

$\Pi(\cup_{i=1,2,..} A_i) = SUP_{i=1,2,...} \Pi(A_i)$ (3)

where SUP indicates the supremum of the concerned values.

We can say that the possibility measure is totally defined, if we can attribute a possibility coefficient to all the singletons of $\Omega$. Consequently, the possibility distribution function $\pi$ defined on $\Omega$, whose values are included in $[0,1]$, such that $\sup_{x \in \chi} \pi(x) = 1$ must be defined. As a result the function $\Pi$ can be defined form the function $\pi$ by:

$\forall A \in \rho(\Omega)\ \ \Pi(A) = \sup_{x \in A} \pi(x)$ (4)

Reciprocally, $\pi$ can be defined form $\Pi$ by:

$$\forall x \in \Omega \quad \pi(x) = \Pi(\{x\}) \tag{5}$$

We should also mention here that the characteristic function of a subset from $\Omega$ can be considered as a possibility distribution $\pi$ defined on $\Omega$. To calculate the possibility degree of the couple $(x, y)$ given that $x \in \Omega_1$ and $y \in \Omega_2$ where $\Omega_1, \Omega_2$ are two non-interactive universes of discourse, the conjoint possibility distribution defined on the Cartesian product $\Omega_1 \times \Omega_2$ should be calculated from:

$$\forall x \in \Omega_1 \ \forall y \in \Omega_2 \quad \pi(x, y) = \min(\pi_\chi(x), \pi_\gamma(y)) \tag{6}$$

In fact, the possibility measure is not sufficient to describe the incertitude of the realization of an event, because sometimes the realization of both the event *A* and its complement $A^C$ could be completely possible simultaneously ($\Pi(A) = 1$ and $\Pi(A^C) = 1$ at the same time). This means that in this particular case it is impossible to take a decision concerning the realization of *A* depending on the estimated possibility measure. For this reason, another function, defined on $\rho(\Omega)$, whose values are included in $[0,1]$ and which is called the necessity measure (denoted *N*) is defined as follows:

Axiom 1: $N(\phi) = 0$ (7)

Axiom 2: $N(\Omega) = 1$ (8)

Axiom 3: $\forall A_1 \in \rho(\Omega) \quad \forall A_2 \in \rho(\Omega)$

$$N(\cap_{i=1,2,....} A_i) = INF_{i=1,2,...} N(A_i) \tag{9}$$

where INF stands for infimum.

### 2.2.1 Relations between Possibility and Necessity Measures

There are some interesting relations between the possibility measure $\Pi$ and the necessity measure $N$ presented in the following equations:

$$\forall A \in \rho(\Omega) \ N(A) = 1 - \Pi(A^C) \tag{10}$$

$$\forall A \in \rho(\Omega) \ N(A) = INF_{x \notin A}(1 - \pi(x)) \tag{11}$$

$$\Pi(A) \geq N(A) \tag{12}$$

$$Max(\Pi(a), 1 - N(A)) = 1 \tag{13}$$

If $N(A) \neq 0$ then $\Pi(A) = 1$ (14)

If $\Pi(A) \neq 1$ then $N(A) = 0$ (15)

$$N(A) \leq \Pr(A) \leq \Pi(A) \tag{16}$$

Where $\Pr(A)$ stands for the probability of any event $A \in \rho(\Omega)$.

### 2.3 Possibility-Based Similarity Estimation

Suppose that we have two objects $O_j$ and $O_k$ containing "S" attributes:

$$O_j = [x_{1j} \quad x_{2j} \quad . \quad . \quad x_{ij} \quad . \quad . \quad x_{Sj}]$$
$$O_k = [x_{1k} \quad x_{2k} \quad . \quad . \quad x_{ik} \quad . \quad . \quad x_{Sk}].$$

Each attribute could take a precise or an imprecise value modeled by its possibility distribution, and this value can be either numerical or nominal. The values of some attributes could be unassigned (missing value). Besides, each attribute is associated with a "tolerance function" defined by an expert as a formula or as a table permitting to describe mathematically to which degree we consider that two values of this attribute are similar. An example of tolerance function is the function that we call "close to". Such a function can be defined by the following formula:

$$\mu_a(a_x, a_y) = 1 - \frac{|a_x - a_y|}{\Delta} \text{ if } |a_x - a_y| \leq \Delta \tag{17}$$
$$\mu_a(a_x, a_y) = 0 \text{ Otherwise}$$

Where $\Delta$ is a variable that influences the slope of the function and consequently the notion of "close to". The tolerance function can be also:

- The function of tolerance "True/false": two values of an attribute are similar if they are identical (similarity equals to 1). If the values are different, the similarity is null, this type of functions is used especially when dealing with nominal variables having independent categories. In the case of ordinal variables we must use the function "close to".
- The "ad hoc" tolerance functions which are defined by the experts to reflect their point of view about the similarities between the attributes.

In our approach the similarity between the two objects $O_j$ and $O_k$ can be estimated by means of two measures: the possibility degree of similarity between $O_j$ and $O_k$ that tells us to which degree it is possible that these vectors are similar, and the necessity degree of similarity of these vectors that tells us to which degree we are certain of their similarity. The probability of the similarity between $O_j$ and $O_k$ exists between the necessity degree that represents the lower limit and the possibility degree that represents the upper limit. To calculate the possibility and the necessity degrees of resemblance, we must calculate the local possibility and necessity degrees between their corresponding attributes and aggregate them by taking their average, for example in order to take a decision concerning the total similarity. The local possibility and necessity degrees of similarity between $x_{ij}$ given by its possibility distribution $\pi_{X_j, x_{ij}}(x_{ij}, y)$ and $x_{ik}$ given by its possibility distribution $\pi_{X_k, x_{ik}}(x, x_{ik})$ for all $i \in \{1,2,...,S\}$ are calculated according to the following relations:

Supposing that $D$ is the definition domain of the considered attribute $x_i$ ($U = D \times D$) and that $\mu$ is the tolerance function associated to this attribute, the conjoint possibility distribution $\pi_D$ is calculated as:

$$\pi_D(x_{ij}, x_{ik}) = \min(\pi_{X_j, x_{ij}}(x), \pi_{X_k, x_{ik}}(y)) \quad (18)$$

In this case, the local possibility degree of similarity $\pi_i$ can be calculated as:

$$\pi_i(x_{ij}, x_{ik}) = SUP_{u \in U}[\min(\mu(u), \pi_D(u))] \quad (19)$$

The local necessity degree of similarity $N_i$ can be calculated as:

$$N_i(x_{ij}, x_{ik}) = INF_{u \in U}[\max(\mu(u), 1 - \pi_D(u))] \quad (20)$$

We consider that if the value of an attribute is given in one object and is unassigned in the other (the case of missing values), it is completely possible that these values are similar $\pi_i = 1$ but we are entirely uncertain $N_i = 0$. Now that the local possibility and necessity degrees are calculated between the attributes, the global possibility and necessity degrees between the objects can be calculated by averaging the local degrees. The average possibility $\Pi_{jk}$ and the average necessity $N_{jk}$ are calculated from the following equations:

$$\Pi_{jk} = \sum_{i=1}^{S} \pi_i \Big/ S \quad (21)$$

$$N_{jk} = \sum_{i=1}^{S} N_i \Big/ S \quad (22)$$

### 2-3-1 Concrete Example of Possibilistic Similarity Estimation

Suppose that we would like to calculate the similarity between two patient records in a medical database. Each record contains patient's age, sex, weight, symptoms, biological analysis ...etc. The values of these attributes could be imprecise, vague, uncertain, or unassigned. In all the cases, these values can be easily modeled by possibility distributions. Actually, even if the value of an attribute was assigned as a probability distribution, we are able to transform it to a possibility distribution by applying Prade-Dubois transformation rule [19-20]. For each attribute, we calculate the possibility degree and the necessity degree of similarity between its assigned values in the first and in the second record. We call these degrees "local degrees" since they are estimated at the attribute level. The average degree of all the local degrees calculated between all the considered attributes of the record is called the global degree of similarity between the records. Let us make things easier by taking numeric values, for this purpose we will take the attribute "age" in the patient record, and will suppose that we consider that the values of two ages are considered similar if the difference between them doesn't exceed ten years old. In other terms, we take the tolerance function (equation 17) whose $\Delta = 10$ (see figure 2). Let us suppose also that the age is assigned in the first record as "is about 40" and in the second record as "is about 50" (see figure 3 in which the value of each age has been modeled by a fuzzy number $\pm 10$).

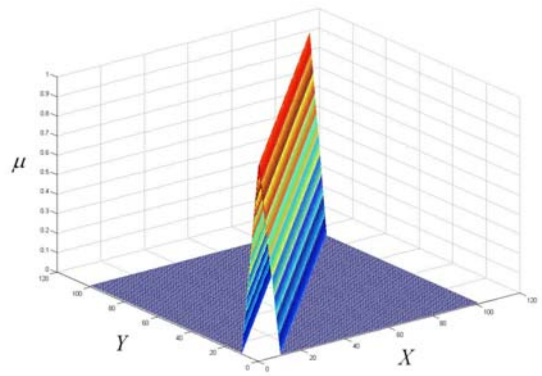


Fig.2 The tolerance function for $\Delta = 10$.

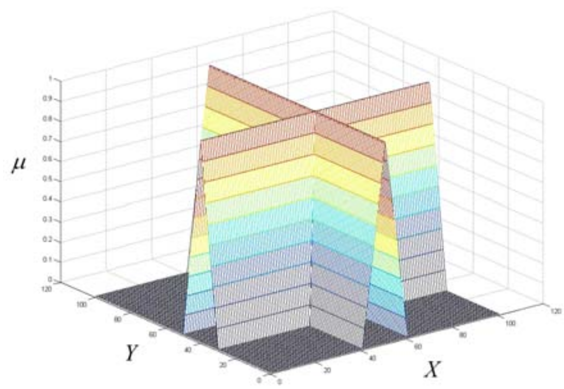


Fig.3 The two ages modeled by fuzzy numbers.

To estimate the local possibility and necessity degrees of similarity, we apply the steps presented in section 2-3 as follows: the conjoint possibility distribution that represents the intersection between the two modeled values of the attribute “age” is calculated using equation 18 (figure 4). The maximum value of the intersection between the tolerance function and the conjoint distribution represents the possibility degree of similarity $\Pi$ (figures 5 and 6). For the values given in this example we find that $\Pi = 0.60$. Then, we use equation 20 to calculate the necessity degree of similarity (figure 7). We find that $N = 2.22 \times 10^{-16}$. Table 1 shows the local possibility and the necessity degrees of similarity of the attribute “age” for other values of $\Delta$.

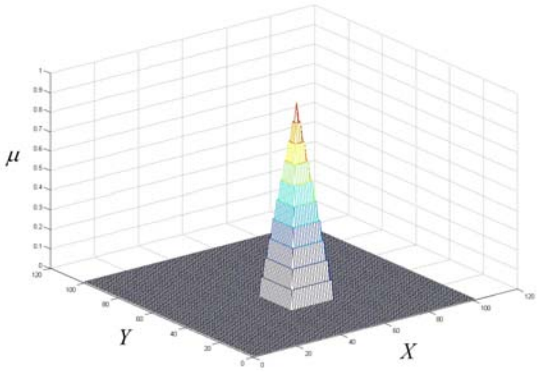

Fig.4 The conjoint possibility distribution.

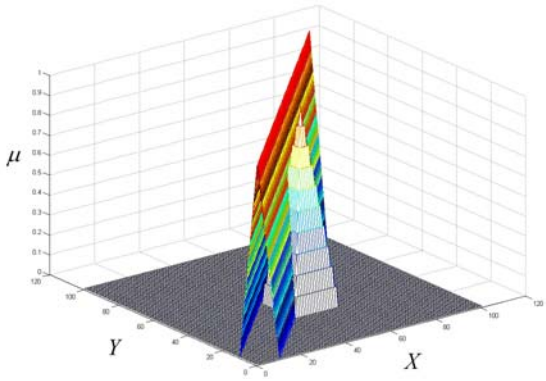

Fig.5 The intersection between the tolerance function and the conjoint possibility distribution.

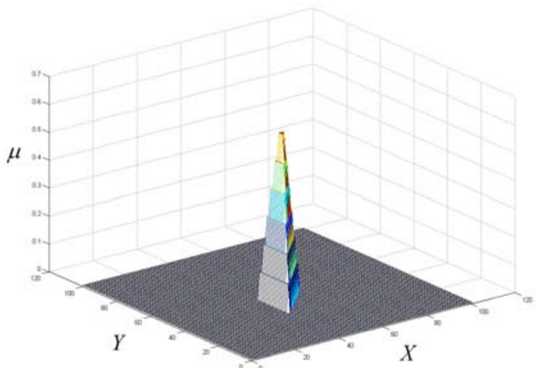

Fig.6 Possibility degree estimation.

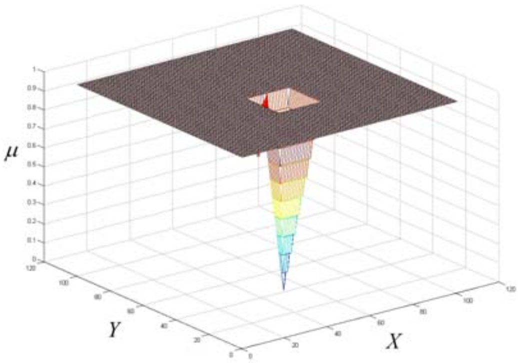

Fig.7 Necessity degree estimation.

| $\Delta$ | $\Pi$ | $N$ |
|---|---|---|
| 20 | 0.70 | 0.30 |
| 30 | 0.80 | 0.40 |
| 50 | 0.84 | 0.60 |
| 99 | 0.90 | 0.76 |

Table 1 Possibility and necessity degrees of similarity of the attribute “age” for different values of $\Delta$

We apply the same steps to all the other attributes of the records taking into account that $\Pi_i = 1$ and $N_i = 0$ if the value of an attribute is assigned in a record and is a missing value in the other record, and that $\Pi_i = 0$ and $N_i = 0$ if the attribute exists in a record and doesn’t exist in the other. This can take place in the databases whose records come from different sources (hospitals) because the attributes of the records that come from a hospital can not be exactly the same as those which come from another one even if all the records characterize the same pathology.

## 3 From Dissimilarity Matrix to Evidential Partition

Belief theory provides a method to model and to quantify the credibility assigned to events for which we ignore the probability of occurrence, by means of its belief functions that introduce degrees in reliability assigned to these events [2-3] and [23]. Let us consider the domain of reference $\Omega$ over which belief coefficients are determined. These coefficients are obtained by distributing a global mass of belief equal to 1 to all the possible events, according to our belief in their occurrence. A Basic Belief Assignment (BBA) on $\Omega$, also called a mass of belief is any function *m* that assigns a coefficient between 0 and 1 to the different parts of $\Omega$ such that:

$$m(\phi) = 0 \tag{23}$$

$$\sum_{A \in \rho(\Omega)} m(A) = 1 \quad (24)$$

Any non-empty part $E$ of $\Omega$ such that $m(E) \neq 0$ is called a focal element or focal proposition.

For any event *A*, we can collect testimonies in its favor and determine the "belief function" ($Bel$) which corresponds to it, i.e. the sum of masses of belief of the focal elements which involve *A*:

$$Bel(A) = \sum_{E \subseteq A} m(E) \quad (25)$$

We can also determine its "Plausibility function" ($Pl$) by taking into account all the focal elements related to *A*, i.e. the sum of the masses of belief of the focal elements which are related to *A* and which make its occurrence possible, as:

$$Pl(A) = \sum_{E \cap A \neq \phi} m(E) \quad (26)$$

Let us now assume that we have two BBAs $m_1$ and $m_2$ representing distinct items of evidence. The standard way of combining them is through Dempster's rule of combination:

$$(m_1 \oplus m_2)(A) = \frac{1}{1-K} \sum_{B \cap C = A} m_1(B) m_2(C) \quad (27)$$

$$K = \sum_{B \cap C = \phi} m_1(B) m_2(C) \quad (28)$$

*K* is called the degree of conflict between $m_1$ and $m_2$ and it represents the degree of disagreement between the two information sources. From a robust dissimilarity matrix we can build an evidential partition $M \equiv (m_{kl})$, $1 \leq k \leq 2^{|\Omega|}$, $1 \leq l \leq n$ using evidential clustering method, where $m_{kl}$ denotes the basic belief of assigning the $l^{th}$ object "$O_l \in O$" to the $k^{th}$ subset of $\rho(\Omega)$. For example if $O = \{O_1, O_2, O_3\}$ and $\Omega = \{\omega_1, \omega_2\}$, then:

$$M = \begin{bmatrix} m_1(\phi) & m_2(\phi) & m_3(\phi) \\ m_1(\{\omega_1\}) & m_2(\{\omega_1\}) & m_3(\{\omega_1\}) \\ m_1(\{\omega_2\}) & m_2(\{\omega_2\}) & m_3(\{\omega_2\}) \\ m_1(\Omega) & m_2(\Omega) & m_3(\Omega) \end{bmatrix} \quad (29)$$

## 3-1 Evidential Clustering

For two BBAs $m_i$ and $m_j$ quantifying one's beliefs regarding the class of two objects $O_i$ and $O_j$, we can combine the vacuous extensions of $m_i$ and $m_j$ in the Cartesian product $\Omega^2 = \Omega \times \Omega$ using:

$$m_{i \times j}(A \times B) = m_i(A) m_j(B) \quad (30)$$
$$\forall A, B \subseteq \Omega \;\; A \neq \phi \;\; B \neq \phi$$

$m_{i \times j}(A \times B)$ is the BBA that describes one's beliefs concerning the class membership of both objects. In $\Omega^2$, "$O_i$ and $O_j$ belong to the class" corresponds to the following subset of $\Omega^2$:

$$S = \{(\omega_1, \omega_1), (\omega_2, \omega_2), ...., (\omega_c, \omega_c)\}.$$

Let $Pl_{i \times j}$ be the plausibility function associated with $m_{i \times j}$. We have [3]:

$$Pl_{i \times j}(S) = \sum_{\{A \times B \subseteq \Omega^2 / \;\; (A \times B) \cap S \neq \phi\}} m_{i \times j}(A \times B) = \sum_{A \cap B \neq \phi} m_i(A) m_j(B)$$
$$Pl_{i \times j}(S) = 1 - \sum_{A \cap B = \phi} m_i(A) m_j(B) = 1 - K_{ij} \quad (31)$$

Given any two pairs of objects $(O_i, O_j)$ and $(O_{i'}, O_{j'})$ it's natural that if $\delta_{ij} \succ \delta_{i'j'} \Rightarrow Pl_{i \times j}(S) \leq Pl_{i'j'}(S)$ or equivalently: $\delta_{ij} \succ \delta_{i'j'} \Rightarrow K_{ij} \geq K_{i'j'}$ i.e. the more dissimilar the objects, the less plausible it is that they belong to the same class and the higher is the conflict between the BBAs. According to Denoeux [3] we can construct *M* easily from *D* as follows: each object $O_i$ in M can be represented as a point in a $2^{|\Omega|}$ – dimensional space and the degree of conflict $K_{ij}$ between two BBAs may be seen as a form of "dissimilarity" between $O_i$ and $O_j$. Therefore we can transpose multidimensional scaling algorithms (MDS) to our problem [3] and [12], by optimizing the evidential partition *M* so that the degrees of conflict $K_{ij}$ reflect the corresponding dissimilarity $\delta_{ij}$. The objective function to be minimized is given by:

$$J(M, f) = I(M, f) + \lambda \sum_{i=1}^{n} H(m_i) \quad (32)$$

$$I(M,f)=\sum_{i\prec j}\left[K_{ij}-f(\delta_{ij})\right]^2 \Big/ \sum_{i\prec j}\left[K_{ij}-\bar{K}\right]^2 \qquad (33)$$

$$H(m)=\sum_{A\in F(m)} m(A)\log_2\left(\frac{|A|}{m(A)}\right) \qquad (34)$$

$I(M,f)$ is the stress function used in the ordinal MDS where $\bar{K}$ is the average degree of conflict and $f$ is any increasing function. *H(m)* is the entropy function (as we would like to extract as much information as possible from the data it is reasonable to require the BBAs to be as informative as possible). Actually, *H(m)* tends to be small when the mass is assigned to few focal sets with small cardinality; $\lambda$ is the penalization coefficient that controls the extent to which the entropy term influences the form of the solution. Increasing $\lambda$ for instance, will result in simpler BBAs with a smaller number of focal sets.

## 4 From Evidential Partition to Fuzzy and Hard Partition

A fuzzy partition may be obtained from the evidential partition by calculating the Pignistic probability function $p_m(\omega_i)$ induced by each BBA $m_i$ and interpreting it as the membership degree:

$$p_m(\omega_i)=\sum_{\{A\subseteq\Omega/\omega_i\in A\}}\frac{m(A)}{|A|} \quad i=1,2,\dots,C \qquad (35)$$

A hard partition can then be easily obtained by assigning each object to the class with the highest Pignistic probability.

## 5 Experiments and Results

Our approach was applied to a medical database concerning the upper gastrointestinal tract (esophagus, stomach, and duodenum), where each object $O_i$ consists of 24 attributes with 145 modalities (or 33 attributes with 206 modalities if a sub-object exists) describing the lesions and the pathologies that the object contains [11]. The characterization of the digestive images has been provided by the same medical expert by means of an interactive qualitative description interface (see figure 8). The attributes of this base could be qualitative, quantitative, or unevaluated (missing values), but they take always precise values. The database contains the following pathologies: Dilated lumen, Stenosis, Extrinsic compression, Web, Ring, Hiatal hernia, Food, Liquid Blood, Blood clot, z-line, spot, Circular Barrett's, Moniliasis, Simple erosion, Ulcer (edge), and Petchial mucosa. For more information about this base, see [6-10].

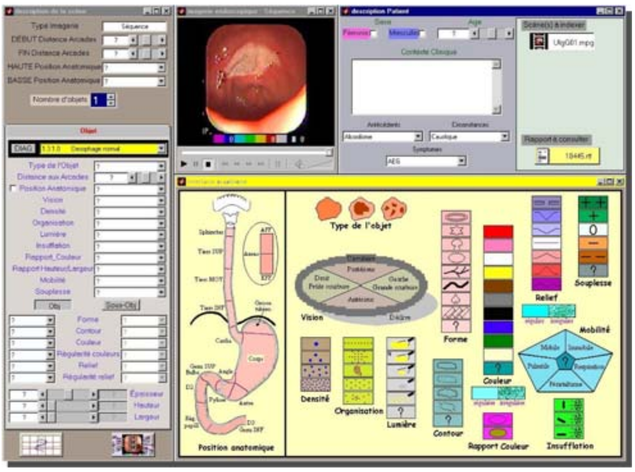

Fig.8 Image description interactive interface.

In our tests the similarity was modeled by the global necessity degrees between the objects stored in a matrix denoted as *X*. For simplicity and clarity graphic representation of our result, we will give as examples subsets of the main data subsets (submatrices of *X*), without loosing the generality of our approach: suppose that $P_1=\{O_1,O_2,\dots,O_8\}$ is the set of the objects whose pathology class is "stenosis (esophagus)", $P_2=\{O_9,O_{10},O_{11},O_{12}\}$ is the set of the objects whose pathology class is "extrinsic compression", $P_3=\{O_{13}\}$ represents the class "web shape", $P_4=\{O_{14},O_{15},O_{16}\}$ represents the class "ring", $P_5=\{O_{17},O_{18}\}$ contains the objects whose class is "Hernia" (see figure 9). Figure 10 represents the partition obtained by applying the steps illustrated throughout this paper to the subset $S_1=\{P_1,P_2\}$ whereas figure 11 presents the partition of $S_2=\{P_1,P_2,P_3,P_4,P_5\}$. In fact, we can construct any combination of sets and depict the corresponding partition and the result will be as expected. For instance, figure 12 depicts the partition of $S_3=\{P_1,P_2,P_4\}$. Contrary to the other clustering methods that don't take in consideration the pre-processing of the imperfect data disregarding the estimation of similarity, which is essential and delicate in the clustering process, our approach achieves the clustering taking into account all the basic phases demanded in the process in a very simple and clear manner, and the results are always as expected. Besides, this approach can be used in text-based image retrieval as well as in clustering. In fact, image retrieval has been mainly studied based on image content using primitives [24]: color, shape, detected contours, texture, transformation coefficient, etc.

In content-based image retrieval there are not imperfect or missing values in the extracted features, while to our knowledge this is the first study that takes into account the imperfect descriptive features of the medical images characterized directly by the doctors or the experts.

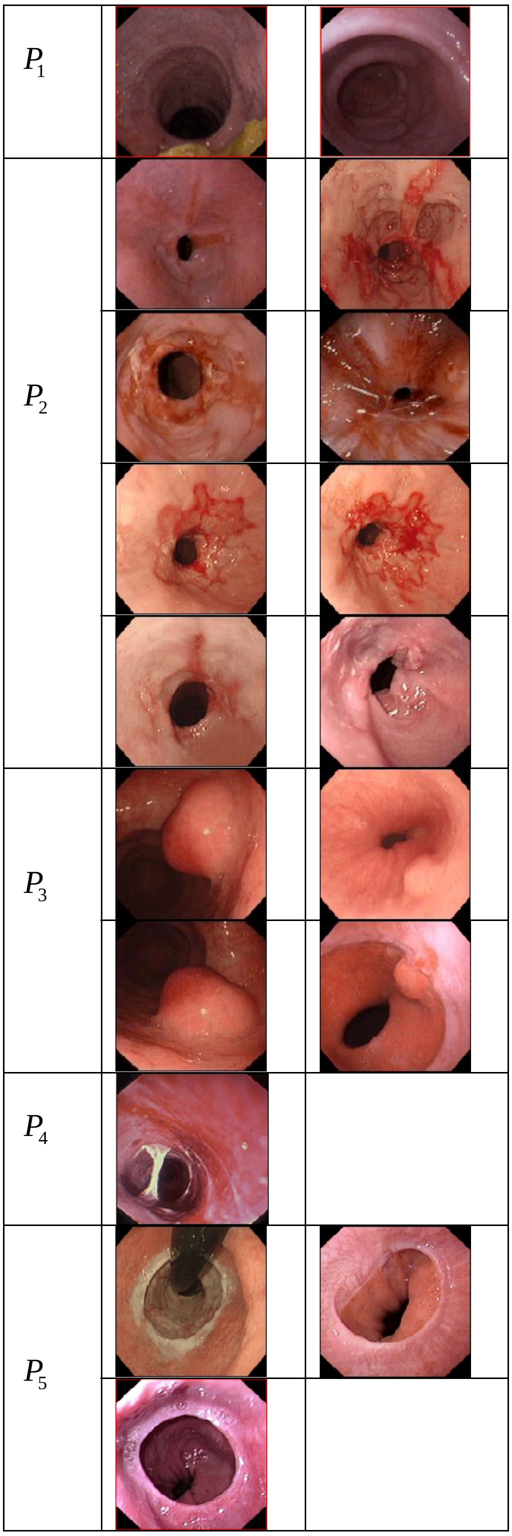

Fig.9 The simplified tested database.

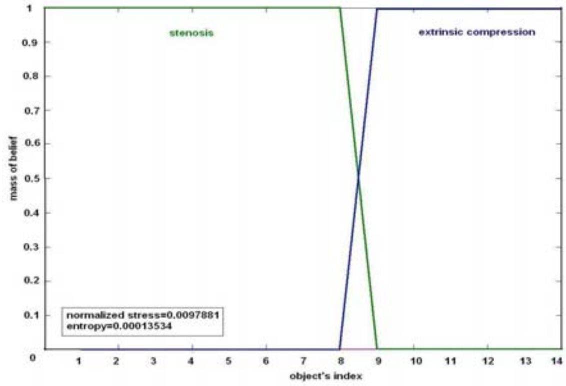

Fig.10 The partition obtained by applying our approach to $S_1$.

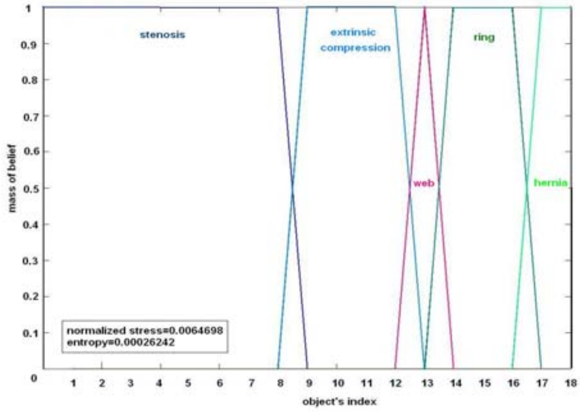

Fig.11 The partition obtained by applying our approach to $S_2$.

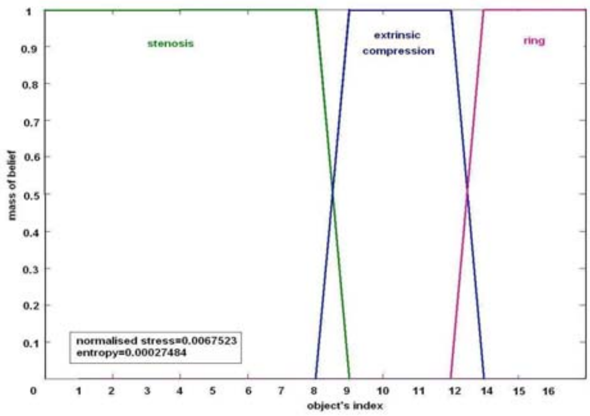

Fig.12 The partition obtained by applying our approach to $S_3$.

## 6 Comparison with Prior Works

Many attempts and methods that aim to overcome the limits and the drawbacks of the traditional measures of similarity have been proposed in the literature. However, theses methods have not been general and they treated very particular cases and databases. The most recent and efficient method among them is the method proposed by Zemirline et al. [14-18], presented briefly as follows:

Supposing that $\Omega$ is the set of all the modalities of the attributes of the cases in the casebase and that the class (pathology) of each case in this base is known:
For each class and for all the cases belonging to the considered class, the normalized frequency of appearance of each element of $\Omega$ is calculated in order to construct this class membership function represented by the histogram. The membership functions of all the classes of the casebase form the knowledge base, from which we calculate the similarity as follows:
Supposing that $f_{A_i}$ is the frequency of appearance of the modality $i$ in the set of cases belonging to the class "*A*", and $e_j$ is the set of the modalities that describe the case $j$ . $\mu_A$ is the membership degree to class "*A*" calculated by equation 36:

$$\mu_A(f_{A_i}) = f_{A_i} \Big/ \max_{j \in \Omega}(f_{A_j}) \qquad (36)$$

The similarity can be calculated by equation 37 (note that the proposed similarity is asymmetric)

$$sim_A(e_i, e_j) = \sum_{k \in e_i, k \in e_j} \mu_A(f_{A_k}) \Big/ \sum_{k \in e_i} \mu_A(f_{A_k}) \qquad (37)$$

The major restriction of Zemirline's method is that it supposes that there is a sufficient number of cases that belong to each class in order to build a reliable knowledge base, whereas in reality, sometimes we have only two or three cases of some pathologies in the database, and consequently no reliable membership functions (knowledge base) could be build basing on these objects. Actually, even if we have a considerable number of some cases, nothing can guarantee that these cases represent all the possible models of the considered pathology. Furthermore, this method can not deal with the imperfection of data (imprecision, uncertainty, or the missing values) though this imperfection could change entirely the knowledge base which the authors try to construct. Moreover, this method can not deal with all the types of data that we can find in databases (like the ordinal data for example).

# 7 Discussion and Conclusion

In this paper, we proposed an approach that depends mainly on belief theory, possibility theory, and Pignistic probability theory. By means of possibility theory, the similarity estimation can be easily carried out between objects having heterogeneously-assigned (symbolic and numeric) and imperfect (missing, imperfect, and uncertain) data without any complicated preprocessing steps needed to deal with these types of data. Furthermore, the possibility-based similarity estimation is very general and can be applied to any other problem in data mining (segmentation, classification, association, seriation, etc.). The belief theory-based clustering proposed by Denorex [3] has been shown to be very simple and efficient in getting very satisfactory results whether when it was applied to well-defined similarity matrices assigned by experts in [3] or when it is applied to the possibility-based similarity matrix in this paper. Herein, the proposed approach was applied to a digestive database. In reality, thanks to its generality, it could be applied to other types of pathologies or even in other domains without any modification or limitation. Moreover, this approach could be developed to be used in content-based case retrieval or in case-based reasoning in order to take a decision or a diagnosis by the doctor. Additionally, this method could be a very useful tool in image understanding, organization and retrieval when the attributes of an object represent descriptions of the pathologies that exist in the images. Finally, this approach provides the necessary framework to get use of the achieved information stored in the electronic health records whose number is in permanent increase thanks to the cheap storage support and the fast advances in technology, in order take the appropriate decision or the correct diagnosis [21].

# 8 Perspectives

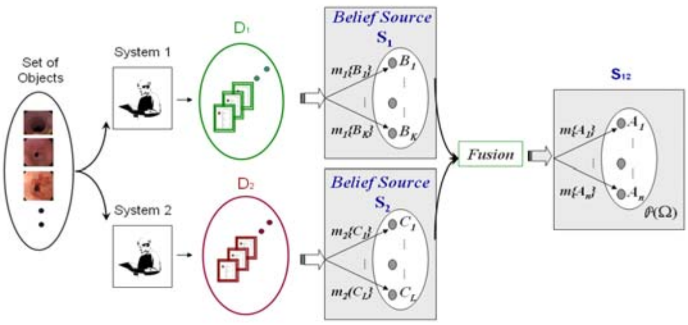


Fig.13 Data fusion

In addition to the perspectives that we mentioned briefly in the previous section, this method can be very useful in combining different similarity matrices in a meaningful manner. For example, supposing that we have a set of objects (a set of endoscopic images of the stomach for instance) and that we have two different experts (doctors) as schematized in figure 13. The first doctor describes the texture, the colors, and the homogeneity of the lesions that appear in the images, while the second describes the form, the diameters, the organization, and the dimensions of these lesions. In other words, we suppose that the same object could be represented in several manners. Thanks to our method that extracts from each representation of the objects the corresponding evidential partition, and thanks to the considerable number of the available fusion methods that enable us to easily combine several evidential partitions (like the methods of Dempester, Shafer, Yager, Dubois and Prade [3] and [22-23]) we can easily combine the two descriptions of the images. In the remote sensing domain like another example, we can combine in the same way the descriptions provided by a radar sensor that shows some aspects of the considered scene in an image and the descriptions provided by an infrared sensor that provides another type of aspects and details of the same scene, in order to take a decision concerning the class of the detected objects.

*References:*

[1] G. Bisson, *La similarité: une notion symbolique/ numérique. Apprentissage symbolique-numérique*, tome 2. Eds Moulet, Brito. Edition CEPADUES. 2000.

[2] B. Bouchon-Meunier, *La logique floue et ses applications*, Addison Wesley France, chapter 2, 1995.

[3] T. Denoeux, M. Masson, *EVCLUS: evidential clustering of proximity data*, Systems, Man, and Cybernetics, Part B, IEEE Transactions, vol. 34, pp. 95-109, 2002.

[4] H. Rakoto, J. Hermosillo, M. Ruet, *Integration of experience based decision support in industrial processes*, IEEE conference, pp. 6-12, vol.7, 2002.

[5] D. Dubois, H. Fargier, H. Prade, *Possibility theory in constraint satisfaction problems*: handling priority, preference and uncertainty. Applied Intelligence, vol. 6. pp. 287-309, 1996.

[6] C. le Guillou, J-M. Cauvin, *From endoscopic imaging and Knowledge to semantic formal images*, Springer, computer science, vol. 4370, pp. 189-201. 2007.

[7] C. Le Guillou, JM. Cauvin, B. Solaiman: *Upper Digestive Endoscopic Scene Analyze.* 23rd Conference of the IEEE Engineering in Medicine and Biology Society, vol. 4, pp. 3855-3858, 2001.

[8] JM. Cauvin, C. Le Guillou, K. Keller: *Similarity and diagnosis in digestive endoscopy.* Proceedings of the 13th World Congress Of Gastroenterology, vol. 7, issue 4, pp. 256-262, 2005.

[9] JM. Cauvin, C. Le Guillou, B. Solaiman: *Computer-assisted diagnosis system in digestive endoscopy.* IEEE Trans Inf Technol Biomed, vol. 7, issue 4, pp. 256-262, 2003.

[10] JM. Cauvin , C. Le Guillou , B. Solaiman: *Diagnostic reasoning by classification in upper digestive tract endoscopy*. World Congress on Medical Physics and Biological Engineering, vol. 1, pp. 31-34, 2000.

[11] http://i3se009d.univ-brest.fr/
Password: view2006

[12] I. Borg, P.J.F. Groenen. *Modern multidimensional scaling: theory and application second edition.* Springer, Springer series in statistics, chapters: 7, 9, and 11. 2005.

[13] P. Szolovits. *Uncertainty and decision in medical informatics*, methods of information in medicine, vol. 34, pp. 111-121, 1995.

[14] A. Zemirline, *Définition et fusion de systèmes diagnostic à l'aide d'un processus de fouille de données : Application aux systèmes diagnostics* (Thesis), Université de Rennes 1, 2008.

[15] L. Lecornu , A. Zemirline, C. Le Guillou,: *Hybrid rule and knowledge based diagnostic system fusion.*. International Journal of Biomedical Engineering and Technology (IJBET), 2007.

[16] A. Zemirline , L. Lecornu , C. Le Guillou :*Data Mining System applied in Endoscopic Image Base*, pp. 1357-1362, ICTTA 2006, 2006.

[17] A. Zemirline , L. Lecornu, B. Solaiman: *Nouvelle méthode d'extraction de règles de classification multi-labels.*. Troisième Atelier Qualité des Données et des Connaissances, Belgique, 2007.

[18] A. Zemirline , L. Lecornu, B. Solaiman. *A new hybrid fusion method for diagnostic systems.* The 11th IPMU International Conference, 2006.

[19] D. Dubois, L. Foulloy, G. Mauris, and H. Prade. *Probability-possibility transformations, triangular fuzzy sets and probabilistic inequalities*. Reliable Computing, vol. 10 : pp. 273–297, 2004.

[20] D. Dubois, H. Prade, and S. Sandri. On possibility/probability transformations. Proceedings of the Fourth Int. Fuzzy Systems Association World Congress (IFSA'91), Brussels, Belgium, pp. 50–53, 1991.

[21] J. Neves, M. Santos, J. Machado, Electronic *Health Records and Decision Support Local and Global Perspectives*, Volume 5, Issue 8, WSEAS transactions on biology and biomedecine, August 2008.

[22] I. Bloch, *Fusion d'information du signal et des images (book)*, Hermès Science, 2003.

[23] S. Rashwan, *The Possibilistic Correlation-Dependent Fusion Methods for Optical Detection*, WSEAS transactions on biology and biomedicine, Vol. 4, Issue 5, May 2007.

[24] L. Zhang, L. Xi, B. Zhou, *Image Retrieval Method Based on Entropy and Fractal Coding,* WSEAS transactions on systems, vol. 7, Issue 1, 2008.

-------------------------------