Towards a Possibilistic Classification of Gastroenterology Patterns in a Complex Environment

Anas DAHABIAH, John PUENTES, and Basel SOLAIMAN
TELECOM Bretagne, Département Image et Traitement de l'Information, Brest, FRANCE
anas.dahabiah@telecom-bretagne.eu

Abstract: - In this paper, several types of imperfection and uncertainty that may affect the information elements in any pattern recognition system will be pondered, whether when object classification depends on datasets, or when it relies on expert knowledge sets. Afterwards, all the discussed imperfection aspects will be processed within a unified framework, fundamentally based on possibility theory and fuzzy relation composition rules. At last, various concrete illustrative examples applied to gastroenterology patterns will be given and discussed.

Key-Words: - Possibility Theory, Fuzzy Relation Composition, Information Imperfection and Uncertainty, Knowledge Base.

1 Problem Description

Information imperfection is one of the most important problems remained unsolved within a unified complete framework in the field of classification and pattern recognition. This thorny issue can mainly be materialized in three essential types. The first one may be encountered at the level of the descriptors of the objects themselves, called the features, the characteristics, or the attributes, that can take any imperfect informational content due to the flood of data on the one hand, or because of the outputs of the other automatic systems which precede this step [1]. For instance, the value of a given attribute can be given as an imprecise value as the age of a patient is between 25 and 30 (quantitative imprecision), or as the pathology is either hernia grade I or hernia grade II (qualitative imprecision). The values of the other values could also be assigned via probability, evidence, or even possibility distributions [2]. It is also possible to find some missing values in the descriptors of the objects that can complicate the process. All these forms of imperfection concerning the object descriptors must be considered when designing any robust classification system.

The second type of imperfection encountered in the classification systems is the consequence of the ambiguous knowledge of the experts concerning the resemblance and the tolerance which must be carried out during the processing, i.e., sometimes, experts' opinion and viewpoints ought to be taken into account when classifying the objects, and this is called "personalization". This process enables the experts to describe to which extent he or she considers that the values of a given attribute are similar in a fuzzy manner. For example, taking the patient record as an object, some measurements and analysis could take very small values like 0.0021, 0.0022, …, 0.0028. In this case, the doctor may consider two values of such attribute are similar if their difference doesn't exceed 0.0001, while for another attribute that takes its values between 2×10^6 and 6×10^6, perhaps if the difference between two values doesn't exceed 1000, they are considered similar. Of course, if the difference is null, they are completely similar, and this similarity decreases when the difference increases until it diminishes beyond the value 1000.

These two aspects of imperfection (at the levels of object's descriptors, and expert knowledge) can affect and complicate the measuring of similarity between the objects of the dataset and the training set, which plays a vital role in any classification system.

The third type of imperfection can occur when labeling the training patterns in the learning dataset within the framework of the supervised classification [1][3]. In this case, it isn't evident to be capable to assignee each object to only one class (label, decision, or hypothesis). On the contrary, it may be associated with various labels via different strength degrees (membership degrees). This can usually take place, when the object classes are assigned by means of an automatic system [4]. This issue enables the users to take account of the real state of the objects, avoiding the complexity to find the appropriate tools that permit to only take one decision concerning the object class. The doctor is authorized for example to simultaneously belief in different sickness, pathologies, lesions, or medicines, with different trust degrees, according to the evidence that he or she has (gotten via the descriptors in the patient record, the medical images made of the concerned organ, etc.).

Unfortunately, there isn't until now any work that considers these three aspects of imperfection within a unified simple framework, though they may usually be encountered together in the real very large databases that one may handle when achieving different data mining tasks and techniques.

2 Prior Works

In spite of its importance stressed in various recent researches, the aforementioned problem has partially been addresses in the literature, i.e. the third type of imperfection has been pondered in the interesting works of classifier designing using an evidential approach [3], but with some conditions and constraints that assume that the sum of all object class membership degrees must be equal to 1 in order to be able to calculate the belief masses, even if it isn't always easy to satisfy this requirement. In addition, the first two types of imperfection haven't been considered in these works. Nonetheless, the need to deliberate them in order to compute the similarity and to achieve the different steps of any mining system [5][6], or any artificial intelligence and machine learning process [7] has been pointed out explicitly in the recent researches [8]. Accordingly, we proposed a method essentially based on possibility theory that takes the first two aspects of imperfection [2] [11] [12], then we improved this technique by taking the uncertainty in class assignment into account [1].Nevertheless, all the proposed methods are limited to work with the systems which satisfy the unity class membership sum condition.

Along with possibility theory, fuzzy relation composition rules will be used in this paper to ameliorate and to generalize the proposed classification system by taking account of all the types of imperfection in a simple, sophisticated, constraint-free design. In the following, we introduce the necessary mathematical bases for the proposed method presented in section 4, clarified by an illustrative example in section 5 and a brief conclusion accompanied with some perspectives in the last section.

3 Basic Mathematical Background

In the following, we will briefly explain two important notions in fuzzy set theory on which our approach is essentially based. The first one that is related to the fuzzy propositions [11] will be very useful to calculate the similarity between objects having imperfect information elements, by taking at the same time the ambiguous knowledge of the expert concerning the resemblance between two values of each descriptor. The second issue that presents the main rule of fuzzy relation composition [12] will be utilized to compose both the possibility-based and the necessity-based fuzzy relations related to the resemblance between the observed and the training objects, with the fuzzy relation that describes the training object class membership, in order to calculate at the end the possibility and the necessity degrees of the relation between the observed objects and each category in the class set.

3.1 Fuzzy Proposition

For a variable given via a 3-tuple information element $(V,\ \Omega,\ T_V)$, where V is the variable name defined on the universe Ω and the set $T_V = \{A_1, A_2, ...\}$ of the basic fuzzy characterization of V, "*V is A*" defined by means of a normalized fuzzy set A of Ω is called an elementary or atomic fuzzy proposition.

The compound fuzzy proposition is obtained by combining several atomic fuzzy propositions like "*V is A*" and "*W is B*", etc. The simplest compound fuzzy proposition is a conjunction of elementary fuzzy propositions "*V is A and W is B*" for two variables V and W respectively defined on the universes Ω_1 and Ω_2 (like for instance, the glycaemia level is abnormal and the cholesterol level is high"). It is associated with the Cartesian product $\Omega_1 \times \Omega_2$ of the fuzzy sets of Ω_1 and Ω_2, characterizing the pair (V, W) on $\Omega_1 \times \Omega_2$. Its truth value is defined by $\min(\mu_A(\omega), \mu_B(\omega))$ or more generally by $T(\mu_A(\omega), \mu_B(\omega))$ for a t-norm *T,* in any (ω_1, ω_2) of $\Omega_1 \times \Omega_2$. Such a fuzzy proposition is very common in rules of knowledge-based systems and in fuzzy control.

Similarly, we can combine elementary propositions by a disjunction of the form "*V* is *A* or *W* is *B*". The truth value of the fuzzy proposition is defined by $\max(\mu_A(\omega), \mu_B(\omega))$ or more generally by $\perp(\mu_A(\omega), \mu_B(\omega))$ for a t-conorm $\perp$, in any (ω_1, ω_2) of $\Omega_1 \times \Omega_2$.

We can estimate the veracity of a fuzzy proposition "*V is A*" defined via $\mu_A(\omega)$, given a referential fuzzy proposition *"V is B"* defined via $\mu_B(\omega)$ using the possibility measure $\Pi(A, B)$, and the necessity measure $N(A, B)$, defined as:

$$\Pi(A,B) = \sup_{\omega\in\Omega} \min(\mu_A(\omega), \mu_B(\omega)) \quad (1)$$

$$N(A,B) = \inf_{\omega\in\Omega} \max(\mu_A(\omega), 1-\mu_B(\omega)) \quad (2)$$

These two equations will be applied to the compound fuzzy proposition formed from the two given values of any attribute and the fuzzy referential proposition materialized by the fuzzy viewpoint of the expert described via a tolerance function.

3.2 Fuzzy Relation Composition

For two universes Ω_1 and Ω_2, the fuzzy relation R defined on the Cartesian product $\Omega_1 \times \Omega_2$ is a fuzzy set defined also on the product set $\Omega_1 \times \Omega_2$ with a membership function $\mu_R(x,y)$ that reflects the strength of the relation between $x \in \Omega_1$ and $y \in \Omega_2$:

$$\mu_R : \Omega_1 \times \Omega_2 \to [0,1] \quad (3)$$

$$R = \{((x,y), \mu_R(x,y)) / \ \mu_R(x,y) \geq 0, \ x \in \Omega_1, \ y \in \Omega_2\}$$

As the crisp relations, fuzzy relations could be represented via bipartite graph, coordinate diagrams, diagraph, and matrices associated with strength weights.

For two fuzzy matrices $A = [\alpha_{ij}]$ and $B = [\beta_{ij}]$, we can perform operations like:

- Sum (+)
  $$A + B = Max[\alpha_{ij}, \beta_{ij}] \quad (4)$$
- Max product ($\circ$)
  $$A \circ B = AB = \underset{k}{Max} \ [Min(\alpha_{ij}, \beta_{ij})] \quad (5)$$
- Scalar product λA where $0 \leq \lambda \leq 1$.

For two fuzzy relations R_1 and R_2 defined on the sets Ω_1, Ω_2, and Ω_3 where $R_1 \subseteq \Omega_1 \times \Omega_2$, and $R_1 \subseteq \Omega_2 \times \Omega_3$. The composition of these relations (Fig. 1) denoted as $R_1 \circ R_2 = R_1 R_2$ is expressed by the relation from Ω_1 to Ω_3, and can be defined by the following:

For $(x,y) \in \Omega_1 \times \Omega_2$, $(y,z) \in \Omega_2 \times \Omega_3$:

$$\mu_{R_1 \circ R_2}(x,z) = \underset{y\in\Omega_2}{Max} \ \lfloor Min(\mu_{R_1}(x,y), \mu_{R_2}(y,z)) \rfloor \quad (6)$$

$R_1 \circ R_2$ that forms this elaboration is a subset of $\Omega_1 \times \Omega_3$, i.e. $R_1 \circ R_2 \subseteq \Omega_1 \times \Omega_3$.

If the fuzzy relations R_1 and R_2 are represented by fuzzy matrices M_{R_1} and M_{R_2}, then the fuzzy matrix $M_{R_1 \circ R_2}$ that corresponds to $R_1 \circ R_2$ is obtained from the max product of M_{R_1} and M_{R_2} as:

$$M_{R_1 \circ R_2} = M_{R_1} \circ M_{R_2} \quad (7)$$

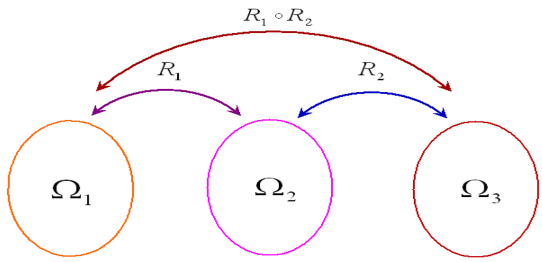


Fig. 1 Fuzzy relation composition

4 Problem Formulation and Solution

Let us suppose that we want to classify a set of observed patterns $\Theta = \{A_j^1, A_j^2, ..., A_j^i, ..., A_j^n\}$ depending on a training pattern set $T = \{A_k^1, A_k^2, ..., A_k^l, ..., A_k^t\}$ into the class set $\Omega = \{\omega_1, \omega_2, ..., \omega_c, ..., \omega_N\}$ (Fig. 2), given that A represents a multivariate feature vector of the considered pattern, whose variables could have imperfect (imprecise, ambiguous, uncertain, or even missing) values, and could be heterogeneous (quantitative, qualitative, ordinal, etc.). Each attribute may be associated with a function called the tolerance function that describes to which extent the expert think that two values of a given attribute are similar [1]. We will also suppose that the class membership of the training pattern is assessed by an expert or by another automatic system as a possibility distribution [1][3]. Specifically, each example A_k^l for $l = 1,2,...,t$ is associated with a possibilistic label modeled as: $[\mu_l(\omega_1) \ \ \mu_l(\omega_2) \ \ ... \ \ \mu_l(\omega_c) \ \ ... \ \ \mu_l(\omega_N)]$.

In the first phase, the similarity between the observed patterns and all the patterns of the training set is calculated via the necessity degree that the two patterns are similar (lower bound) and the possibility degree of resemblance (the upper degree) as we proposed in [1]. These degrees are the average of inter-attribute possibility and necessity degrees that take into account, if required, specific physicians' points of view, to adjust the similarity of each attribute between the patterns modeled by the tolerance function [2].

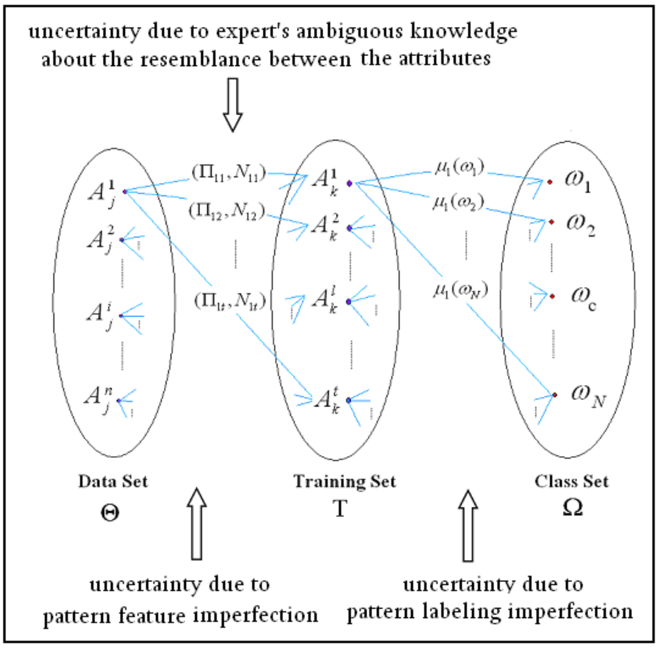


Fig. 2 Information uncertainty in a general pattern recognition system

The inter-attribute possibility and necessity degrees of similarity between the attribute value a_{fj} given in A_j^i, $\forall i \in \{1,2,...,n\}$ modeled by its possibility distribution (represented on the axis y) $\pi_{A_j, a_{fj}}(a_{fj}, y)$ and the value a_{fk} given in A_k^l, $\forall l \in \{1,2,...,t\}$ provided by its possibility distribution (represented on the axis x) $\pi_{A_k, a_{fk}}(x, a_{fk})$, $\forall f \in \{1,2,...,S\}$ (S is the number of the attributes in each pattern), are calculated as follows:

Supposing that D is the definition domain of the considered attribute ($U = D \times D$) and that Λ is the tolerance function assigned to this attribute, the conjoint possibility distribution π_D is calculated by:

$$\pi_D(a_{fj}, a_{fk}) = \min(\pi_{A_j, a_{fj}}(x), \pi_{A_k, a_{fk}}(y)) \tag{8}$$

In this case, the inter-attribute possibility and necessity degrees of similarity Π_f and N_f can be calculated as:

$$\Pi_f(a_{fj}, a_{fk}) = \sup_{u \in U}[\min(\Lambda(u), \pi_D(u))] \tag{9}$$

$$N_f(a_{fj}, a_{fk}) = \inf_{u \in U}[\max(\Lambda(u), 1 - \pi_D(u))] \tag{10}$$

For the missing values, we consider that if the value of an attribute is given in a pattern and is unassigned in the other (the case of missing values), it is completely possible that these values are similar $\Pi_f = 1$ but we are entirely uncertain $N_f = 0$.

Notice that measuring the similarity between the values of the attributes assigned via probability distributions is very simple and straightforward, because these distributions can easily be transformed into possibility distributions using any transformation in the literature that satisfies Zadeh consistency principle like the well-known transformation of Prade-Dubois [13].

On the other hand, if the values of the attributes are assigned as imprecise or ambiguous values described via membership functions, we can consider that the values of these functions are numerically equal to the possibility degrees at the level of the singletons of the frame of discernment according to the epistemic distribution principle (fuzzy restriction), and consequently, the corresponding possibility distributions can easily be deduced to measure the similarity. Actually, measuring the possibilistic similarity modeled by the possibility and the necessity degrees of resemblance is always simple, straightforward, and can adopt any type or nature of the attributes' values.

In this scene, three obvious fuzzy relations can be found. The first one R_1 represents the relation between the observed patterns in the dataset and the training objects in the learning set via the necessity degree of similarity, and its fuzzy matrix will be denoted as $M_{R_1} = [\Pi_{il}]$. the second relation R_2 reflects this resemblance via the possibility degree of similarity and can be described by a fuzzy matrix $M_{R_2} = [N_{il}]$. The last relation R_3 is already defined via $M_{R_3} = [\mu_l(\omega_c)]$ that contains the membership degrees of the training object classes.

Using the fuzzy relation composition rules, we propose to compose R_1 with R_3 (where $M_{R_1 \circ R_3} = M_{R_1} \circ M_{R_3}$) to find the membership degree of each observed object to each class in the discernment frame based on necessity degree, and then we compose R_2 with R_3 (where $M_{R_2 \circ R_3} = M_{R_2} \circ M_{R_3}$) to find the membership degree of each observed object to each class in the discernment frame based on possibility degree.

In $M_{R_1 \circ R_3}$ and $M_{R_2 \circ R_3}$ respectively, one can find the lower and the upper bound of the membership degree of any observed object to the classes in the information content set.

5 Illustrative Example

Let us suppose that our system consists of three elements $\{A_j, A_k^1, A_k^2\}$ where A_j is the observed object, while A_k^1 and A_k^2 are the training patterns. Each object is described by three attributes $\{a_1, a_2, a_3\}$ as follows: a_1 is a nominal binary attribute that takes its values in the set {positive, negative}, a_2 is an imprecise quantitative measure, and a_3 is characterized based on other measures or on other automatic systems via a probability distribution. The values assigned to these three attributes in the three aforementioned patterns are depicted in Fig. 3-a.

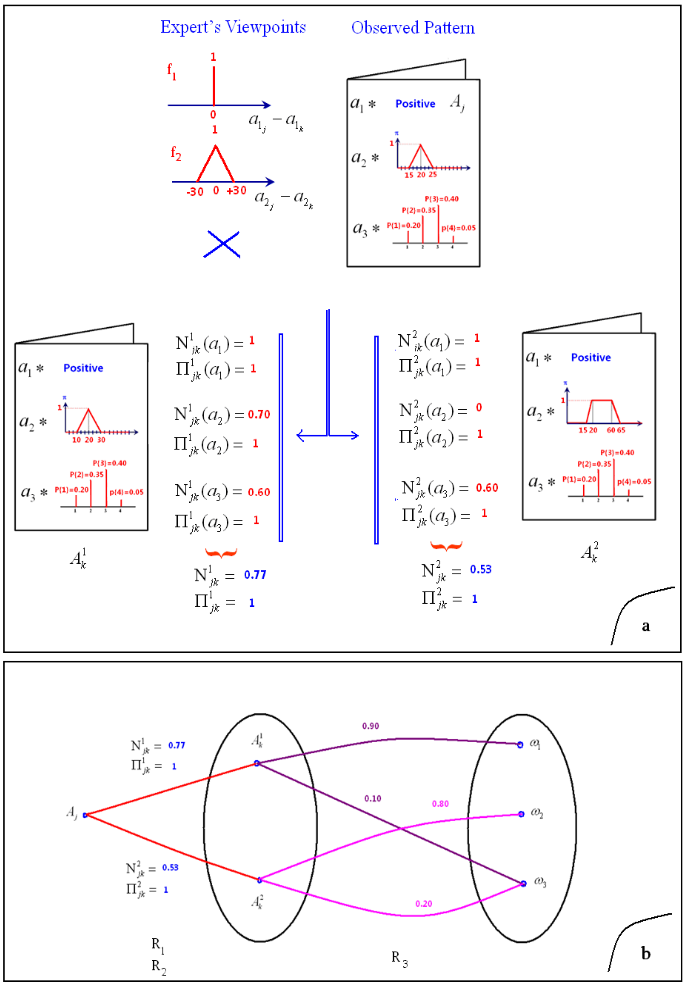


Fig. 3 Illustrative example: a- the observed object and the training patterns, with their assigned values, inter-variable possibility and necessity degrees, and the global degrees. b- the pattern recognition system and the fuzzy relations

In order to estimate the similarity between these three objects – which is a fundamental step in any recognition system - we assume that experts' viewpoints about the resemblance between each two variables will be modeled by tolerance functions. As schematized in Fig. 3-a, the true/false tolerance function is designed for the first attribute. In other words, in the expert's opinion, only if the two values of this attribute are identical, then they are considered completely similar, elsewhere they are totally dissimilar. Concerning the second attribute, the similarity is modeled by f_2 as illustrated by Fig. 3-a. According to f_2, two values of an attribute are totally similar if the difference between them is null. This similarity decreases when the difference increases until ± 30. Beyond this value, the similarity is equal to 0. Regarding the last attribute, we suppose that the expert didn't give an opinion to formulate its similarity.

As one might see in this simplified example, measuring the similarity between these pattern that contain heterogeneous values (qualitative, quantitative, etc.), and imperfect information elements (imprecise, probabilistic, etc), is extremely complex and uncertain using the conventional and the prior measures and approaches, particularly when experts' points of view must be taken into account, in addition to information imperfection. On the contrary, this calculation is completely straightforward and efficient when applying the proposed unified framework.

The values of inter-attribute similarities modeled by the similarity possibility degrees and the similarity necessity degrees are shown in Fig. 3-a. These values have been averaged in order to calculate the inter-pattern similarity between A_j and A_k^l, $(\forall l \in \{1,2\})$, modeled by Π^l_{jk} and N^l_{jk}.

According to the calculated values, we obtain two relations R_1 and R_2 described via the fuzzy matrices:

$$M_{R_1} = \begin{array}{cc} & \begin{array}{cc} A_k^1 & A_k^2 \end{array} \\ A_j & [0.77 \quad 0.53] \end{array} \quad and \quad M_{R_2} = \begin{array}{cc} & \begin{array}{cc} A_k^1 & A_k^2 \end{array} \\ A_j & [1.0 \quad 1.0] \end{array}$$

These relations must be composed with the relation R_3 given via the matrix:

$$M_{R_3} = \begin{array}{cccc} & \omega_1 & \omega_2 & \omega_3 \\ A_k^1 & 0.90 & 0 & 0.10 \\ A_k^2 & 0 & 0.80 & 0.20 \end{array}$$

Applying the fuzzy relation composition rule (eq. 7) gives:

$$M_{R_1 \circ R_3} = A_j \quad \begin{array}{ccc} \omega_1 & \omega_2 & \omega_3 \\ {[0.77} & 0.53 & 0.20] \end{array}$$

and

$$M_{R_2 \circ R_3} = A_j \quad \begin{array}{ccc} \omega_1 & \omega_2 & \omega_3 \\ {[0.90} & 0.80 & 0.20] \end{array}$$

In other terms: $A_j(\omega_1) = [0.77,\ 0.90]$, $A_j(\omega_2) = [0.53,\ 0.80]$, $A_j(\omega_3) = 0.20$.

As we notice, this result describes the membership strength between the observed patterns and the categories of the class set, providing a logical and expected response, by using very simple steps.

In spite of its simplicity, this example can be applied to many other problems and applications more complicated, following the same techniques of computation, regardless of the number of the observed or the training patterns, and independently of the number of their descriptors.

7 Interesting Applications in Knowledge Databases

In the following we present one of the most important applications of the proposed method in knowledge databases, in which we intend to classify a given object depending on the imperfect subjective knowledge of the expert. In order to simplify the illustration, our explanation will always be joined with medical examples, and to guarantee a better understanding of the proposed strategy [14] [15], concrete numeric examples will be presented at the end of this paragraph.

Given a knowledge database in which subjective descriptions of the experts of the domain about the classes of certain cases are provided in such a way that experts' uncertainty are modeled in a qualitative manner, and given the descriptions of a certain object, our goal is to take a decision concerning the class of this object by fusing its own imperfect descriptions and those provided in the knowledge database. In a gastroenterology knowledge databases, classes could be for instance a set of given pathologies, denoted as $T = \{P_1, P_2, \ldots\}$. Each of them is described by a set of S attributes $X = \{x_1, x_2, \ldots, x_S\}$. Each attribute x_i is described in a qualitative manner, and takes its values from a given set called the attribute domain denoted as Ψ_i. This set contains all the possible values of this attribute. For instance, the domain of the attribute "esophagus color" can be Ψ=*{red, pink, white, yellow, brown, gray, black, blue, green, translucent}*.

In figures 4 and 5, we give two examples of two different described classes (pathologies) [15] [16]. The first one is "esophagus hiatal Hernia", and the second one is "esophagus Mallory-Weiss". For clarity, we suppose that these classes are composed of only seven patient's attributes *{sex, age, ...}*. The first attribute's domain is Ψ_1=*{male, female},* the seconds one "age" takes its values in Ψ_2=*{<20, 20-30, 30-40, 40-50, 50-60, 60-70,70-80, >80},* etc.

For a given pathology, the expert doctor can describe his knowledge concerning each possible value of each attribute in the class via qualitative descriptions defined upon a set called the set of frequencies. In this example this set is *F = {never, exceptional, few, usual, always}*. For instance, in the examples given in figures 4 and 5, concerning the age of the patient, the first pathology is usual if the age exceeds the forty, and can happen in few cases before forty and exceptionally before twenty, whereas, the second pathology usually takes place between twenty and fifty and in few cases in the other ages. It can also be noticed that according to pathology's class and the joined descriptions, doctors can assign the class to a lesion "called endoscopic finding" from a set of predefined lesions Ω=*{ulcer, Z-line, red blood clot, red blood liquid, ...}*. Figure 6 depicts an example of an object used in this context [16] [17] [18]. The object can be a patient record, a medical image, a video, etc. In this example, the object is the doctor descriptions of the gastroenterology image of a patient, in which the pathology is ""esophagus hiatal Hernia" (the class described in figure 4), and the lesion is "Hiatus Hernia".

Our aim is to classify objects like the one in figure 6, denoted as $X_j = [x_{1j} \ \ldots \ x_{ij} \ \ldots \ x_{Sj}]^t$ to the lesions of the set $\Omega = \{\omega_1 \ \ldots \ \omega_c \ \ldots \ \omega_N\}$ depending on the descriptions of the knowledge data base of the N_P pathologies $T = \{P_1 \ \ldots \ P_l \ \ldots \ P_{N_P}\}$, where $j = 1,\ 2,\ \ldots, N_o$, and N_o is the total number of the objects. Each object has *S* attributes, and each attribute x_{ij} ($\forall j$) is defined upon its domain set $\Psi_i = \{A_i^1 \ \ A_i^2 \ \ \ldots \ \ A_i^{m(i)}\}$ where *i* is the order of the attribute and *m(i)* is the number of the properties (characteristics) of the i^{th} attribute. *N* stands for the number of the lesions in the class set. For the generality of the proposed strategy we will suppose that the pathologies are assigned to the lesion using membership degrees. Figure 7 shows the general scheme of our work, joined with the associated notations.

Disease : Esophagus Hiatal Hernia

Hernie hiatale

Sex

Male	no interest
Female	no interest

Age

< 20 years	exceptional
20-30 years	few
30-40 years	few
40-50 years	usual
50-60 years	usual
60-70 years	usual
70-80 years	usual
> 80 years	usual

Complementary Procedure

Biology	no interest
Biopsy	no interest
Methylene blue coloration	no interest
Toluidine blue coloration	no interest
Clot takeoff	no interest
Echo-endoscopy	no interest
Clinical Examination	no interest
Other functional	no interest
Manometry / pH-determination	no interest
Other morphology	no interest

Clinical Context

Previous history

Alcohol addiction	no interest
Immuno-depression	no interest
Congenital abnormality	no interest
Non neoplastic duodenal diseases	no interest
Non neoplastic gastric diseases	no interest
Liver diseases	no interest
Infectious diseases	no interest
Neurologic diseases	no interest
Non neoplastic esophageal diseases	no interest
Pancreatic diseases	no interest
Psychiatric diseases	no interest
Splenic diseases	no interest
Non malignant neoplastic diseases	no interest
Malignant neoplasms: other organs	no interest
Malignant neoplasms: eso-sto-duo	no interest
Tobacco addiction	no interest
Previous Surgery: other organ	no interest
Previous Surgery: biliopancreatic	no interest
Previous Surgery: duodenum	no interest
Previous Surgery: stomach	no interest
Previous Surgery: esophagus	no interest
Previous Varices Sclerotherapy	no interest
Previous radiotherapy	no interest

Clinical situations

Caustic injury	no interest
Foreign body	no interest
Screening	no interest
Gastro-toxic drugs	no interest
Not fasting	no interest
Traumatology	no interest

Signs and symptoms

Body Weight Loss	no interest
Anemia	no interest
Unconsciousness	no interest
Diarrhea	no interest
Pain	no interest
Dyspepsia	no interest
Dysphagia	no interest
Digestive bleeding	no interest
Malabsorption syndromes and enteropathies	no interest
Biliopancreatic signs	no interest
Nausea / vomiting	no interest
Regurgitation	usual
Gastroesophageal reflux symptoms	usual

Endoscopic Findings

1.3.1.1.6 Hiatus Hernia	always

Fig. 4 An example of the class description in a knowledge qualitative database

Disease : Esophagus Mallory-Weiss

Syndrome Mallory-Weiss

Sex

Male	no interest
Female	no interest

Age

< 20 years	few
20-30 years	usual
30-40 years	usual
40-50 years	usual
50-60 years	few
60-70 years	few
70-80 years	few
> 80 years	few

Complementary Procedure

Biology	no interest
Biopsy	no interest
Methylene blue coloration	no interest
Toluidine blue coloration	no interest
Clot takeoff	no interest
Echo-endoscopy	no interest
Clinical Examination	no interest
Other functional	no interest
Manometry / pH-determination	no interest
Other morphology	no interest

Clinical Context

Previous history

Alcohol addiction	no interest
Immuno-depression	no interest
Congenital abnormality	no interest
Non neoplastic duodenal diseases	no interest
Non neoplastic gastric diseases	no interest
Liver diseases	no interest
Infectious diseases	no interest
Neurologic diseases	no interest
Non neoplastic esophageal diseases	no interest
Pancreatic diseases	no interest
Psychiatric diseases	no interest
Splenic diseases	no interest
Non malignant neoplastic diseases	no interest
Malignant neoplasms: other organs	no interest
Malignant neoplasms: eso-sto-duo	no interest
Tobacco addiction	no interest
Previous Surgery: other organ	no interest
Previous Surgery: biliopancreatic	no interest
Previous Surgery: duodenum	no interest
Previous Surgery: stomach	no interest
Previous Surgery: esophagus	no interest
Previous Varices Sclerotherapy	no interest
Previous radiotherapy	no interest

Clinical situations

Caustic injury	no interest
Foreign body	no interest
Screening	no interest
Gastro-toxic drugs	no interest
Not fasting	no interest
Traumatology	no interest

Signs and symptoms

Body Weight Loss	no interest
Anemia	no interest
Unconsciousness	no interest
Diarrhea	no interest
Pain	no interest
Dyspepsia	no interest
Dysphagia	no interest
Digestive bleeding	always
Malabsorption syndromes and enteropathies	no interest
Biliopancreatic signs	no interest
Nausea / vomiting	usual
Regurgitation	no interest
Gastroesophageal reflux symptoms	no interest

Endoscopic Findings

1.3.1.2.4 Blood (red)	few
1.3.1.2.4.a Blood (clot)	few
1.3.1.3.0.b Z line	usual
1.3.1.6.1 Ulcer	always
1.3.2.2.0.a Blood (red)	few
1.3.2.2.0.b Blood (clot)	few
1.3.2.6.1.a Ulcer (edge)	always

Fig. 5 Another example of the class description in a knowledge database

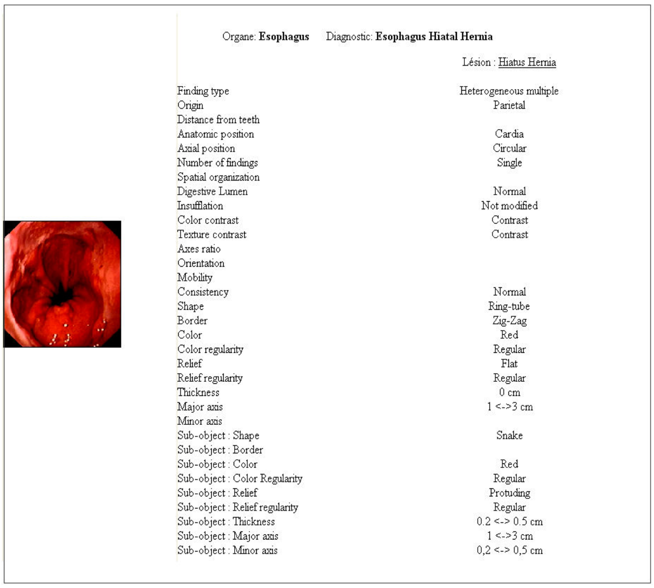


Fig. 6 An example of the descriptions of the objects in a gastroenterology database

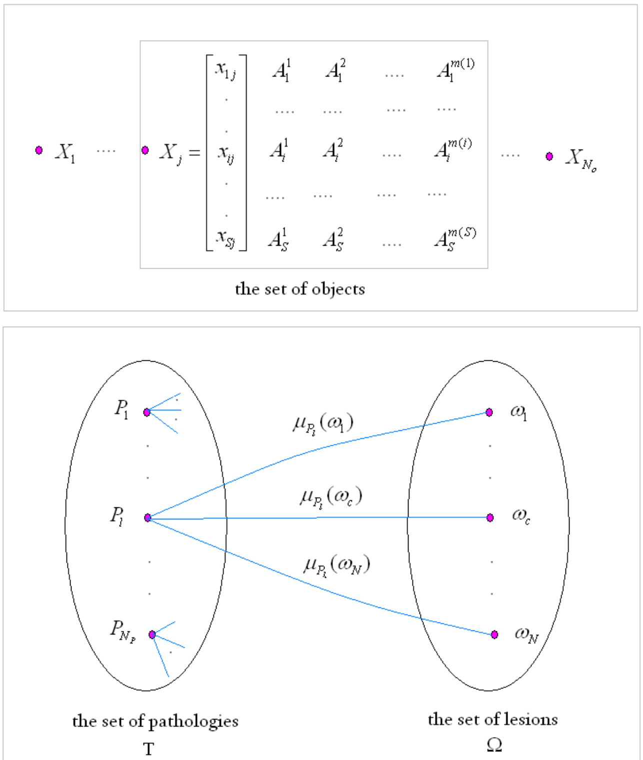


Fig. 7 The sets of patterns, pathologies, and lesions. Each pattern must be classified to the lesions according to its descriptions, the pathology descriptions, and the pathology membership degree to the considered lesions

For each pathology P_l in the knowledge dataset, where $l=1,..,N_p$, the expert qualitatively describes the frequencies of all the possible values of each attribute. i.e. for each element of the set $\Psi_i=\{A_i^1 \quad A_i^2 \quad \dots \quad A_i^{m(i)}\}$ that represents the attribute's domain of the feature x_{ij} in the object X_j, where *i=1,..,S*, the doctor assigns an ordinal qualitative value from the set of frequencies *F*, like the values "never", "exceptional", "few", "usual", always", etc., to reflect his opinion and uncertainty regarding the frequencies of each possible value in the concerned pathology, $\Psi'_i=\{f_{P_l}(A_i^1) \quad f_{P_l}(A_i^2) \quad \dots \quad f_{P_l}(A_i^{m(i)})\}$, where each element belongs to the set *F*. As the elements of the set *F* are ordinal, then we can associate to each of them the possibility degree $\pi_r=\dfrac{r-1}{R-1}$, where *r* is the rank of the considered property and *R* is the total numbers of the elements of the set of frequencies (See the example in figure 8).

F	never	exceptional	……	always
r (rank)	1	2	……	R
$\pi_r=\dfrac{r-1}{R-1}$	0	$\dfrac{1}{R-1}$	……	1

F	never	exceptional	few	usual	always
r (rank)	1	2	3	4	5
$\pi_r=\dfrac{r-1}{R-1}$	0	1/4	1/2	3/4	1

Fig. 8 A general and a numeric example of calculating the possibility degree

Accordingly, for each $f_{P_l}(A_i^g)$, for $g=1,2,\dots,m(i)$, that represents the expert description of all the g^{th} property of the i^{th} attribute in the pathology P_l, the qualitative value can be replaced by $\pi_r(f_{P_l}(A_i^g))$ (figure 9). Then this value can be replaced by the possibility and the necessity measures as follows: for every property, the possibility measure of this characteristic in the considered pathology $\Pi_{P_l}(A_i^g)$ is equal to the possibility degree $\pi_r(f_{P_l}(A_i^g))$, whereas the necessity measure of its occurrence in this pathology can be calculated from the equation:

$$N_{P_l}(A_i^g)=1-\max_{all\ \ G\in\{1,2,\dots,m(i)\}/g}\pi_r(f_{P_l}(A_i^G))$$

In other words, the necessity of the occurrence of a given possible value of an attribute is equal one minus the maximum value of the possibility degrees of the other possible values of this attributes.

Figure 9 depicts the main forms of the knowledge dataset in details.

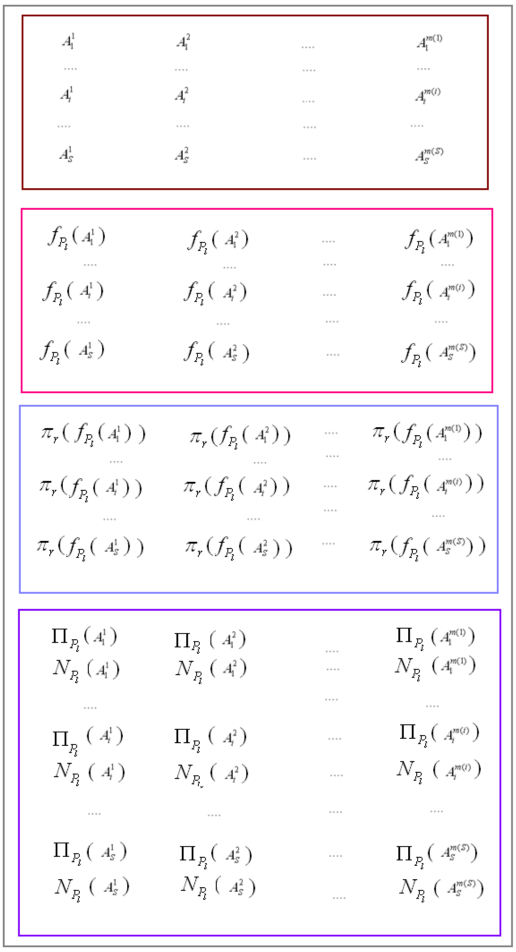


Fig. 9 The different steps of calculating the possibility and necessity measures of all the possible values of all the attributes in a given pathology

To classify the objects of the database to the corresponding lesions, depending on the descriptions of the knowledge base, two fuzzy relations must be combined using the fuzzy composition rule [14] [15]. The first one is composed of the possibility and the necessity measures calculated between each object in the dataset and each class in the knowledge set. These measures reflects the possibility and the necessity that a given object can have a given class (pathology) and can be computed as the average of the possibility and the necessity measures of the given values in all descriptions of the objects, numerically estimated from the knowledge set. The second fuzzy relation can be assigned by the expert in the knowledge set to describe the relation between the pathologies and the lesions (figure 10).

This can be better understood by using a concrete simple example. Assume that we would like to classify an object that contains two attributes: the color assigned to "white", and the shape assigned to "triangle" (figure 10-a), given that the first attribute's domain is Ψ_{color}=*{white, gray, black}*, and the second attribute's domain is Ψ_{shape}=*{Rectangle, triangle, circular}*, and the class universe is $\Omega = \{L_1,\ L_2,\ L_3\}$ (*L* stands for lesion, for example).

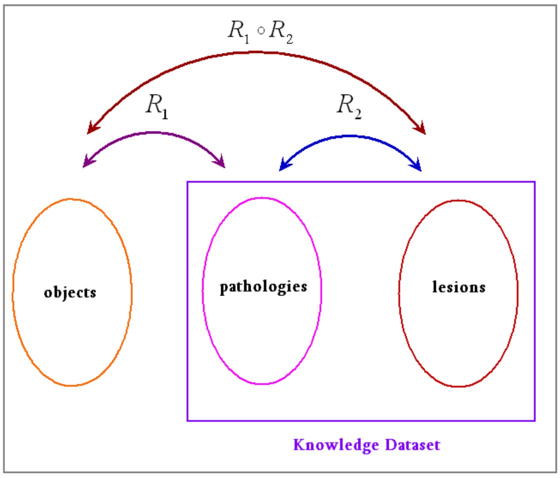


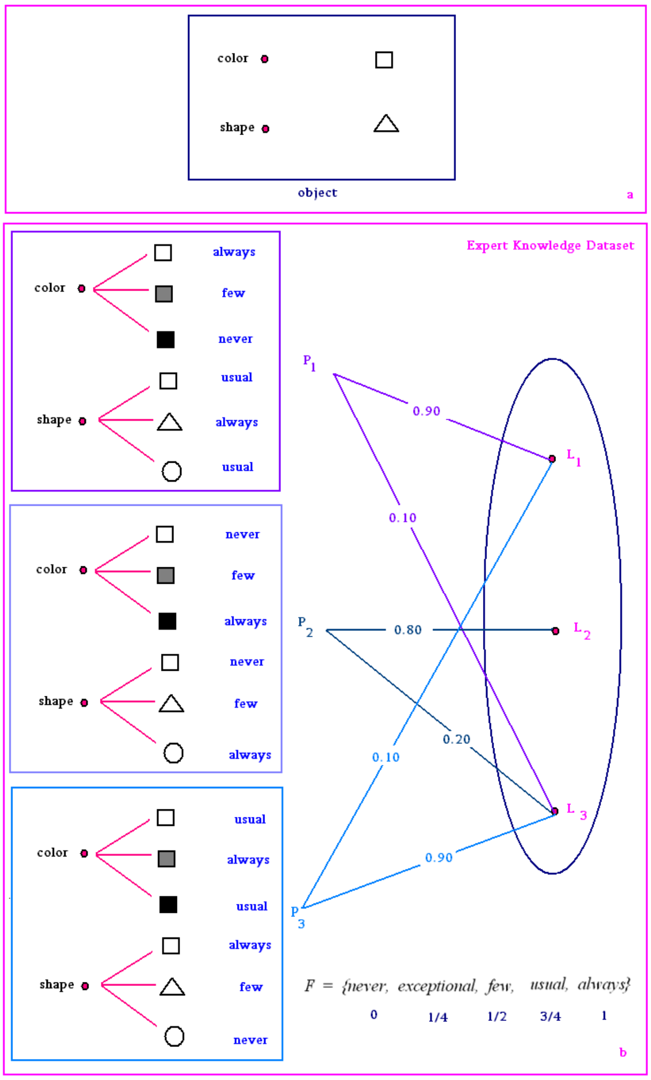
Fig. 9 Illustrative scheme of the last step of combining the two fuzzy relations

The classification will be carried out depending on the knowledge database described in figure 10-b, supposing that we have three classes of pathologies $T = \{P_1,\ P_2,\ P_3\}$, and each pathology is described via qualitative values defined upon *F={never, exceptional, few, usual, always}* to describe the frequency of each possible value of each attribute. Accordingly, each pathology will be assigned to the lesions in the class set Ω via membership degrees (possibility distribution) as shown in figure 10-b. For example, regarding the pathology P_1 in which the color white is the dominant value of the color, and the shape "triangle" is the most frequent followed by the "circular" and "rectangular" which can be usually found, there is a very great possibility 90% that the lesion is L_1, and a very little possibility 10% to be L_3, but it is impossible that this pathology has L_2 as a lesion.

To classify the given object (figure 10-a) by taking account of the given descriptions of the knowledge base (figure 10-b), the describing values of the modalities of each attribute that belongs to the set *F* must be transformed to possibility and necessity measures defined upon the set [0, 1] as explained above. Consequently we got the values depicted in figure 10-c.

Now that the possibility and necessity measures have been computed, the first fuzzy set between the object and the pathologies can be estimated as follows: as the value of the first attribute in the object is "white", then its possibility to belong to the pathology P_1 is 1, to

P_2 is 0, and to P_3 is 3/4. While the necessity to P_1 is 1/2, to P_2 and to P_3 is 0. As the value of the second attribute in the object is “triangle”, then according to the knowledge base its possibility to belong to the pathology P_1 is 1, to P_2 and P_3 is 1/2. While the necessity to P_1 is 1/4, to P_2 and P_3 is 0. To calculate the possibility and the necessity measure that the object belongs to a given pathology, we compute the average of the possibility and that of the necessity measures. Thus, the object belong to P_1 with a possibility degree equals to (1+1)/2=1 and necessity degree equals to (1/2+1/4)/2=0.375, to P_2 the possibility is equal to 0.25 and the necessity is null, to P_3 the possibility is equal to 0.625 while the necessity is null. The measure (shown in figure 10-d) represents the first fuzzy relation. The second one is represented by the membership degrees assigned by the experts to describe the relations between the pathologies and the lesions (figure 10-d).



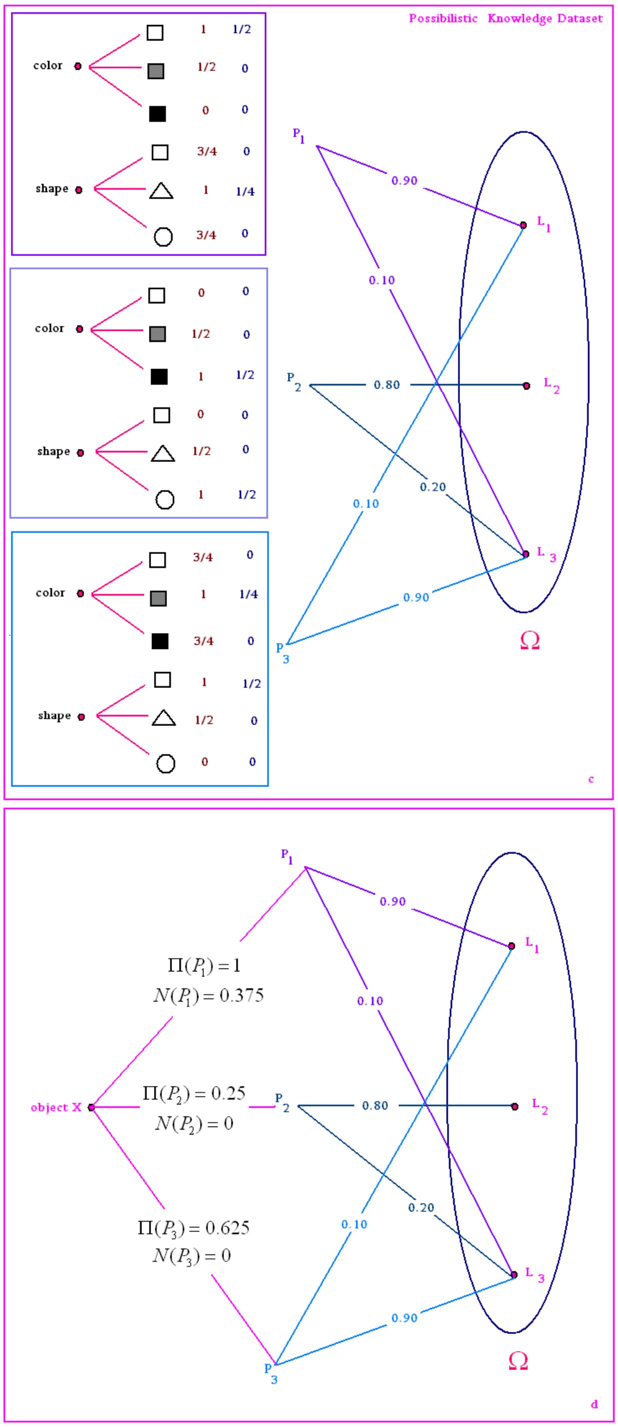


Fig. 10 An illustrative numeric example, a) the described object, b) the expert knowledge database, c) the possibilistic knowledge database, d) the two fuzzy relations that must be combined: the first one represents the possibility and the necessity degrees that the object belong to the pathologies, and the second one represents the relations between the pathologies and the lesions

Using the fuzzy relation composition rule to combine these two relations, we get:

$$\begin{array}{cc} & \begin{array}{ccc} P_1 & P_2 & P_3 \end{array} \\ object & [1 \quad 0.25 \quad 0.625] \\ & \begin{array}{ccc} P_1 & P_2 & P_3 \end{array} \\ object & [0.375 \quad 0 \quad 0\ ] \end{array} \circ \begin{array}{c|ccc} & L_1 & L_2 & L_3 \\ \hline P_1 & 0.90 & 0 & 0.10 \\ P_2 & 0 & 0.80 & 0.20 \\ P_3 & 0.10 & 0 & 0.90 \end{array} =$$

$$\begin{array}{l} \\ object\ \ possibility \\ object\ \ necessity \end{array} \begin{array}{c} \begin{array}{ccc} L_1 & L_2 & L_3 \end{array} \\ \begin{bmatrix} 0.90 & 0.25 & 0.625 \\ 0.375 & 0 & 0.10 \end{bmatrix} \end{array}$$

As expected, we can notice that the described object belongs to the first lesion with high possibility and necessity degree since it is highly similar to the first pathology that belongs to this category.

In figure 11, we give another example of a knowledge database that consists of three classes; each of them is described by three different attributes. The first one takes four possible values, the second takes 3 values, and the last one takes 6 values. In this example, we calculate the possibility and the necessity degrees of each possible value of each attribute in each pathology, according to the method proposed in this paper.

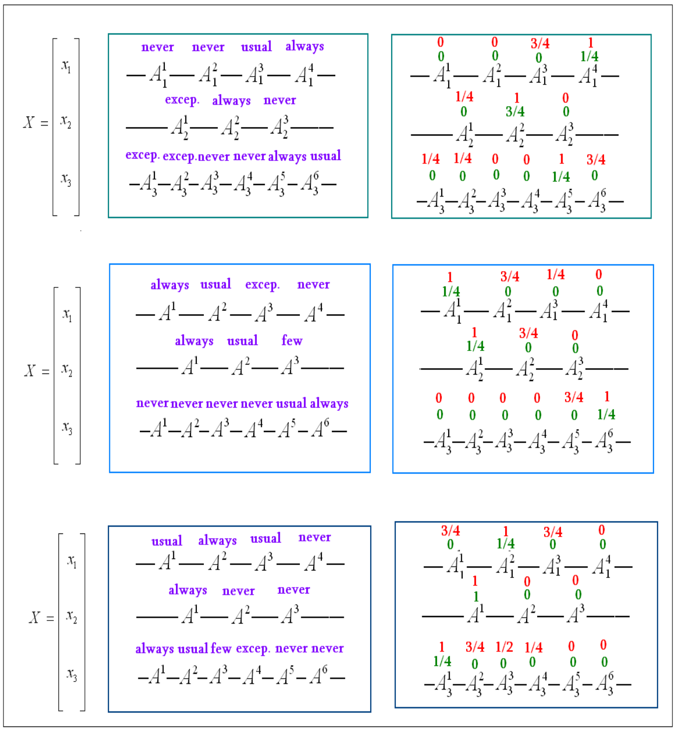


Fig. 11 An example of transforming the expert knowledge dataset to a possibilistic one

The clear notice that one can get from these two examples besides the example presented in section 5 is that the computation and the estimation of the classes is extremely simple and straightforward in spite of all the types of imperfection that can be encountered in the classification system.

Actually, in some bases, the expert knowledge can be modeled via probability distributions, especially when it comes from automatic systems. In this case, probability distribution can easily be transformed to possibility distributions using an appropriate transformation, and then the proposed method can directly applied to these possibility degrees.

6 Discussion and Perspectives

In this paper, three main types of information element imperfection have been treated within a simple unified framework, essentially based on the mathematical tools provided by the fuzzy set theory, namely the possibility theory, and the fuzzy relation composition rules. All of these tools fundamentally get use of the main basic computation operations (like the sum, the maximum, the minimum, etc.) of the processors or the logic circuits, and can easily be evaluated using matrix structures. This issue can assure a fast processing and a rapid computation of the proposed approach, which is an important factor in pattern recognition and data mining, where the volume of the databases that we handle can be notably large, and where the number of the descriptors which describe the patterns of these bases may be significantly considerable.

Herein, the approach has been applied to overcome the complexity and the difficulties in classification resulting from the imperfection in the available information elements. But in general, it can be applied to many other data mining tasks and applications using the same technique and reasoning. For example, the possibilistic similarity proposed in the first phase of our method can be apply in clustering, retrieving, case-based reasoning, seriation, etc., without any modifications or other pre-processing steps concerning the imperfect information elements.

The various examples presented in this paper prove the simplicity and the generality of the proposed approach when classifying complex objects based whether on datasets, or on knowledge sets. Contrary to our approach, all the proposed methods in the literature stand incapable to easily process the complex and hostile conditions proposed within a unified, constraint-free, straightforward framework. Furthermore, these examples can be applied to other types of knowledge or data sets, in the military, biological, etc. domains.

References:

[1] A. Dahabiah, J. Puentes, B. Solaiman, Imperfect Pattern Recognition Using the Fuzzy Measure Theory, *Springer LNCS*, Vol. 5788, 2009, pp. 101-108

[2] A. Dahabiah, J. Puentes, B. Solaiman, Possibilistic Similarity Estimation and Visualization, *Springer LNCS*, Vol. 5766, 2009, pp. 273-280

[3] T. Denoeux, Neural Network Classifier Based on Dempster-Shafer Theory, IEEE *Transactions on Systems, Man, and Cybernetics*, Vol. 30, No. 2, 2000, pp. 131-150

[4] A. Dahabiah, J. Puentes, B. Solaiman, Digestive database evidential clustering based on possibility theory. *WSEAS transactions on biology and biomedicine*, vol. 5, no. 9, pp. 239-248.

[5] B. Bouchon-Meunier, M. Rifqi, M-J Lesot, Similarities in Fuzzy Data Mining: From a Cognitive View to Real-World Applications, *Springer LNCS*, Vol. 5050, 2008, pp. 349-367

[6] E. Pecalska, R. Duin, The Dissimilarity Representation for Pattern Recognition, *World Scientific Publishing*, 2005, Chapters 1,5,6, and 9

[7] E. L. Rissland, AI and Similarity, *IEEE Intelligent Systems*, Vol. 21, No. 3, 2006, pp. 39-49

[8] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, C. Roux, Multimodal Information Retrieval Based on DSmT : Application to Computer-Aided medical Diagnosis, *Advances and Applications of DSmT for Information Fusion,* Vol. 3, edited by F. Smarandache, J. Dezert Chapter 18, 2008, pp. 471-502

[9] A. Dahabiah, J. Puentes, B. Solaiman, Venous Thrombosis Supervised Image Indexing and Fuzzy Retrieval, *Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007, pp. 4528-4531

[10] Dahabiah, A., Puentes, J., Solaiman, B., Possibilistic Pattern Recognition in a Digestive Database for Mining Imperfect Data. *WSEAS Transactions on Systems*, vol. 8, no. 2, pp. 229-240.

[11] B. Bouchon-Meunier, La logique floue et ses applications, *Addison-Wesley France*, Collection: Vie Artificielle, Chapters 1,2, and 3, 1990

[12] J. Kacprzyk, Fuzzy Relation and Composition, *Springer First Course on Fuzzy Theory and Application, Advances in Soft Computing*, Vol. 7, 2005, Chapter 3, pp. 53-90

[13] M. H. Masson, T. Denoeux, Inferring a Possibility Distribution from Empirical Data, *Fuzzy Sets and Systems*, Vol. 157, No. 3, 2006, pp. 319-340

[14] A. Zemirline, L. Lecornu, B. Solaiman, A New Hybrid Fusion Method for Diagnostic Systems, *IPMU*, Chapter 28, 2006, pp. 389-401

[15] C. Le Guillou, J.M. Cauvin, From Endoscopic Imaging and Knowledge to Semantic Formal Images, *Springer*, LNCS, Vol. 4370, 2007, pp. 189-201

[16] J.M. Cauvin, C. Le Guillou, B. Solaiman, M. Robaszkiewicz, P. Le Beux, C. Roux, Computer-assisted diagnosis system in digestive endoscopy. *IEEE Transactions on Information Technology in Biomedicine*, Vol. 7, No. 4, 2003, pp. 256-262

[17] A. Dahabiah, J. Puentes, B. Solaiman, Digestive Casebase Mining Based on Possibility Theory and Linear Unidimensional Scaling. *8th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases*, 2009, 218-223

[18] A. Dahabiah, J. Puentes, B. Solaiman, Possibilistic Evidential Clustering. *8th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases*, 2009, 212-217