II. Data Analysis:
progress and recent trends

Probabilist, possibilist and belief objects for knowledge analysis

Edwin Diday
University Paris IX Dauphine and INRIA, Domaine de Voluceau, Rocquencourt F-78153 Le Chesnay Cedex, France

The main aim of the symbolic approach in data analysis is to extend problems, methods and algorithms used on classical data to more complex data called "symbolic objects" which are well adapted to representing knowledge and which are "generic" unlike usual observations which characterize "individual things". We introduce several kinds of symbolic objects: Boolean, possibilist, probabilist and belief. We briefly present some of their qualities and properties; three theorems show how Probability, Possibility and Evidence theories may be extended on these objects. Finally, four kinds of data analysis problems including the symbolic extension are illustrated by several algorithms which induce knowledge from classical data or from a set of symbolic objects.

Keywords: Knowledge analysis, symbolic data analysis, metadata, metaknowledge, probability, possibility, evidence theory, uncertainty logic.

1.1. Introduction

If we wish to describe the fruits produced by a village, by the fact that "The weight is between 300 and 400 grammes and the color is white or red and if the color is white then the weight is lower than 350 grammes", it is not possible to put this kind of information in a classical data table where rows represent villages and columns descriptors of the fruits. This is because there will not be a single value in each cell of the table (for instance, for the weight) and also because it will not be easy to represent rules (if..., then...) in this table. It is much easier to represent this kind of information by a logical expression such as:

$$a_i = [weight = [300, 400]] \wedge [color = \{red, white\}]$$

$$\times \wedge [\text{if } [color = white] \text{then} [weight \leqslant 350]],$$

where a_i, associated to represent the ith village, is a mapping defined on the set of fruits Ω such that for a given fruit $\omega \in \Omega$, $a_i(\omega)$ = true if the weight of ω belongs to the interval [300,400], its color is red or white and if it is white then its weight is less than 350 gr. Following the terminology of this paper, a_i is a kind of symbolic object; "symbolic" because a_i is described by an expression which contains operators different from those used with classical numbers, "object" because it is considered to be an individual object for a statistic of a higher level unit; if we have a set of 1000 villages represented by a set of 1000 symbolic objects $a_1, \ldots, a_{1000}$, an important problem is to know how to apply data analysis or statistical methods to it. For instance, what is a histogram or a classification or a probability law for such a set of objects? The aim of symbolic data analysis Diday [7,8] is to provide tools for answering this problem.

In some fields a Boolean representation of the knowledge ($a_i(\omega)$ = true or false) is sufficient to get the main information, but in many cases we need to include uncertainty to represent the real world with more efficiency. For instance, if we say that in the ith village "the color of the fruits is often red and seldom white" we may represent this information by $a_i = [color = often\ red,\ seldom\ white]$. More generally, in the case of Boolean objects or objects where frequency appears, we may write $a_i = [color = q_i]$ where q_i is a characteristic function in the Boolean case, and a probability measure in the second case. More precisely, in the Boolean case, if $a_i = [color = \{red, white\}]$ we have $q_i(red) = q_i(white) = 1$ and $q_i = 0$, for the other colors; in the probabilist case, if $a_i = [color = 0.9\ red,\ 0.1\ white]$ we have $q_i(red) = 0.9$, $q_i(white) = 0.1$. If an expert says that the fruits are red we may represent this information by a symbolic object $a_i = [color = q_i]$, where q_i is a "possibilist" function in the sense of Dubois and Prade [4]; we will have, for instance, $q_i(white) = 0$, $q_i(pink) = 0.5$ and $q_i(red) = 1$. If an expert who has to study a representative sample of fruits from the ith village, says that 60% are red, 30% are white and the color is unknown for 10% which were too rotten, we may represent this information by $a_i = [color = q_i]$ where q_i is a belief function such that $q_i(red) = 0.6$ $q_i(white) = 0.3$ and $q_i(o) = 1$, where o is the set of possible colors. Depending on the kind of mapping q_i used, a_i is called a Boolean, probabilist, possibilist or belief object. In all these cases a_i is a mapping from Ω (the set of fruits) to [0,1]. Now, the problem is to know how to compute $a_i(\omega)$; if there is doubt about the color of a given fruit ω, for instance, if the expert says that "the color of ω is red or pink" then, ω may be described by a characteristic function r and represented by a symbolic object $\omega^s = [color = r]$ such that $r(red) = r(pink) = 1$ and $r = 0$ for the other colors. Depending on the kind of knowledge that the user wishes to represent, r may be a probability, possibility or belief function. Having $a_i = [color = q_i]$ and $\omega^s = [color = r]$ to compute $a_i(\omega)$ we introduce a comparison function g such that $a_i(\omega) = g(q_i, r)$ measures the fit between q_i and r. What is the meaning of $a_i(\omega)$? May we say that $a_i(\omega)$ measures a kind of probability, possibility or belief that ω belongs to the class of fruits described by a_i when q_i and r, depending on the background knowledge, are characteristic, probability, possibility or belief functions respectively? To answer this question we need to extend a_x (where x represents a kind of background knowledge) to a_x^* defined on a_x a set of symbolic objects and to define set operators $OP_x = \{\cup_x, \cap_x, c_x\}$ in a_x adapted to x. If we say that classic sets represent a knowledge level of order 0: probability, possibility and belief, a knowledge level of order 1, the question was now to know if a_x^* represents a knowledge level of order 2. In other words, if it is a probability of probability, a possibility of possibility a belief of belief respectively associated with the corresponding operators OP_x; theorems 1, 2, 3 show that this is the case, if OP_x and some functions g_x and f_x are well chosen.

In probability theory, very little is said about events which are generally identified as parts of the sample space Ω. In computer science, object oriented languages consider more general events called objects or "frames" defined by intension. In data analysis (multidimensional scaling, clustering, exploratory data analysis etc.) more importance is given to the elementary objects which belong to the sample Ω than in classical statistics where attention is focused on the probability laws of Ω; however, objects of data analysis are generally identified with points of $\mathbb{R}^p$ and hence are inadequate to treat complex objects coming for instance from large data bases, and knowledge bases. Our aim is to define complex objects called "symbolic objects", inspired by those of object oriented languages in such a way that data analysis becomes generalized into knowledge analysis. Objects may be defined intensionally by the properties of a generic element of the class that they represent; we distinguish these kinds of objects rather than "elementary observed objects" which characterize "individual things": for instance "the customers of my shop" instead of "a customer of my shop", a "species of mushroom" instead of "the mushroom that I have in my hand". Symbolic objects extend classical objects of data analysis in two ways: first, in case of "elementary objects" which represents individual things, by giving the possibility of introducing in their definition, structured information (see the case of "horde" in section 2 for the description of an image), probabilities (subjective or objective), possibilities (in case of vagueness and imprecision for instance), belief (in case of probabilities only known on parts and to express ignorance); second, in case of objects which are described intensionally, by the same possibilities as in the case of elementary objects, plus the possibility of expressing variation for the values taken by each variable among the member of their extension ([*color* = {*red*, *white*}]) and also by expressing constraints between these values with rules (if [*color* = *white*] then [*weight* $\leq$ 350]).

By extending data analysis methods to symbolic objects this paper makes a bridge between several domains: "data analysis and statistics" (where limited interest has, as yet, been shown in treating this kind of objects), "statistical data bases" (where symbolic objects may be considered as "metadata" which means data on the data), "management of uncertainty in knowledge-based systems" (where the emphasis is now more on knowledge representation and reasoning then on data analysis), "learning machine" (where this kind of objects as input and classical methods of data analysis has been neglected) and more generally in AI (where the results here obtained, in theorems 1, 2, 3, concern metaknowledge or knowledge on knowledge).

We have not used the notion of "predicates" from classical logic, firstly, because by using only mappings or functions, things seem more understandable,

Ω	y_1
w_1	1
w_2	2
w_3	2
w_4	1

Ω	y_2
w_1	1
w_2	2
w_3	2
w_4	3

Δ_1	a_1
δ_1^1	1
δ_2^1	2

Δ_2	a_2
δ_1^2	1
δ_2^2	2
δ_2^2	3

Figure 1. Any element of D, B or a may be considered as a symbolic object.

especially to statisticians; secondly, because they cannot be used easily in the case of probabilist, possibilist and belief objects where uncertainty is present.

2. Symbolic objects

2.1. DEFINITION OF SYMBOLIC OBJECTS

We denote Ω a set of elementary things called "individual objects", Δ a set of possible descriptions of Ω, y a mapping $\Omega \to \Delta$ (see figure 1) which associates to any $\omega \in \Omega$ its description $\delta = y(\omega)$; D is a set of description of subsets of Ω, Y_Ω is a mapping $P(\Omega) \to D$, where $P(\Omega)$ is the power set of Ω, which associates to any $\Omega' \subseteq \Omega$ its description $d \in D$; Y is a mapping $P(\Omega) \to P(\Delta)$ such that $Y(\Omega') = \Delta'$ iff $\Delta' = \{y(\omega)/\omega \in \Omega'\}$; Y_Δ is a mapping $P(\Delta) \to D$ which associates to any $\Delta' \subseteq \Delta$ a description $d \in D$ which satisfies at least the following property: $Y_\Delta(\Delta') \subseteq D$; A is a set of mappings $\Omega \to L$ where $L = \{true, false\}$ in this section (more generally $L = [0,1]$ in section 3); h_Ω is a mapping $D \to A$ such that $h_\Omega(d) = a$ where a is the mapping $\Omega \to \{true, false\}$ such that $a(\omega) =$ true iff $y(\omega) = \delta \in d$; B is the set of mappings $D \to L = \{true, false\}$ such that $h_\Delta(d) = b$ where b is the mapping $\Delta \to \{true, false\}$ such that $b(\delta) = true$ iff $\delta \in d$; we denote $\mathsf{a} = h_\Omega(D)$ and $B = h_\Delta(D)$; Z is a mapping $B \to \mathsf{a}$ such that $Z(b) = a$ iff $a = b \circ y$.

An intension of a set of individual objects $\Omega' \subseteq \Omega$ may be defined by $d = Y_\Omega(\Omega')$, $a = h_\Omega(Y_\Omega(\Omega'))$, or $b = h_\Delta(Y_\Omega(\Omega'))$. In section 2.4 we compare these different kinds of intension. The extension of a in Ω is a subset of Ω denoted $\mathrm{Ext}(a/\Omega)$ and defined by $\mathrm{Ext}(a/\Omega) = \{\omega \in \Omega / a(\omega) = true\}$; the extension of b is a subset of Δ defined by $\mathrm{Ext}(b/\Delta) = \{\delta \in \Delta / b(\delta) = true\}$; the extension of $d \in D$ in X is denoted $\mathrm{Ext}(d/X)$; by definition, we set $\mathrm{Ext}(d/\Omega) = \mathrm{Ext}(a/\Omega)$ and $\mathrm{Ext}(d/\Delta) = \mathrm{Ext}(b/\Delta)$.

E_Δ is the mapping $B \to P(\Delta)$ such that $E_\Delta(b) = \mathrm{Ext}(b/\Omega)$, E_Ω is the mapping $\mathsf{a} \to P(\Omega)$ such that $E_\Omega(\mathsf{a}) = \mathrm{Ext}(a/\Omega)$. All these mappings are summarized in figure 1.

In statistics or in classical data analysis we study a knowledge base defined by the pair (Ω, Δ) such that the units are pairs (ω, δ) where $\omega \in \Omega$ is an individual object described by $\delta \in$ D.

In symbolic data analysis we study a knowledge base (W, X) where W is a subset of $P(\Omega)$ and X is an intension space included in D, B or $\mathfrak{a}$. Notice that in probability theory, probabilities are usually defined on the set (Δ, $P(\Delta)$).

A symbolic object is a set of properties concerning a subset of Ω. Any element of D, B or $\mathfrak{a}$ may be considered as a symbolic object; in the next section we give an example which illustrates the mappings and sets which have been defined in this section.

2.2. THE CASE WHERE DESCRIPTIONS ARE CARTESIAN PRODUCTS

In this special case we assume that Ω is described by $\Delta = O_1 \times \ldots \times O_p$ where O_i is a domain containing a set of possible values (the color of fruits, for instance) and $D = P(O_1) \times \ldots \times P(O_p)$; it results in the finite case, that card $P(\Delta)$ = card $2^{\Pi_i \mathrm{card}\, O_i}$ and $D = 2^{\Sigma_i \mathrm{card}\, O_i}$; hence D, which is included in $P(\Delta)$, is generally much smaller than $P(\Delta)$.

In this case, if $d = (V_1, \ldots, V_p)$ where $V_i \subseteq O_i$ and $h_\Delta(d) = b$, then we denote $b = \wedge_i [X_i = V_i]_\Delta$, which means that if $\omega = (x_1, \ldots, x_p)$, $b(\omega) = true$ iff the statements $x_i \in V_i$ are true; if, moreover, $h_\Omega(d) = a$ we have $a(\omega) = \wedge_i [y(\omega) \in V_i]_\Omega$ which may be written $a(.) = \wedge_i [y(.) \in V_i]_\Omega$, which is simplified to $a = \wedge_i [y = V_i]_\Omega$.

EXAMPLE

Ω is a set of fruits, Δ is the set of all possible descriptions of the fruits by their color and their weight; hence if O_1 is the set of possible weights and O_2 is the set of possible colors we have $\Delta = O_1 \times O_2$; Ω is the set whose elements are the fruits produced by a village; Y_Ω associates to the set of fruits $\Omega' \subseteq \Omega$ of a village, the smallest interval V_1 of weights in which they take their values and the union of their color V_2; hence we have $Y_\Omega(\Omega') = V_1 \times V_2 = d$, $a = h_\Omega(d) = [y_1 = V_1]_\Omega \wedge [y_2 = V_2]_\Omega$ and $b = h_\Delta(d) = [x_1 = V_1]_\Delta \wedge [x_2 = V_2]_\Delta$ where for instance, as in the example of the introduction: $V_1 = [300, 400]$ and $V_2 = \{\text{red, white}\}$; $Y(\Omega')$ is the set of descriptions Δ' of the fruits of the village and $Y_\Delta(\Delta') = d = V_1 \times V_2$.

$E_\Delta(b) = \mathrm{Ext}(b/\Delta)$ is the set Δ'' of the descriptions δ such that $b(\delta) = true$ and so on, such that $\delta \in d = V_1 \times V_2$; hence, $Y(\Omega') = \Delta' \subseteq \Delta''$; $E_\Omega(a) = \mathrm{Ext}(a/\Omega)$ is the set Ω'' of individual objects $\omega \in \Omega$ such that $a(\omega) = true$ and so on, such that $y_1(\omega) \in V_1$ and $y_2(\omega) \in V_2$, hence $\Omega' \subseteq \Omega''$.

2.3. THE CASE WHERE DESCRIPTIONS ARE CARTESIAN PRODUCTS WITH CONSTRAINTS

Constraints may appear in order to describe more precisely a set $\Omega' \subseteq \Omega$ of individual objects; for instance, in the example of the introduction we have added to the description $a = [y = [300,400]] \wedge [color = \{red, white\}]$ the constraint [if [*color* = *white*] then [*weight* $\leqslant$ 350]]. Other kinds of constraints may appear to avoid incoherences in the description of a set $\Omega' \subseteq \Omega$; for instance, if Ω' is a set of mushrooms with or without hat and one of the descriptions concerns the color of the hat, we must add the condition that there is no color of hat when there is no hat.

2.4. COMPARISONS BETWEEN THE SETS OF INTENSIONS D, $\mathfrak{a}$, B, C

These comparisons depend on the choice of y and Δ. In order to simplify, we assume that $D \subseteq P(\Delta)$, it is then easy to show that h_Δ is a bijection (which is not the case for h_Ω if y is not bijective). If y is surjective it is easy to prove that Z is injective and if y is injective that Z is surjective; therefore, if y is bijective Z becomes a bijection between B and $\mathfrak{a}$.

Two natural choices for Δ are the following: the first, denoted Δ_1, is the set of descriptions with constraints (for instance, coherent descriptions); the second, denoted Δ_2, is the set of all possible (realisable or unrealisable) descriptions. When y is bijective and $\Delta = \Delta_1$, $\Omega = \Omega_1$ is the set of all coherent or "observable" individual objects; when y is bijective and $\Delta = \Delta_2$, then $\Omega = \Omega_2$ is the set of all "possible" (realisable or unrealisable) individual objects; Ω_2 is called the set of "possibilities". In practice we have $\Omega = \Omega_0$, the set of "observed" individual objects which is not in bijection with the sets Δ_1 or Δ_2 as several individual objects may have the same description and also as some description of Δ_1 or Δ_2 may correspond to no individual object of Ω_0 ; hence we have to consider also the case where y is not bijective. We denote C the set of 1-complexes introduced by Michalski et al. [14] which elements are logical expressions of the kind $c = \wedge_i [X_i = V_i]$ where the statement $[X_i = V_i]$ means "value of X_i is one of the elements of V_i "; from the definition of B it results that $b = \wedge_i [X_i = V_i]_\Delta$ is equal to $c = \wedge_i [X_i = V_i]$ iff $\Delta = \Delta_2$. The comparison between the different sets of intensions is given in figure 2, where the sign $\Leftrightarrow$ means the existence of a bijection.

2.5. COMPLETE SYMBOLIC OBJECTS AND LATTICES ON $\mathfrak{a}$, B AND C

When we associate to an element $\Omega' \in P(\Omega)$ a symbolic object $Y_\Omega(\Omega') = d \in D$ the extension of d in Ω which is $E_\Omega(h_\Omega(d))$ contains Ω' as it is the set of $\omega \in \Omega$ such that $y(\omega) \in d$; in other words, we have $\Omega' \subseteq E_\Omega(a)$ with $a = h_\Omega(Y_\Omega(\Omega'))$; in the particular case where $\Omega' = E_\Omega(a)$, we say that a is a complete symbolic object; similarly, we say that b is a complete symbolic object iff $\Omega' = E_\Delta(b)$ with $b = h_\Delta(Y_\Omega(\Omega'))$. We

Δ	y is a bijection	y is not a bijection
Δ_1	$\mathfrak{a}$ <=> $\mathfrak{B}$ <≠>C	$\mathfrak{a}$ <≠> $\mathfrak{B}$ <≠>C
Δ_2	$\mathfrak{a}$ <=> $\mathfrak{B}$ <=>C	$\mathfrak{a}$ <≠> $\mathfrak{B}$ <=>C

Figure 2. Comparison between the sets of intensions; in any case $B \Leftrightarrow D$; C is the Michalski set of 1-complexes

denote $\mathfrak{a}_c$ (resp. B_c) the set of complete symbolic objects included in $\mathfrak{a}$ (resp. B). We define a partial order on a set of symbolic objects by stating that a symbolic object s_1 is lower than a symbolic s_2 iff the extension of s_1 is contained in the extension of s_2. We define the supremum (resp. infimum) of two symbolic objects s_1, s_2 which description is respectively $d_1 = O'_1 \times \ldots \times O'_p$ and $d_2 = O''_1 \times \ldots \times O''_2$ by $d_1 \cup d_2 = O'_1 \cup O''_1 \times \ldots \times O'_p \cup O''_p$ (resp. $d_1 \cap d_2 = O'_1 \cap O'_2 \times \ldots \times O'_p \cap O''_p$).

The smallest description of $\Omega' \subseteq \Omega$ is the intersection of all the descriptions $d \in D$, such that $E_\Omega(h_\Omega(d)) = \Omega'$. It may be shown that $\mathfrak{a}$, $\mathfrak{a}_c$ (see Diday [7]), and B_c (see Brito [2]) constitute a lattice.

EXAMPLE

Let $\Omega_3^c = \{\omega_1, \omega_2\}$ be described by y: $\Omega_3^c \to O = \{1,2\}$ such that $y(\omega_1) = 1$, $y(\omega_2) = 2$; therefore $y(\Omega_3^c) = \Delta_3 = \{\delta_1, \delta_2\}$ where $\delta_1 = 1$ and $\delta_2 = 2$; it results also that $D = P(O) = \{\{1\},\{2\},\{1,2\},\emptyset\}$.

We define the following symbolic objects of A: $a_1 = [y = 1]_\Omega$, $a_2 = [y = 2]_\Omega$, $a_3 = [y = \{1,2\}]_\Omega$ and $a_4 = [y = \emptyset]_\Omega$; we choose $\Omega = \{\omega_1\}$, therefore $a_1 = a_3$ and $a_2 = a_4$.

We define also the mappings $b_i \in B$ represented by the l-complex $c_i = Z_1(b_i)$: $c_1 = [X = 1]$, $c_2 = [X = 2]$, $c_3 = [X = \{1,2\}]$ and $c_4 = [X = \emptyset]$.

In this case, it is easy to see that the set of complete objects is $\mathfrak{a}_c = \{a_1^c, a_2^c\}$ with $a_1^c = [y = 1]$ and $a_2^c = [y = \emptyset]$. In figures 3(a), (b), (c) we represent three lattices respectively associated to $\mathfrak{a} = \{a_1 = a_3, a_2 = a_4\}$, $\mathfrak{a}_c = \{a_1^c, a_2^c\}$ and $l_c = \{c_1, c_2, c_3, c_4\}$.

2.6. CHOICE OF THE KNOWLEDGE BASE FOR A SYMBOLIC DATA ANALYSIS

We have seen in section 2.1 that a knowledge base is a pair (Ω, X) where $X = \mathfrak{a}$ or B or C; so, a natural question is to ask in which case we have to use $\mathfrak{a}$, B or C, in practice.

If we wish to take account only of the set of descriptions Δ_2 then, the best choice to make is $X = B$; this happens for instance, when the descriptions of subsets Ω' of Ω (i.e. $\Omega' \in \Omega$) have constraints and do not depend on any sample Ω; this kind of knowledge base is used when we wish to study species in biology, scenarios of

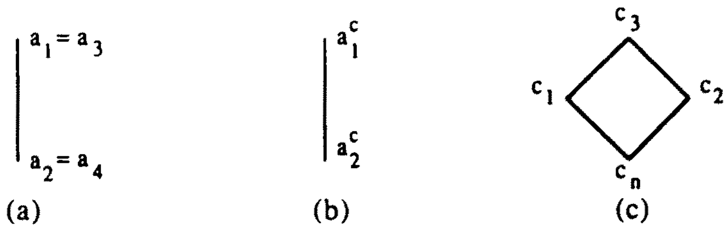

Figure 3. (a), (b), (c) represent respectively the lattice of $\mathfrak{a}$, $\mathfrak{a}_c$ and l_c. In (a) we represent the order $a_1 = a_4 < a_1 = a_3$; in (b) the order $a_2^c < a_1^c$ and in (c) $c_4 < c_1$, $c_4 < c_2$, $c_1 < c_3$, $c_2 < c_3$.

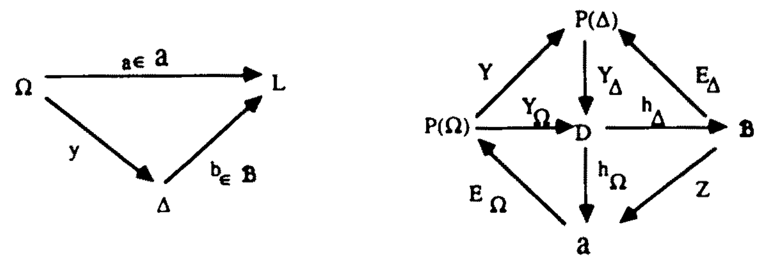

Figure 4.

accidents in transportation, teams in a company (each species, scenario or team is then an element of Ω), independently of any sample set.

If we wish to study a set Ω described without constraints and independently of Ω the best choice is $X = C$. If we wish to take into account the statistical information contained in Ω, the best choice to make is $\mathfrak{a}$; moreover, $\mathfrak{a}$ allows the possibility to compute a more simple lattice (see the previous example in section 2.5) and distances between symbolic objects when the descriptions vary; this may happen for instance when several sensors give different measures on the same set Ω, or when Ω is described by variables the values of which vary with time.

If Ω is described by two mappings y_1 and y_2 such that $y_i(\Omega) = \Delta_i = O_i$, then the mappings $a_i \in \mathfrak{a}_i$ defined by h_Ω: $D_i = P(O_i) \to \mathfrak{a}_i$ when i varies are comparable by using a dissimilarity (for instance $s(a_1,a_2) = \Sigma \{|a_1(\omega) - a_2(\omega_2)| / \omega \in \Omega\}$ whereas the mappings $h_i \in B_i$ defined by $h_{\Delta i}$: $\Delta_i \to B_i$ are not comparable when i varies.

EXAMPLE

Let $\Omega = \{\omega_1,\omega_2,\omega_3,\omega_4\}$ be a set described by two ordinal mappings y_1: $\Omega_1 \to O_1 = \{1,2\}$ and y_2: $\Omega \to O_2 = \{1,2,3\}$, as given in figure 4. Let $a_1 = [y_1 = 1]_\Omega$ and $a_2 = [y_2 = 1]_\Omega$

Considering that O_1 and O_2 are ordered sets, we may compute $s(a_1,a_2) = \Sigma_{\omega\in\Omega} \; a_1(\omega) - a_2(\omega) = 2$, whereas $c_1 = [X_1 = 1]$ and $c_2 = [X_2 = 1]$ are not comparable as they are not defined on the same set of objects, since c_1 is defined on Δ_1 whereas c_2 is defined on Δ_2.

In this paper we focus on the knowledge base $(P(\Omega), \mathfrak{a})$ because $\mathfrak{a}$ is the only set which may take into account the statistical information contained in Ω when y is not injective, and also it may take into account only the descriptions when y is bijective.

On this issue, Brito [2] focuses on the knowledge base (Ω, B) when y is not bijective and $\Delta = \Delta_2$; De Carvalho [5] focuses on the knowledge base $(\Omega, \mathfrak{a})$ when y is bijective and $\Delta = \Delta_1$. In their dissertation, Lebbe and Vignes [13] focus on (Ω, D) with $\Delta = \Delta_1$ and y not bijective.

3. Boolean symbolic objects

In this section, descriptions are Cartesian products; so, we have $\Delta = O_1 \times \ldots \times$

$O_p = O$ and $D = P(O_1) \times \ldots \times P(O_p)$; let y_i be a mapping $\Omega \rightarrow O_i$ which associates to $\omega \in \Omega$ its value $y_i(\omega)$ in the domain O_i; $y = (y_1, \ldots, y_p)$ is a mapping $\Omega \rightarrow \Delta$ such that $y(\omega) = (y_1(\omega), \ldots, y_p(\omega))$. Boolean symbolic objects are symbolic objects considered in the case where L is Boolean (i.e. L = {true, false}). Several kinds of Boolean symbolic objects may be defined in $\mathfrak{a}$: events, assertions, hordes, synthesis; we define them in the following section.

3.1. EVENTS

Let $D_i = P(O_i)$ and h_Ω^i the mapping $D_i \rightarrow \mathfrak{a}$ such that $h_\Omega^i(V_i) = e_i$ where e_i is the mapping $\Omega \rightarrow$ {true, false} such that $e_i(\omega)$ = true iff $y_i(\omega) \in V_i$. By analogy with the denominations used in probability theory (where an "event" is a subset $V_i \subset \Omega$), the basic symbolic object e_i is called an "event". In logical terms we may write $e_i(\omega) = [y_i(\omega) \in V_i]_\Omega$ where $[y_i(\omega) \in V_i]_\Omega$ is the logical proposition which is true iff $y_i(\omega) \in V$; to express the symbolic object e_i, in order to simplify notations, instead of writing $\{\forall\omega, e_i(\omega) = [y_i(\omega) \in V]_\Omega\}$ or $e_i(.) = [y_i(.) \in V]_\Omega$ we write $e_i = [y_i = V_i]_\Omega$ or more simply $e_i = [y_i = V_i]$ by dropping Ω when there is no ambiguity on its choice. For instance, if e_i = [color = {red,white}], then $e_i(\omega)$ = true iff the color of ω is red or white. When $y_i(\omega)$ is meaningless (e.g. the kind of computer used by a company without computers) $V_i = \emptyset$ and when it has a meaning but it is not known $V_i = O_i$. The extension of e_i in Ω denoted by ext (e_i/Ω) is the set of elements $\omega \in \Omega$ such that $e_i(\omega)$ = true.

3.2. ASSERTIONS

An assertion is a conjunction of events; more precisely, it is defined by the mapping h_Ω: $D = D_1 \times \ldots \times D_p \rightarrow \mathfrak{a}$ such that if $V = (V_1, \ldots, V_p)$ where $V_i \subseteq O_i$ then $h_\Omega(V) = a$ such that $a(\omega)$ = true iff $y(\omega) \in V$.

In logical terms we may write $a(\omega) = \wedge_i [y_i(\omega) \in V_i] = \wedge_i e_i(\omega)$; in conformity with the notation for an event, an assertion a is denoted $a = \wedge_i [y_i = V_i]$. For instance, if a = [color = {red, white}] $\wedge$ [height = [0, 15]], $a(\omega)$ = true iff ω is red or white and its height is between 0 and 15. The extension of an assertion denoted ext(a/Ω) is the set of elements of Ω such that $\forall i$, $y_i(\omega) \in V_i$.

3.3. HORDES AND SYNTHESIS OBJECTS

A "horde" is a symbolic object which is used when we need to describe a structure composed of several elements of Ω related together, for instance, when we need to express relations between elements of a picture that we wish to describe. It is defined by the mapping $h_\Omega = D \rightarrow H$ where H is the set of mappings $\Omega^p \rightarrow$ {true, false}, such that $h_\Omega(V) = H$ where $V = (V_1, \ldots, V_p)$ and H(u) = true where $u = (u_1, \ldots, u_p)$, iff $y_i(u_i) \in V_i$; such a horde is denoted $H = \wedge_i [y_i(u_i) = V_i]$.

Notice that if we add the constraint $u_1 = u_2 = \ldots = u_p$ a horde becomes an assertion. The extension of H in Ω^p is Ext $(H/\Omega^p) = \{\omega \in \Omega^p / H(\omega) = \text{true}\}$.

For instance, if Ω is a set of people in a town, $H = [y_1(u_1) = 1] \wedge [y_2(u_2) = 2] \wedge [y_3(u_1) = [30,35]] \wedge [\text{neighbour}(u_1, u_2) = \text{yes}]$ means that u_1 is a man, u_2 is a woman and both are neighbours.

A *"synthesis object"* is a conjunction or a semantic link between hordes denoted in the case of conjunction by $s = \wedge_i h_i$ where each horde may be defined on a different set Ω_i by different descriptors. For instance Ω_1 may be individuals, Ω_2 location, Ω_3 kind of job etc. All these objects are detailed in Diday [8].

EXAMPLE

Ω is a set of mushrooms, described by their color and their length; they are represented by two variables col_t: $\Omega \to O_c$ and ℓ_t: $\Omega \to O_\ell$ which depend upon the time t. In order to simplify we suppose that at any time, they may take only two colors and only two classes of length, such that $O_{col} = \{1, 2\}$ and $O_\ell = \{1, 2\}$. At time t_1 and t_2 we obtain the tables (a) and (b) given below for a set of two mushrooms $\Omega_1 = \{\omega_1, \omega_2\}$; table (c) represents the values taken by the elements of the set of decrivable object Ω at a given time.

Let a_1, a_2, c be three assertions where c is a ℓ-complex and c a complex such that

$$a_{t1} = [\text{col}_{t1} = 1] \wedge [\ell_{t1} = 1,2]; \quad a_{t2} = [\text{col}_{t2} = 1] \wedge [\ell_{t2} = 1,2];$$

$$c = [X_1 = 1] \wedge [X_2 = 1,2].$$

By definition, a_{t1} and a_{t2} are mappings $\Omega \to \{\text{true, false}\}$ such that

$$a_{t1}(\omega_1) = [\text{col}_{t1}(\omega_1) \in \{1\}] \wedge [\ell_{t1}(\omega_1) \in \{1,2\}] = \text{true};$$

similarly way we obtain

$$a_{t1}(\omega_2) = \text{false}, \; a_{t2}(\omega_1) = \text{false}, \; a_{t2}(\omega_2) = \text{true},$$

$$c(x_1) = c(x_2) = \text{true}, \; c(x_2) - c(x_1) = \text{false}.$$

Ω	col_{t1}	ℓ_{t1}
ω_1	1	1
ω_2	2	1

Table (a)

Ω	col_{t2}	ℓ_{t2}
ω_1	2	1
ω_2	1	2

Table (b)

O	O_1	O_2
x_1	1	1
x_2	2	1
x_3	1	2
x_4	2	2

Table (c)

It results that $ext(a_{t1}/\Omega) = \{\omega_1\}$; $ext(a_{t2}/\Omega) = \{\omega_2\}$ and $ext(c/O) = \{x_1, x_3\}$.

We may also define three hordes as follows:

$$h_1 = [col_{t1}(u_1) = 1] \wedge [\ell_{t2}(u_2) = 1, 2],$$

$$h_2 = [col_{t2}(u_1) = 1] \wedge [\ell_{t2}(u_2) = 1, 2] \text{ where } u_i \in \Omega;$$

$$h_c = [X_1(u_1) = 1] \wedge [X_2(u_2) = 1, 2] \text{ where } u_i \in O^1.$$

Therefore it is easy to see that $Ext(h_1/\Omega) = \{(\omega_1, \omega_1), (\omega_1, \omega_2)\}$, $Ext(h_2/\Omega) = \{(\omega_2, \omega_1), (\omega_2, \omega_2)\}$; $Ext(c, O) = \{(x_1, x_1), (x_1, x_2), (x_1, x_3), (x_1, x_4), (x_3, x_1), (x_3, x_2), (x_3, x_3), (x_4, x_4)\}$.

4. Modal objects

4.1. INTERNAL AND EXTERNAL MODAL OBJECTS

Suppose that we wish to use a symbolic object to represent individuals of a set satisfying the following sentence: "It is possible that their weight be between 300 and 500 grammes and their color is often red, seldom white"; this sentence contains two events $e_1 = [\text{weight} = [300, 500]]$, $e_2 = [\text{color} = \{\text{red, white}\}]$ which lack the modes *possible*, *often* and *seldom*; a new kind of events, denoted f_1 and f_2, is needed if we wish to introduce them: $f_1 =$ *possible* $[\text{weight} = [300, 500]]$ and $f_2 = [\text{color} =$ {*often* red, *seldom* white}]; we can see that f_1 contains an *external* mode *possible* affecting e_1 whereas f_2 contains *internal* modes affecting the values contained in e_2. Hence, it is possible to describe informally the sentence by a modal assertion object denoted $a = f_1 \wedge_x f_2$ where $\wedge_x$ represents a kind of conjunction related to the background knowledge of the domain. The case of modal assertions of the kind $a = \wedge_i f_i$ where all the f_i are events with external modes has been studied, for instance, in Diday [7]. This paper is concerned with the case where all the f_i contain only internal modes.

4.2. A FORMAL DEFINITION OF INTERNAL MODAL OBJECTS

Let x be the background knowledge and

- M^x a set of modes, for instance $M^x = \{\text{often, sometimes, seldom, never}\}$ or $M^x = [0, 1]$.
- $Q_i = \{q_i^j\}_j$ a set of mappings q_i^j from O_i in M^x, for instance $O_i = \{\text{red, yellow, green}\}$, $M^x = [0, 1]$ and $q_i^j(\text{red}) = 0.1$; $q_i^j(\text{yellow}) = 0.3$; $q_i^j(\text{green}) = 1$, where the meaning of the values 0.1, 0.3, 1 depends on the background knowledge (for instance q_i^j may express a possibility, see section 5.1).
- y_i is a descriptor (the *color* for instance); it is a mapping from Ω in Q_i. Notice that in the case of Boolean objects y_i was a mapping from Ω in O_i, and not Q_i.

EXAMPLE

If O_i and M^x are chosen as in the previous example and the color of ω is red then $y_i(\omega) = r$ means that $r \in Q_i$ be defined by a characteristic mapping r: r (red) = 1, r(yellow) = 0, r (green) = 0.

- $OP_x = \{\cup_x, \cap_x, c_x\}$ where $\cup_x, \cap_x$ express a kind of union and intersection between subsets of Q_i, and $c_x(q_i)$ (sometimes denoted $\bar{q}_i$)), is the complementary of $q_i \in Q_i$. To gain insight into the notion of union $\cup_x$, we may say that $q_1 \cup_x q_2$ is a "generalisation" of the observation q_1, q_2 given, for instance, by two experts or two sensors.

We denote by Q_i^x the smallest stable set for OP_x (e.g. Q_i^x is the set of any $*_x$ or c_x combination of elements $q_i^j \in Q_i$).

If $Q_x \subseteq Q_i^x$, we denote Q the mapping $Q = \cup_x \{q / q \in Q_x\}$. The complementary of Q_x in Q_i^x is $c(Q_x) = 1 - Q$.

EXAMPLE

If $q_i^j \in Q_i$ and $Q_i^j \subseteq Q_i$

$$q_i^1 \cup_x q_i^2 = q_i^1 + q_i^2 - q_i^1 q_i^2$$

$$q_i^1 \cap_x q_i^2 = q_i^2 q_i^2 \text{ where } q_i^1 q_i^2(v) = q_i^1(v) q_i^2(v); \quad c_x(q_i) = 1 - q_i$$

Intuitively, if q_i^j is the probability distribution of the words contained in a text T_i^j, then $q_i^1 \; q_i^2(v)$ is the probability of getting v among two words drawn independently one in T_i^1 and the other in T_i^2; if $P_2 > P_1$, it is less "general" to draw one word among P_1 words drawn among P_1 texts independently, than to draw one word among P_2 words drawn independently in P_2 texts.

This choice of OP_x is "Archimedian" because it satisfies a family of properties studied by Shweizer and Sklar [18] and recalled by Dubois and Prade [10]. In section 6.2 we use these operators in order to define probabilist objects.

- g_x^i is a "comparison" mapping from $Q_i^x \times Q_i^x$ in an ordered space L^x. In this paper g_x^i will not depend on i and will be denoted simply g_x.

EXAMPLE

$L^x = M^x = [0,1]$ and $g_x(q_i^1, q_i^2) = \langle q_i^1, q_i^2 \rangle$ the scalar product.

- f_x is an "aggregation" mapping from $P(L^x)$, the power set, of L^x in L^x. For instance, $f_x(\{L_1, \ldots, L_n\}) = \text{Max } L_i$.

Let $\{y_i\}$ be a set of descriptors and $\{q_i^j\}_j \subseteq Q_i^x$. Now we are able to give the formal definition of an internal modal object (called "im" object). It is a symbolic object with $D = P(Q_1^x) \times \ldots \times P(Q_p^x)$ and $h(d) = a$ where $d = (\{q_1^j\}_j, \ldots, \{q_p^j\}_j)$ and a is an im assertion defined as follows:

DEFINITION

Given OP_x, g_x and f_x, an im assertion is a mapping from Ω in an ordered space L^x, denoted $a = \wedge[y_i = \{q_i^j\}j]$, such that, if $\omega \in \Omega$ is described for any i by $y_i(\Omega) = r_i$ then a is given by: $\{\forall\omega \in \Omega,\ a(\omega) = f_x(\{g_x(\cup_{jx} q_i^j, r_i)\}i)\}$.

We denote by $\mathfrak{a}_x$ the set of im objects associated to background knowledge x, and by φ the mapping from Ω in $\mathfrak{a}_x$ such that $\varphi(\omega) = \omega^s = \wedge_{ix}[y_i = y_i(\omega)]$.

By convention, in all this paper an event $[y_i = \{q_i^j\}_j]$ may also be denoted $[y_i = q_i^1, q_i^2 \ldots]$; Notice also that it results from the definition that $[y_i\ \{q_i^j\}_j]$ is equivalent to the event $[y_i = \cup_{jx} q_i^j]$: in other words, by using the preceding notation, the event $[y_i = Q_x]$ will be considered to be equivalent to $[y_i = Q]$.

The x-union of two assertions a_1, a_2 denoted $a_i = \wedge_{ix}[y_i = q_i^j]$ is defined by $a_1 \cup_x a_2 = \wedge_{ix}[y_i = q_i^1 \cup_x q_i^2]$; more generally we have $\cup_{jx} a_j = \wedge_{ix}[y_i = \cup_{jx} q_i^j]$; hence, it results with our convention that $\cup_{jx} a_j = \wedge_i[y_i = \{q_i^j\}_j]$. The intersection of assertions is defined similarly: $\cap_{jx} a_j = \wedge_i[y_i = \cap_{jx} q_i^j]$. The operators OP_x extended on $\mathfrak{a}_x$ will be studied in greater depth in section 9.

There are at least two ways to define the extension of an im object a. The first consists in considering that each element $\omega \in \Omega$ is more or less in the extension of a according to its weight given by $a(\omega)$; in this case the extension of a denoted Ext (a/Ω) will be the set of pairs $\{(\omega, a(\omega)) \cup \omega \in \Omega\}$. The second requires a given threshold α and then, the extension of a will be Ext $(a\ \Omega, \alpha) = \{(\omega, a(\omega))\ \omega \in \Omega, a(\omega) \geqslant \alpha\}$.

4.3 SEMANTICS OF IM OBJECTS

In addition to the modes, several other notions may be expressed by an im object a:

(a) *Certainty*: $a(\omega)$ is not true or false as for Boolean objects but expresses a degree of certainty.

(b) *Variation*: this appears at two levels in an im object denoted $a = \wedge_{ix}[y_i = \{q_i^j\}_j]$: first within each q_i^j, for instance if y_i is the color, q_i^1 (red) = 0.5, q_i^1 (green) = 0.3 means that a variation exists between the individual objects which belong to the extension of a (for instance a species of mushrooms) where some are red and others are green; second, for a given description y_i and $v \in O_i$, between the $q_i^j(v)$ when j varies (each q_i^j (v) expresses for instance the variation of the color v between different kinds of species).

(c) *Doubt*: if we say that the color of a species of mushroom is red "or" green, it is an "or" of variation, but if we say that the color of the mushroom which is in my hand is red "or" green, it is an "or" of doubt.

Hence, if we describe $\omega \in \Omega$ by $\varphi(w) = \omega^s = \wedge_j\, [y_i = y_i\,(\omega)]$ where $y_i\,(\omega) = \{r_i^j\}_j$ we express a vagueness or an imprecision in each r_i^j and a doubt among the r_i^j provided, for instance, by several experts.

4.4. AN EXAMPLE OF BACKGROUND KNOWLEDGE EXPRESSING "INTENSITY"

Here the background knowledge x is denoted i, for intensity. Each individual object $\omega \in \Omega$ is a manufactured object described by two features y_1, which expresses the degree of "roundness" and "flatness", and y_2, the "heaviness": $O_1 = \{$flat, round$\}$, $O_2 = \{$heavy$\}$; $M^i = \{$*very*, *quite*, *a little*, *very little*, *nil*$\}$.

Let a and ω^s be defined by:

$$a = [y_1 = \textit{a little}\ \text{flat},\ \textit{quite}\ \text{rounded}] \wedge_i [y_2 = \textit{a little}\ \text{heavy}]$$

$$\omega^s = [y_1 = \textit{quite}\ \text{rounded}] \wedge_i [y_2 = \textit{very}\ \text{heavy},\ \textit{quite}\ \text{heavy}].$$

(The user has a doubt for ω between *very* and *quite* heavy).

The problem is to know if it is acceptable to say that ω belongs to the class of manufactured objects described by a.

Hence q_1^1 (flat) = *a little*; q_1^1 (rounded) = *quite*; q_2^1 (heavy) = *a little*, r_1^1 (flat) = *nil*; r_1^1(rounded) = *quite*; r_2^1(heavy) = *very*, r_2^2 = (heavy) = *quite*. A given taxonomy Tax which expresses the background knowledge on the values of M^i makes it possible to say that Tax (*very*, *quite*) = *somewhat*; hence if we set $r_1^1 \cup_i r_2^2(v) = \text{Tax}(r_2^1(v),\ r_2^2(v))$, we have $r_2^1 \cup_i r_2^2$ (heavy) = Tax (*very*, *quite*) = *somewhat*.

We define L^j by L_1 = not acceptable, L_2 = acceptable, L_3 = completely acceptable and we suppose that the comparison mapping g_i is given by a table T_{gi} such that $g_i(q_1^1, r_1^1) = T_{gi}$ ((*a little* flat, *quite* rounded), (*nil* flat, *quite* rounded)) = acceptable; $g_i(q_2^1,\ r_2^1 \cup_i r_2^2) = T_{gi}$ (*a little* heavy, *somewhat* heavy) = not acceptable.

Finally if we set $f(\{L_j\}) = \text{Min}\ L_j$ and $L_1 < L_2 < L_3$, we obtain $a(\omega) = f_i\,(g_i\,(q_1^1,\ r_1^1),\ g_i\,(q_2^1,\ r_2^1 \cup_i r_2^2)) = f_i$ (acceptable, not acceptable) = not acceptable.

Notice that more complex objects may occur when instead of only one, as in the preceding definition, several events concern the same variable; for instance if we have $a = \wedge_i\, a_i$ with $a_i = \wedge_{l,x}\, [y_i = q_i^l]$; in this case, it is necessary to introduce a third mapping h from $P(L^x)$ in L^x such that $a_i\,(\omega) = h\,(\{g(q_i^l,\ r_i)\}_l)$; hence, more generally, if $a = \wedge_{i,x}\, a_i = \wedge_{i,x} \wedge_{l,x} [y_i = q_i^l]$ then $a(\omega) = f_x\,(\{a_i\,(\omega)\}_i) = f_x\,(\{h_x\,(\{g_x\,(q_i^l,\ r_i)\}_l)\}_i)$. The following example may be omitted in a first lecture, its aim is to build an assertion a_i formed by the conjunction of the events for which extension at level $\frac{1}{2}$ contains a given $\omega \in \Omega$.

EXAMPLE

Let $M_i^x = [0,1]$, $O_i = \{v_1, v_2\}$, and Q_i be the set of probability measures

$P(O_i) \to [0,1]$; y is a mapping from a set Ω in Q_i and $\omega \in \Omega$ is described by $\omega^s = [y_i = r]$ is such that $r(v_1) = r(v_2) = \frac{1}{2}$; the set of im events $e_i = [y = q_i]$ such that $a_i(\omega) \geqslant \frac{1}{2}$ is defined by the set of probability measures q_i which satisfy the inequality $e_i(\omega) = f_x(g_x(q_i, r)) \geqslant \frac{1}{2}$; if f_x is the mean and g_x is the scalar product we get $e_i(\omega) = \text{Mean}(\{\langle q_i, r\rangle\}) = \langle q_i, r\rangle$ as there is only one variable. Hence q_i has to satisfy the following inequality: $e_i(\omega) = \langle q_i, r\rangle = q_i(v_1)r(v_1) + q_i(v_2)r(v_2) \geqslant \frac{1}{2}$ which is equivalent to $\frac{1}{2}q_i(v_1) + \frac{1}{2}q_i(v_2) \geqslant \frac{1}{2}$, which is satisfied by any event e_i, as $q(v_1) + q(v_2) = 1$ for any measure of probability q defined on O_i. If $a_i = \wedge_{fx}\{e_i^f / e_i^f(\omega) \geqslant \frac{1}{2}\}\}$ then $a_i(\omega) = h_x(\{e_i^f(\omega)\}_f)$; if $h_x = \text{Min}$ then $a_i(\omega) = \text{Min}(\{e_i^f(\omega)\}_f) = \frac{1}{2}$.

5. Possibilist objects

5.1. THE POSSIBILIST APPROACH

Here we follow Dubois and Prade [10] in giving the main idea of this approach.

DEFINITION

This is a mapping Π from $P(\Omega)$ the power set of Ω in [0, 1] such that:

(1) $\Pi(\Omega) = 1 \qquad \Pi(\phi) = 0$

(2) $\forall A, B \subseteq \Omega \ \Pi(A \cup B) = \text{Max}(\Pi(A), \Pi(B))$

A measure of necessity is a mapping from $P(\Omega)$ in [0, 1] such that:

(3) $\forall A \subseteq \Omega \ N(A) = 1 - \Pi(\overline{A})$.

The following properties may then be shown:

$$N(\phi) = 0;\ N(A \cap B) = \text{Min}(N(A), N(B));\ \Pi(\cup_i A_i) = \text{Max}_i(\Pi(A_i));$$
$$N\cap_i A_i) = \text{Min}_i(N(A_i));\ \Pi(A) \leqslant \Pi(B) \text{ if } A \subseteq B;\ \text{Max}(\Pi(A), \Pi(\overline{A})) = 1;$$
$$\text{Min}(N(A), N(\overline{A})) = 0;\ \Pi(A) \geqslant N(A);\ N(A) > 0 \text{ implies } \Pi(A) = 1;$$
$$\Pi(A) < 1 \text{ implies } N(A) = 0;\ \Pi(A) + \Pi(\overline{A})\ 1 \text{ and } N(A) + N(\overline{A}) \leqslant 1.$$

EXAMPLE

We define $\Pi_E(A)$ (resp. $N_E(A)$) as the possibility (resp. the necessity) that $\omega \in A$ when $\omega \in E$. We say that $\Pi_E(A) = 1$ if this possibility is true and $\Pi_E(A) = 0$ if not. Hence Π_E and N_E are mappings from $P(\Omega)$ in {0, 1}. It is then easy to show that Π_E and N_E satisfy the three conditions of their definition.

The theory of possibility models several kinds of semantics; generally possibilities valuate vague observations of inaccessible characteristics, for instance:

(i) The physical possibility: this expresses the material difficulty for an action to occur. For instance if several experts have described that an athlete has the possibility $\Pi(\{200\})=0.8$ of carrying 200kg and the possibility $\Pi(\{250\})=0.5$ of carrying 250kg; then, for these experts, the possibility of carrying 200 or 250kg for this athlete will be $\Pi(\{200\}\cup_p\{250\})=\text{Max}(\{200\},\{250\})=0.8$.

ii) The possibility as a concordance with actual knowledge: "it is possible that it will rain or snow today".

iii) The non-astonishment: for instance, "the "typicality" for the color of a flower to be yellow or brown".

5.2. A FORMAL DEFINITION OF POSSIBILIST OBJECTS

Here the background knowledge x is denoted p for possibility.

DEFINITION

A possibilist assertion denoted $a_p = \wedge_i\ [y_i = \{q_i^j\}_j]$ is an im assertion which takes its values in $L^p = [0, 1]$ such that

- $\forall i\ Q_i$ is a set of measures of possibility.
- OP_p: $\forall i,\ q_i^1,\ q_i^2 \in Q_i\quad q_i^1 \cup_p q_i^2 = \text{Max}\ (q_i^1,\ q_i^2)$; $q_i^1 \cap_p q_i^2 = \text{Min}\ (q_i^1,\ q_i^2)$; $c_p(q) = 1 - q$ denoted also $\bar{q}$.
- g_p: $g_p\ (q_i^1,\ q_i^2) = \sup\{\text{Min}\ (q_i^1\ (v),\ q_i^2\ (v))/v \in O_i\}$
- f_p: $\forall L \subseteq [0, 1]\ f_p\ (L) = \text{Max}\ (\ell/\ell \in L)$

Notice that OP_p is defined as in fuzzy sets and g_p has also been proposed by Zadeh [19]. Notice also that $q_i^1 \cap_p q_i^2$ is not necessarily a measure of possibility.

It is also possible to define a "necessitist" assertion a_n (thanks to M.O. Menessier, D. Dubois and H. Prade, for their useful remarks which have allowed me to improve this point) by setting: $a_n = 1 - \bar{a}_p$ where $\bar{a}_p = \wedge_{ip}\ [y_i = \bar{q}_i]$ and $\bar{q}_i = c_p(q_i) = 1 - q_i$.

This results in $a_n\ (\omega) = 1 - f_p(\{g_p\ (\bar{q}_i,\ r_i)\}_i)$ and then

$$\begin{aligned} a_n(\omega) &= 1 - \underset{i}{\text{Max}}\ g_p(\bar{q}_i, r_i) \\ &= 1 - \text{Max}\{\sup\{\text{Min}(\bar{q}(v), r_i(v)/v \in O_i\}\}_i \\ &= \text{Min}\{1 - \{\sup \text{Min}(\bar{q}(v), r_i(v))/v \in O_i\}\}_i \\ &= \text{Min}\{\inf\{1 - \text{Min}(\bar{q}_i(v), r_i(v)/v \in O_i\}\}_i \\ &= \text{Min}\ \inf\{\text{Max}(q_i(v),\ 1 - r_i(v))/v \in O_i\} \end{aligned}$$

and then finally $a_n\ (\omega) = \text{Min}\ g_n\ (q_i,\ \bar{r}_i)$.

It results that a necessitist object is defined by $OP_n = \{\cup_n, \cap_n, c_n\}$ where $\cup_n$ is $\cap_p$, $\cap_n$ is $\cup_p$ and c_n is c_p, $g_n\ (q_i,\ r_i) = \inf\{\text{Max}\ (q_i\ (v),\ \bar{r}_i(v))/v \in O_i\}$ and $f_n = \text{Min}$.

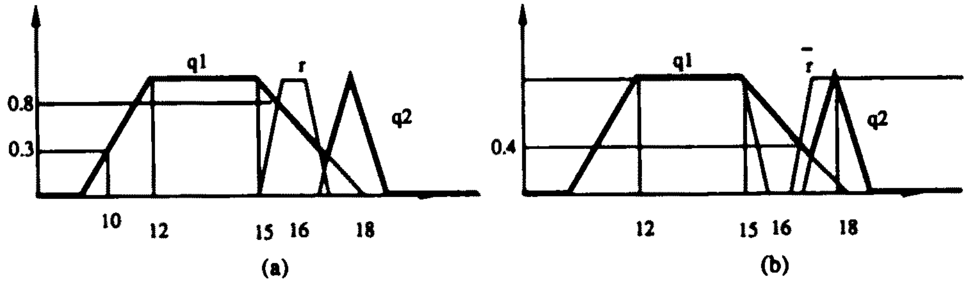

Figure 5. (a) $q_1 \cup q_2 = \mathrm{Max}(q_1, q_2)$, (b) $\bar{r}_1 = 1 - r_1$.

EXAMPLE

An expert describes a class of objects by the following possibilist assertion (restricted, to simplify, to a single event):

e_p = [height = [around [12, 15], about {18}]]. An elementary object ω is defined by ω^s = [height = close to 16].

The question is to find the possibility and necessity of ω knowing e_p, in the case where e_p and ω^s may be written: e_p = [height = q_1, q_2] and ω^s = [height = r_1] where q_1, q_2, r_1 are possibilist mappings from O = [0, 20] in [0, 1] defined by the background knowledge in figure 5. This means that an object of height 14 (resp. 10) has a possibility 1 (resp. 0.3). It is then possible to compute the possibility of ω by

$$e_p(\omega) = g_p(q_1 \cup_p q_2, r_1) = \sup\{\mathrm{Min}(q_1 \cup_p q_2(v), r_1(v))/v \in O\} = 0.8.$$

The necessity of ω is given by:

$$e_n(w) = g_n(q_1 \cup_p q_2, (r_1), = \inf\{\mathrm{Max}(q_1 \cup_p q_2(v), \bar{r}_1(v))/v \in O = 0.4.$$

This example shows that possibilist objects are able to represent not only certainty, variation and doubt but also inaccuracy (around, about, close to); it is also possible to use vagueness, in representing for instance "high" or "heavy" by a measure of possibility.

5.3. THE PARTICULAR CASE OF BOOLEAN OBJECTS

A Boolean object $a = \wedge_i [y_i = V_i]$ is an im object $a_b = \wedge_i [y_i = q_i]$ where q_i is the characteristic mapping of $V_i \subseteq O_i$, $OP_b = \{\cup_b, \cap_b, c_b\}$is such that $q_1 \cup_b q_2 = \mathrm{Max}(q_1, q_2)$, $q_1 \cap_b q_2 = \mathrm{Min}(q_1, q_2)$ and $c_b(q) = 1 - q$. There are two choices for g_b and f_b: $(g_b, f_b) = (g_p, f_p)$ or $(g_b, f_b) = (g_n, f_n)$. If $\omega^s = \wedge_i [y_i = r_i]$ where r_i is the characteristic mapping of $y_i(\omega) \subseteq O_i$, (there is doubt if $y_i(\omega)$ is not reduced to a single element), it is then easy to show that in the possibilist choice $y_i(\omega) \cap V_i \neq \phi \Leftrightarrow$

$a_b(\omega) = 1$ and in the necessitist choice $y_i(\omega) \subseteq V_i \Leftrightarrow a_b(\omega) = 1$. If we denote $|a|_\Omega$ the set of elements of Ω such that $a(\omega) =$ true, we have $|a|_\Omega = \text{Ext}(a_b/\Omega, \alpha)\ \forall a \in]0,1]$, for both choices.

6. Probabilist objects

6.1. THE PROBABILIST APPROACH

First we recall the well known axioms of Kolmogorov:

If $C(\Omega)$ is a σ-algebra on Ω (i.e. a set of subsets stable for countable intersection or union and for complementation). We say that p is a measure of probability on $(\Omega, C(\Omega))$ if

(i) $p(\Omega) = 1$
(ii) $p(\cup_i A_i) = \Sigma\, p(A_i)$ if $A_i \in C(\Omega)$ and $A_i \cap A_j = \phi$.

There are several semantics which follow these axioms: for instance luck in games, frequencies, some kind of uncertainty by subjective probability. Let Q_i be a set of measures of probabilities defined on $(O_i, C(O_i))$. We suppose that the $\omega^s = \wedge_i[y_i = y_i(\omega)]$ are such that $y_i(\omega) \in Q_i$. We recall that Q_i^x has been defined in section 4.1.

6.2. A FORMAL DEFINITION OF PROBABILIST OBJECTS

DEFINITION

A probabilist assertion is an im assertion which takes its values in $L^{pr} = [0,1]$

OP_{pr}: $\forall q_i^1, q_i^2 \in Q_i\ q_i^1 \cup_{pr} q_i^2 = q_i^1 + q_i^2 - q_i^1 q_i^2;\ q_i^1 \cap_{pr} q_i^2 = q_i^1 q_i^2$ which is the mapping which associate to $v \in O_i$, $q_i^1(v)\, q_i^2(v)$; $c_{pr}(q) = \bar{q} = 1 - q$.
g_{pr}: $\forall \{q_i^1, q_i^2\} \in Q_i^x \times Q_i\ \ g_{pr}(q_i^1, q_i^2) = \langle q_i^1, q_i^2 \rangle = \Sigma\, \{q_i^1(v)\, q_i^2(v) / v \in O_i\}$.
f_{pr}: $f_{pr}(\{L_i\}) =$ mean of the L_i.

Notice that if there are some characteristic dependencies between variables, then, an event of the form $[y_i = q_i]$ may represent them; for instance, if the expert wishes to describe the dependencies between y_1, y_3, y_7, then this information may be represented by the event denoted $[y_{137} = pr(y_1, y_3, y_7)]$ where $pr(y_1, y_3, y_7)$ represents the joint probability of y_1, y_3, y_7; this event is of the form $[y_i = q_i]$ where $y_i = y_{137}$ and $q_i = pr(y_1, y_3, y_7)$. In the case where "causalities" or "influences" among sets of variables are given by the expert to describe a symbolic object, propagation techniques (see Pearl [15]), Lauritzen and Spiegelhalter [12] may be used which induce other mappings g_{pr} and f_{pr}.

To give an intuitive idea of the notion of union and intersection of measures of probabilities it is easy to see that if q_i^1 and q_i^2 are the measures of probabilities associated to two dices, $q_i^1 \cup_{pr} q_i^2(v)$, with $v \in O_i$, is the probability that the event v occurs, for one dice **or** (not exclusive) for the other, $q_i^1 \cap_{pr} q_i^2(v)$ is the probability that the event v occurs for both dices when the two dices are thrown independently. This comes from the fact that if (X_1, X_2) is a pair of random variables $\Omega \to O_i \times O_i$ where $O_i = \{1, 2, \ldots, 6\}$ with probability $(q_i^1 q_i^2)$, then the probability that the number j occurs in both dices thrown independently is $q_i^1 \cap_{pr} q_i^2(j) = Pr((X_1, X_2) = (j, O) \cap (O, j)) = Pr((X_1, X_2) = (j, j)) = Pr(X_1 = j) Pr(X_2 = j) = q_i^1(j) q_i^2(j)$; the probability that the number j occurs in one or the other dice is: $q_i^1 \cup_{pr} q_i^2(j) = Pr((X_1, X_2) = (j, O_i) \cup (O_i, j)) = Pr((X_1, X_2) = (j, O_i)) + Pr((X_1, X_2) = (O_i, j)) - Pr((X_1, X_2) = (O_i, j) \cap (j, O_i)) = q_i^1(j) q_i^2(O_i) + q_i^1(O_i) q_i^2(j) - q_i^1(j)\ q_i^2(j) = (q_i^1 + q_i^2 - q_i^1 q_i^2)(j)$.

Notice that $q_i^1 \cup_{pr} q_i^2$ is not a measure of probability because even if $q_i^1 \cup_{pr} q_i^2(v) \in [0, 1]$ the sum of the $q_i^1 \cup_{pr} q_i^2(v)$ on O_i is larger than 1. Also, $q_i^1 \cap_{pr} q_i^2$ is not a measure of probability because the sum of the $q_i^1 \cap_{pr} q_i^2$ (v) on O_i may be lower than 1. We have defined g on Q_i x Q_i^x and not on Q_i^x x Q_i^x as for a general im object, because for instance, g $(q_i^1 \cup_{pr} q_i^2, \cup_j q_i^j)$ may become larger than 1; but notice that in this case, it is easy to transform $q_i^1 \cup_{pr} q_i^2$ in a probability measure by dividing it by the sum of the $q_i^1 \cup_{pr} q_i^2(v)$ on O_i.

EXAMPLE

An object ω is described by its color $y_1(\omega)$ which may be red or blue and its roundness $y_2(\omega)$ which may be round or flat.

Let $a = [y_1 = q_1^1, q_1^2] \wedge_{pr} [y_2 = q_2]$ and $\omega^s = [y_1 = r_1] \wedge_{pr} [y_2 = r_2]$ where q_1^1 (red) = 0.9;

q_1^1 (blue) = 0.1; q_1^2 (red) = 0.5; q_1^2 (blue) = 0.5; q2 (round) = 0.2; q2 (flat) = 0.8. It results that a is described by two kinds of objects: either often red and rarely blue, or red or blue with equal probability.

By using $q_1^3 = q_1^1 \cup_{pr} q_1^2 = q_1^1 + q_1^2 - q_1^1 q_1^2$ we obtain

$$q_1^3(\text{red}) = 0.9 + 0.5 - 0.9 \times 0.5 = 0.95,$$

$$q_1^3(\text{blue}) = 0.1 + 0.5 - 0.1 \times 0.5 = 0.55.$$

If r_1 and r_2 are defined as follows:

r_1 (red) = 1, r_1 (blue) = 0; r_2 (round) = 1, r_2 (flat) = 0, it results that a $(\omega) = g_{pr}(q_1^3, r1) \wedge_{pr} g_{pr}(q_2, r_2) = (0.95 \times 1 + 0.55 \times 0) \wedge_{pr} (0.2 \times 1 + 0.8 \times 0) = 0.95 \wedge_{pr} 0.20 + \frac{1}{2}(0.95 + 0.20) = 0.57$, which represents the average probability that an instance of the class of objects described by a be ω and may be interpreted as a kind of membership degree for ω to the im object defined by a.

7. Belief objects

7.1. THE BELIEF FUNCTION FORMALISM

At the origin of this theory we may mention at least the work of Choquet [4] on "Capacities of order 2" and Dempster [6] on "upper and lower probabilities induced by a multivalued mapping". The basic notions of this formalism are in Schafer's book [17]: "A mathematical theory of evidence" which is "still a standard reference for this theory" (Schafer [16]). First a "probability assignment" function m from $P(\Omega)$ (the power set of Ω, supposed finite) in [0,1] is defined by: $\Sigma\{m(V)/V \in P(\Omega)\} = 1$ and $m(\phi) = 0$; then a belief function Bel: $P(\Omega) \to [0,1]$ is defined by:

$$\text{Bel}(A) = \sum m(V)/V \in P(\Omega),\ V \subseteq A.$$

A "body of evidence" is viewed as a pair (F,m) where m is a probability assignment function and $F = \{V \in P(\Omega)/m(V) \neq 0\}$ is the set of "focal" elements. Given a body of evidence it is possible to define exactly a belief function; it is also possible to define a "plausibility" function Pl : $P(\Omega) \to [0,1]$ such that:

$$\text{Pl}(A) = \sum \{m(V)/V \in P(\Omega), V \cap A \neq \phi\}$$

and then we have: $\text{Bel}(A) = 1 - \text{Pl}(\bar{A})$.

It may be proved (Schafer [17]) that we have the following properties: Bel is a belief function iff:

(i) $\text{Bel}(\Omega) = 1$

(ii) $\text{Bel}(\phi) = 0$

(iii) $\text{Bel}(A_1 \cup \ldots \cup A_n) \geqslant \sum_i \text{Bel}(A_i) - \sum_{i<j} \text{Bel}(A_i \cap A_j)$

$$+ \ldots = \sum_{\substack{I \subseteq \{1,\ldots,n\} \\ I \neq \phi}} (-1)^{|I|+1} \text{Bel}\left(\bigcap_{i \in I} A_i\right)$$

As a consequence of (iii) we get:

$$\text{Pl}(A_1 \cap \ldots \cap A_n \leqslant \sum_i \text{Pl}(A_i) - \sum_{i<j} \text{Pl}(A_i \cup A_j) + \ldots$$

Given a belief function Bel, the basic probability assignment function m related to Bel is obtained by:

$$\forall A \subseteq P(\Omega) \qquad m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} \text{Bel}(B).$$

Given two belief functions Bel_1 and Bel_2, their orthogonal sum $\text{Bel}_1 \oplus \text{Bel}_2$, also known as Dempster's rule of combination, is defined by their associated probability assignments:

$$m_1 \oplus m_2 (A) = \sum_{V_1 \cap V_2 = A} m_1(V_1)\, m_2(V_2) / \sum_{V_1 \cap V_2 \neq A} m_1(V_1)\, m_2(V_2)$$

As a special case, we get a generalization of Bayes rule of conditioning, which is known as Dempster's conditioning:

$$\mathrm{Bel}(A/B) = \frac{\mathrm{Bel}(A \cup \overline{B} - \mathrm{Bel}(\overline{B})}{(1 - \mathrm{Bel}(\overline{B}))}$$

We have the following link with probability and possibility theories: it may be shown that if F contains only singletons then Bel is a classical probability measure. Dempster [6] said that Pl and Bel may be viewed as upper and lower probabilities. Schafer [17] has shown that if F contains only a nested sequence of subsets $V_1 \subseteq V_2 \subseteq \ldots \subseteq V_n$ then we have: Bel $(A \cap B) = \mathrm{Min}(\mathrm{Bel}(A), \mathrm{Bel}(B))$ and $\mathrm{Pl}(A \cup B) = \mathrm{Max}(\mathrm{Pl}(A), \mathrm{Pl}(B))$ and hence, in this case, Bel and Pl satisfy respectively the properties of necessity and possibility measures. Given a probability measure pr, it may be shown that there exists a possibility, necessity, belief and plausibility function respectively denoted pos, nec, bel, pl, such that nec $\leqslant$ bel $\leqslant$ pr $\leqslant$ pl $\leqslant$ pos.

The theory of evidence models several kinds of knowledge:

(i) *Probability*: as said by Pearl [15]: "Belief functions result from assigning probabilities to sets rather than to individual points".

EXAMPLE

A machine is able to compute the average number of vehicles whose speeds vary within a set of a priori given intervals for instance $V_1 =]0,110]$. Sometimes this machine may fail to give the speed but still be able to give the number of vehicles which pass on the road. If the machine gives for instance the following percentage: 0.40 for speeds which belong in the interval V_1, 0.50 for speeds which belong in $V_2 = \{\text{speed} > 110\}$ and 0.10 for unknown speeds, we may represent this information by a belief function q with body of evidence (F,m) such that $F = \{V_1, V_2, \mathbb{R}^+\}$, $m(V_1) = 0.40$, $m(V_2) = 0.50$, $m(\mathbb{R}^+) = 0.10$. Then we have, for instance, bel $([0,130]) = 0.40$ and $P_\ell\,([0,130]) = 0.40 + 0.50 = 0.90$.

(ii) *Testimony*: if two witnesses observe the same event A, then by using the Dempster rule it may be shown that the belief in A increases. If one observes A and the other B with $A \neq B$ and $A \cap B \neq \phi$ then it may be shown that the belief in A and B decreases. If $A \cap B = \phi$ the belief in A and B decreases more than in the preceding case and the higher the belief in B, the lower the belief in A.

EXAMPLE

After an accident observed by two witnesses, the first one is almost sure that the speed of the vehicle was in the interval $V_1 =]0,100$ km] and the second witness who was further away, thinks the same thing but is less sure. Hence, each witness may be represented by a belief function, the first one by q_1, with body of evidence $\{F_1, m_1\}$ such that $F_1 = [V_1, \mathbb{R}^+]$, $m_1(V_1) = 0.90$ and q_2 defined by $\{F_2, m_2\}$ such that: $F_2 = F_1$ and $m_2(V_1) = 0.70$. Then by using the Dempster rule we get:

$$q_1 \oplus q_2(V_1) = q_1(V_1) + q_2(V_1) - q_1(V_1)q_2(V_1) = 0.90 + 0.70 - 0.63 = 0.97.$$

7.2 A FORMAL DEFINITION OF "BELIEF OBJECTS"

Following Dubois and Prade [11], we define the union and intersection of two bodies of evidence (F_1, m_1) and (F_2, m_2) as follows:

$$\forall A \in P(\Omega), \quad m_1 \cup_{bel} m_2(A) = \sum_{V_1 \cup V_2 = A} m_1(V_1)m_2(V_2);$$

$m_1 \cap_{bel} m_2(A) = \sum_{V_1 \cap V_2 = A} m_1(V_1)\, m_2(V_2)$ which is consistent with Dempster's rule if the term $m_1 \cap m_2(\emptyset)$ (which reflects the amount of dissonance between the sources or their independence) is eliminated. In the following definition we denote by q_i^j a belief function with body of evidence (F_i^j, m_i^j).

DEFINITION

A belief assertion denoted $a_{bel} = \wedge_{i\,bel}[y_i = \{q_i^j\}_j]$ is an im assertion which takes its values in $L^{bel} = [0,1]$ such that:

- $\forall i$, Q_i is a set of belief functions defined on O_i
- OP_{bel}: $\forall i$, $q_i^1, q_i^2 \in Q_i$ $q_i^1 \cup_{bel} q_i^2(V) = \sum_{A \subseteq V} m_i^1 \cap_{bel} m_i^2(A)$; $q_i^1 \cap_{bel} q_i^2(V) = \sum_{A \subseteq V} m_i^1 \cup_{bel} m_i^2(A)$; the complement is defined by $c_{bel}(q_i^j)(V) = \bar{q}_i \sum_{A \subseteq V} m_i^1(A)$ where $\bar{m}_i^1(A) = m_i^j(A)$).

g_{bel}: $g_{bel}(q_i^1, q_i^2) = \sum \{m_i^1 \cap_{bel} m_i^2(V_2)\ V_2 \subseteq V_1, (V_1, V_2) \in F_1 \times F_2\}$
f: the mean.

Notice that the union and intersection of belief functions remain belief functions (unlike in the case of probabilities and possibilities).

As in the case of probabilist objects, the choice of the function f may be more general; we have chosen the mean in order to simplify. It is also possible to define a plausibilist object by

$OP_{p\ell}$: $q_i^1 \cup_{p\ell} q_i^2(V) = \sum_{A \subseteq V \neq \emptyset} m_i^1 \cap m_i^2(A)$; $q_i^1 \cap_{p\ell} q_i^2(V) = \sum_{A \cap V \neq \emptyset} m_i^1 \cup m_i^2(A)$ and $c_{p\ell}(q_i) = \bar{q}_i$ is defined as in the belief case.

$g_{p\ell}$: $g_{p\ell}(q_i^1, q_i^2) = \sum \{m_i^1(V_1)\, m_i^2(V_2) / V_1 \cap V_2 \neq \emptyset, (V_1, V_2) \in F_1 x F_2\}$ and f remains the mean.

The following properties may then be shown: $q_i^1 \cap_{bel} q_i^2 = q_i^1\, q_i^2$, because $q_i^1 \cap_{bel} q_i^2(V) = \sum_{A \subseteq V} m_i^1 \cup_{bel} m_i^2(A) = \sum_{V_1 \cup V_2 = A \subseteq V} m_i^2(V_2) = \sum_{V_1 \subseteq V} m_i^1(V_1) \sum_{V_2 \subseteq V} m_i^2(V_2)$.

We have also $g_{bel}(q_i^1, q_i^2) = \sum_{V_1 \in F1} m_i^1(V_1)\, q_i^2(V_1)$; $g_{p\ell}(q_i^1, q_i^2) = \sum_{V_2 \in F2} m_i^2(V_2)\, p\ell_i^1(V_2) = \sum_{V1 \in F1} m_i^1(V_1)\, p\ell_i^2(V_2)$ where $p\ell_i^j(V_j) = \sum_{V \cap V_j \neq \emptyset} q_i^j(V)$; hence $g_{p\ell}$ is symmetric whereas g_{bel} is not; it is also easy to show that $\forall A \in P(\Omega)$ $q_i^1 *_{bel} q_i^2(A) = 1 - q_i^1 *_{p\ell} q_i^2(\bar{A})$.

If two experts observe the same event A and are associated to the belief functions q_i^1, q_i^2 with $F_i^1 = F_i^2 = \{A, O\}$, then it may be shown that:

$$q_i^1 \cup_{bel} q_i^2 = q_i^1 + q_i^2 - q_i^1 q_i^2.$$

Let us give a simple example.

EXAMPLE

Several transportation experts define an accident scenario between a car and a bicycle by a belief function q_1 concerning the speed of the car. Knowing q_1 we are able to define a belief object $a = [speed = q_1]$ where the body of evidence of q_1 is $\{F_1, m_1\}$ such that $F_1 = \{V_1, O\}$, where O is the set of possible speeds and $V_1 \subseteq O$ is an interval of speed (for instance, $V_1 = [100, 120]$ km/h). Now suppose that a witness observes an accident and says that it is defined by a belief function q_2 with body of evidence $\{F_2, m_2\}$ such that $F_2 = \{V_2, O\}$. If we wish to know how much a given accident defined by $\omega^s = [speed = q_2]$, satisfies the scenario defined by a, we have to compute $a(\omega)$; as a is a belief object, by definition we have: $a(\omega) = \sum_{V \in F1} m_1(V)\, q_2(V) = m_1(V_1)\, q_2(V_1) + m_1(O)\, q_2(O)$, $a(\omega) = m_1(V_1)\, q_2(V_1) + m_1(O)$. Hence if $V_2 \subseteq V_1$, then $a(\omega) = m_1(V_1)\, m_2(V_2) + m_1(O)$ and the higher the witness' belief in V_2 the more ω satisfies the scenario defined by a; if $V_1 \subseteq V_2$ then $a(\omega) = m_1(O)$ and the greater the ignorance of the expert who has defined the scenario, the more ω satisfies the scenario.

8. Some qualities and properties of symbolic objects

8.1. ORDER, UNION AND INTERSECTION BETWEEN IM OBJECTS

It is possible to define a partial preorder $\leqslant_\alpha$ on the im objects by: $a_1 \leqslant_\alpha a_2$ iff $\forall \omega \in \Omega$, $\alpha \leqslant a_1(\omega) \leqslant a_2(\omega)$.

We deduce from this preorder an equivalence relation R by: a_1 R a_2 iff Ext $(a_1/\Omega, \alpha)$ = Ext $(a_2/\Omega, \alpha)$ and a partial order denoted $\leq_\alpha$ and called "symbolic order" on the equivalence classes induced from R.

We say that a_1 inherits from a_2 or that a_2 is more general than a_1, at the level α, iff $a_1 \leq_\alpha a_2$ (which implies $\text{Ext}_a(a_1/\Omega, \alpha) \subseteq \text{Ext}_a(a_2/\Omega, \alpha)$).

We call intension at the level α of a subset $\Omega_1 \subset \Omega$ the symbolic object b defined by the conjunction of events whose extension at the level α contains Ω_1. The symbolic union $a_1 \cup_{x,a} a_2$ (resp. intersection $a_1 \cap_{x,\alpha} a_2$) at the level α is the intension of Ext $(a_1/\Omega, \alpha) \cup$ Ext $(a_2/\Omega, \alpha)$ (resp. Ext $(a_2/\Omega, \alpha) \cap$ Ext $(b / \Omega, \alpha)$).

8.2. SOME QUALITIES OF SYMBOLIC OBJECTS

As in the Boolean case, see Diday [7], Brito and Diday [1], it is possible to define different kinds of qualities of symbolic objects (refinement, simplicity, completeness etc.). For instance, we say that a symbolic object s is "complete" iff the properties which characterize its extension are exactly those whose conjunction defines the object; in other words s is a complete symbolic object if it is the intension of its extension. More intuitively, if I can see some white dogs and I state "I can see some dogs", my statement does not describe the dogs in a complete way, since I am not saying that they are white.

On the other hand, the simplicity at level α of an im object is the smallest number of elementary events whose extension at level α coincides with the extension of s at the same level.

When an operator $\cup_x$ has to be defined in a domain related to a specific semantic which induces the notion of similarity between symbolic objects, it seems natural to require that it should satisfy the following intuitive properties:

(a) The union of two symbolic objects is more general than each one separately; in other words, the extension of the union of two symbolic objects contains the extension of each one.

(b) The union of an object with itself has an extension which contains the extension of this object.

(c) The more two objects are similar the less they are general.

(d) The most opposite objects (i.e. opposite in all the variables which define them) have a union whose extension contains every one.

(e) The union of two similar objects must reject, from its extension, objects which are not similar to them.

In case of intersection, analogous "natural" conditions may be defined, they express the inverse conditions, for instance: the intersection of two symbolic objects is less general than each one.

In case of probabilist and possibilist objects, it is easy to see that condition (a) is satisfied, since when q_1 and q_2 are two probabilist measures, we have: $q_1 \cup_{pr} q_2 = q_1 + q_2 - q_1 q_2 \geqslant q_k$ for $k = 1, 2$. When q_1 and q_2 are possibilist measures we have $q_1 \cup_p q_2 = \text{Max}(q_1, q_2) \geqslant q_k$ for $k = 1, 2$.

If $a_j = \wedge_i [y_i = q_i^j]$ we get $a_1 \cup_x a_2 = \wedge_i [y_i = q_i^1 \cup_x q_i^2]$ and $\forall \omega \in \Omega$: $\omega^s = \wedge_i [y_i = r_i]$ we have:

$a_1 \cup_x a_2(\omega) = f_x(\{g_x(q_i^1 \cup_x q_i^2, r_i)\}_i)$; hence,
$a_1 \cup_{pr} a_2(\omega) = \text{Mean}\{\sum_{v \in O_i} q_i^1 \cup_{pr} q_i^2(v)\, r_i(v)\}_i \geqslant \text{Mean}\{\sum_{v \in O_i} q_i^k(v)\, r_i(v)\}_i = a_k(\omega)\}$ with $k = 1, 2$.

Similarly in case of possibilities we have:

$a_1 \cup_p a_2(\omega) = \text{Max}\{\text{Max}_{v \in O_i} \text{Min}(q_i^1 \cup_{pr} q_i^2(v), r_i(v))\} \geqslant \text{Max}_i\{\text{Max}_{v \in O_i}\text{Min}(q_i^k(v), r_i(v))\} = a_k(\omega)$ with $k = 1, 2$.

It is also easy to see that the probabilist and possibilist intersection satisfies the inverse condition.

Condition (b) is proved in case of probabilist objects, by the following argument, in the case of a probabilist assertion reduced to an event, and may be easily generalized (by taking the mean) to the case of a conjunction of several events: let $a = [y = p]$; we have by definition $a \cup_{pr} a = [y = p \cup_{pr} p] = [y = 2p - p^2]$; hence $\forall \omega^s = [y = r]$, we have $a \cup_{pr} a(\omega) = \sum_{v \in O} (2p - p^2)(v)\, r(v) \geqslant \sum_{v \in O} p(v) r(v)$ and so $a \cup_{pr} a(\omega) \geqslant a(\omega)$; therefore $a \cup_{pr} a \geqslant a$.

In case of possibilist objects it is easy to see that $a \cup_p a = a$, since if q is a possibilist measure and $a = [y = q]$, then $a \cup_p a = [y = q \cup_p q] = [y = \text{Max}(q, q)] = a$.

Conditions (c) and (e) depend on the chosen similarity; with the similarity proposed in section 10.1 it may be shown that condition c) is not satisfied by probabilist objects. It is easy to show that d) is satisfied by probabilist and possibilist objects; let $a_i = [y = p_i]$ with $p_i(v_i) = 1$ and therefore $p_i(v_j) = 0$ if $v_i \neq v_j$. It results that in the probabilist case we obtain $\cup_i p_i = \mathbf{1} \in Q_i^{pr}$ where $\mathbf{1}$ is the mapping such that $\forall v$, $\mathbf{1}(v) = 1$, from which it results that for any $\omega^s = [y = r]$ where r is a probability measure, $\cup_{i\,pr} a_i(\omega) = 1$. In the case where the p_i are possibilities we get also $\cup_i p_i = \mathbf{1}$ (which is a possibility), and so, it results also that for any $\omega^s = [y = r]$ where p is a measure of possibility $\cup_{p\,i} a_i(\omega) = 1$; therefore in both cases the union of the most opposite objects are equal to Ω^s, the full object whose extension contains all the elements of Ω.

8.3. SOME PROPERTIES OF IM OBJECTS: LATTICE AND COMPLETENESS

It may be shown, see Diday [9] for instance, that given a level α, the set of im objects is a lattice for the symbolic order and that the symbolic union and intersection define the supremum and infimum of any couple. To do so, f_x, g_x and h_x (see section 3.1) have to be well chosen and we introduce a "full" and an "empty" (which could also be called "top" and "bottom") because they are the most and the less general symbolic object denoted Ω^s and ϕ such that $\forall \omega \in \Omega$, $\Omega^s(\omega) = 1$ and $\phi(\omega) = 0$; it is then easy to see that the extension of Ω^s contains all the elements of Ω (e.g. it is "full") and the extension of ϕ contains no one (e.g. it is "empty").

It may also be shown that the symbolic union and intersection of complete im objects are complete im objects and hence that the set of complete im objects is also a lattice.

9. An extension of possibilities, probabilities and belief assertions on symbolic objects

9.1. DUAL ASSERTIONS

Several kinds of valuations of symbolic objects can be studied. For instance, in case of Boolean objects we obtain a valuation, by setting $\forall A \subseteq \mathfrak{a}_b$, $a^*(A) = \text{card } A / \text{card } \mathfrak{a}_b$, (in this case the O_i must be finite) and a^* satisfy the Kolmogorov axioms; $a^*(A)$ may also be computed by taking into account the constraints which may exists among the variables (see De Carvalho [5], for more details about the constraints). Another possibility may be to consider the x-union or x-intersection of subsets of $\mathfrak{a}_x$ by the following definition where $*_x \in \{\cup_x, \cap_x\}$ $\forall A_x^1$, $A_x^2 \subseteq \mathfrak{a}_x$, $A_x^1 *_x A_x^2 = \{a_1 *_x a_2 / (a_1, a_2) \in A_x^1 A_x^2\}$ and then by studying the link between $a^*(A_x^1 \cup_x A_x^2)$, $a^*(A_x^1 \cap_x A_x^2)$, $a^*(A_x^1)$ and $a^*(A_x^2)$ (where, for instance $a^*(A_x) = \Sigma\{a^*(a_i) / a_i \in A_x\}$).

In this paper, our aim is to extend an im assertion $a = \wedge_{ix} [y_i = q_i]$ (where q_i depends on the choice of x and may be for instance a possibility, a probability or a belief function) to a dual im assertion denoted a^* defined on subsets of $\mathfrak{a}_x$ (the set of im assertions associated to x), and more generally, on "$*_x$-combinations" of such subsets of the kind $A *_x B$ where $*_x \in \{\cup_x, \cap_x\}$ and to show that a^* is itself a kind of possibility, probability or belief function depending on x.

More precisely:

Given $A_x \subseteq a_x$, we have $A_x = \{a / a \in A_x\}$ and to define $A = \cup_x \{a / a \in A_x\}$ we use the set $Q_i^{A_x} \subseteq Q_i^x$ such that $Q_i^{A_x} = \{q_i / a = \wedge_{x} [y_j = q_j] \in A_x\}$; we denote: $q_i^A = \cup_x \{q_i / q_i \in Q_i^{A_x}\}$. We define the $\cup_x$ of im assertions by: $\cup_x \{a / a \in A_x\} = \wedge_{ix} [y_i = q_i^A]$; hence, we have $A = \wedge_{ix} [y_i = q_i^A]$.

We define $a\ell^*$ a "dual" measure of $a\ell = \wedge_{ix} [y_i = q_i^\ell]$ by $a\ell^*(a_j) = f_x(\{g_x(q_i^\ell, q_i^j)\}_i)$; hence, given $A_x^k \subseteq a_x$, we denote $A_k = \cup_x \{a / a \in A_x^k\}$ and we get $a\ell^*(A_k) = f_x(\{g_x(q_i^\ell, q_i^{A_k}$; more generally $a\ell^*(A_1 *_x A_2) = f_x(\{g_x(q_i^\ell, q_i^{A_1} *_x q_i^{A_2})\}_i)$, where $*_x \in \{\cup_x, \cap_x\}$ and $q_i^{A_k} = *_x\{q_i / q_i \in Q_i^{A_x^k}\}$.

9.2. THREE THEOREMS OF META-KNOWLEDGE

The three following results [9] prove the existence of probabilist, possibilist and belief objects defined respectively on probabilist, possibilist and belief objects, themselves defined on Ω. The proof of theorems 1 and 2 is in the appendix, the proof of theorem 3 is long and will be published elsewhere.

(a) In the case of possibilist objects:

THEOREM 1

(i) $a^*(\mathrm{a}_p) = 1$ $\quad a^*(\phi) = 0$

(ii) $\forall A_1, A_2 \subseteq \mathrm{a}_p$ $\quad a^*(A_1 \cup_p A_2) = \mathrm{Max}\,(a^*(A_1), a^*(A_2))$.

(b) In the case of probabilist objects:

THEOREM 2

(i) $a^*(\mathrm{a}_{pr}) = 1$ $\quad a^*(\phi) = 0$

(ii) $\forall A_1, A_2 \subseteq \mathrm{a}_{pr}$ $\quad a^*(A_1 \cup_{pr} A_2) = a^*(A_1) + a^*(A_2) - a^*(A_1 \cap_{pr} A_2)$.

(c) In case of belief objects:

We say that there is independence between the body of evidence of two belief objects a_1 and a_2 iff $\forall i$ the bodies of evidence (F_i^j, m_i^j) associated to q_i^j for $j = 1,2$ are such that $m_i^1 \cap_{bel} m_i^2 (\phi) = 0$, (or in other words, the focal elements $V_i^1 \in F_i^1$, $V_i^2 \in F_i^2$ are such that: $V_i^1 \cap V_i^2 \neq \phi$). The body of evidence of two subsets A_1, A_2 of a_{bel} are said to be independent iff for $\forall i$ and $j = 1,2$ such that $Q_i^j = \cup_{bel} \{q_i^j / q_i^j \in Q_i^{A_j}\}$, the body of evidence of Q_i^1 and Q_i^2 are independent.

THEOREM 3

(i) $a^*(\mathrm{a}_{bel}) = 1,\ a^*(\phi) = 0$

(ii) If $\forall i,\ A_i \subseteq \mathrm{a}_{bel}$ the body of evidence of the A_i's are independent, then:

$$a^*\left(\underset{i \in \{1,\dots,n\}}{\cup}{}_{bel} A_i\right) \geqslant \sum_{I \subseteq \{1,\dots,n\}} (-1)^{|I|+1} a^*\left(\underset{i \in I}{\cap}{}_{bel} A_i\right).$$

(iii) If $\forall A \subseteq \mathrm{a}_{bel}$, $m^*(A) = \dfrac{a^*(A)}{a^*(h(A))} \sum_{B \subseteq A} (-1)^{A-B}\, a^*(h(B))$

where $h(B) = \cap_{bel} \{A_i / A_i = A - \{a_i\},\ a_i \in A \setminus B,\ B \neq A\}$

$h(A) = \cup_{bel} \{A_i / A_i = A - \{a_i\},\ a_i \in A\}$

then m^* is a probability assignment function on a_{bel} (in other words: $m^*: P(a_{bel}) \to [0,1]$ is such that $m^*(\phi) = 0$, $\sum_{A \subseteq a_{bel}} m^*(A) = 1$ and $\forall A \subseteq a_{bel}$, $a^*(A) = \sum_{B \subseteq A} m^*(B)$).

By using m* it is then possible to extend Dempster's rule and Dempster's conditioning on the set of belief assertions.

9.3. SEMANTIC OF a* IN CASE OF PROBABILIST OBJECTS

In case of probabilities $a_1^*(a_2)$ represents intuitively the average probability that the same instance occurs in both entities (e.g. part of Ω) described by a_1 and a_2; it will be high iff $\forall i$ $g(q_i^1, q_i^2) = \sum_v q_i^1(v)\, q_i^2(v)$ is high; more precisely, the more $q_i^1(v)$ and $q_i^2(v)$ are high together or low together and their high values are concentrated on few elements $v \in O_i$, the more $g(q_i^1, q_i^2)$ will be high. If $q_i^1(v)$ is high when $q_i^2(v)$ is low for any i then $g(q_i^1, q_i^2)$ will be low. Notice also, that if we consider that $a^*(A_1 \cap_x A_2)$ is a measure of probabilist specialisation and $a^*(A_1 \cup_{pr} A_2)$ a measure of probabilist generalisation between A_1 and A_2, then theorem 2 shows that, when $a^*(A_1) + a^*(A_2)$ is constant, the more A_1 and A_2 are specialized (e.g. $a^*(A_1 \cap_{pr} A_2)$ high) the less they are general (e.g. $a^*(A_1 \cup_{pr} A_2)$ low).

9.4. SEMANTIC OF a* IN CASE OF POSSIBILIST OBJECTS

If a_1 and a_2 are possibilist objects, $a_1^*(a_2)$ represents intuitively the "possibility" that some individual object "possible" for a_2 be "possible" for a_1; moreover, in the extreme case where a_1 and a_2 are Boolean assertions $a_1^*(a_2)$ measures the possibility that an individual object satisfies simultaneously a_1 and a_2. More precisely, if a_j is a Boolean possibilist object, it may be written $a_j = \wedge_i [y_i = q_i^j]$, where q_i^j is a characteristic mapping such that $q_i^j(v) = 1$ iff $v \in V_i^j$; so a_j may also be written as a Boolean symbolic object: $a_j = \wedge_i [y_i = V_i^j]$; it results (see section 5.3) that $a_1^*(a_2) = \text{Max}_i(\sup\{\text{Min}(q_i^1(v), q_i^2(v)) \mid v \in O_i\} = 1$ iff $\forall i$ $V_i^1 \cap V_i^2 \neq \phi$ which expresses the fact that it is **possible** for a value taken in V_i^2 to be taken in V_i^1. If a_1 is a Boolean necessitist object we have in the Boolean case: $a_1^*(a_2) = \text{Min}(\inf\{\text{Max}(q_i(v), \bar{r}_i(v)) / v \in O_i\}) = 1$ iff $\forall i$ $V_i^2 \subseteq V_i^1$ which expresses the fact that a value taken in V_i^2 is **necessarily** taken in V_i^1.

Notice also, that it is necessary and sufficient that at least for one $v \in O_i$, $q_i^1(v)$ and $q_i^2(v)$ be high together to get a high value of $g_{pos}(q_i^1, q_i^2) = \sup_v \inf(q_i^1(v), q_i^2(v))$.

EXAMPLE

We have several documents to classify, which are characterized by the frequency of some given words.

Probabilist objects: by using the frequencies, we associate to each document d_i a measure of probability q_i and a probabilist assertion a_i. It is then easy to see that $a_i^*(a_j)$ is the probability that the same word occurs for both documents d_i and d_j, it will be high if in documents d_i and d_j the frequencies are concentrated on few words and high for the same words.

Possibilist objects: some words may appear but out of context and some other, important for some documents, may not appear; so, taking into account the context, an expert associates with each word a measure of possibility; therefore each document d_i may be represented by a possibilist assertion a_i and $a_i^*(a_j)$ will be high iff at least for one word, the possibilities are simultaneously high for both documents d_i and d_j.

9.5 SEMANTICS OF a^* IN THE CASE OF BELIEF OBJECTS

The meaning of $a_1^*(a_2)$ may be interpreted as a "belief of belief" or the "conviction" of someone, denoted E_1, whose belief is represented by a_1, concerning the belief of someone else, denoted E_2, whose belief is represented by a_2.

EXAMPLE

For i = 1,2, let be $a_i = [y = q_i]$ where q_i is a belief function $O \to [0,1]$ with body of evidence (F_i, m_i) and $F_1 = F_2 = \{A, B, O\}$ with $A \cap B = f$; then we have:

$$a_1^*(a_2) = g_{bel}(q_1, q_2) = \sum_{V \in F_1} m_1(V) q_2(V)$$
$$= m_1(A) m_2(A) + m_1(B)\, m_2(B) + m_1(O). \tag{1}$$

Following a classical example given by Schafer [16], suppose that: I am expert E_1, Betty is expert E_2, A = "a tree limb fell on my car", B = "No limb fell on my car". Suppose that Betty tells me a tree limb fell on my car (therefore $m_2(A) = 1$, $m_2(B) = 0$); knowing that my subjective probability that Betty is reliable is $p = 0.9$ (so, my subjective probability that she is not reliable is $1 - p = 0.1$), I say that her testimony alone justifies a 0.9 degree of belief that a tree limb fell on my car (therefore $m_1(A) = 0.9$, $m_1(B) = 0$, $m_1(O) = 0.1$); then, it results from (1) that my belief on her belief is $a_1^*(a_2) = 1$; this is justified since my belief gives me no reason to reject the belief of Betty as $m(B) = 0$. If I have some reason to belief in B, then $m_1(B) \neq 0$ and my belief on her belief $a_1^*(a_2) = m_1(A) + m_1(O)$ becomes smaller than 1 (as $m_1(A) + m_1(B) + m_1(O) = 1$).

Notice that "my subjective probability that Betty is reliable" is equal to my belief on her belief (i.e. $a_1^*(a_2) = 0.9$) in the two following cases: i) $m_1(A) = 0.9$, $m_1(B) = 0.1$ and $m_2(A) = 1$, ii) $m_1(A) = 1$ and $m_2(A) = 0.9$, which corresponds to intuition.

More generally, we can see that the conviction of E_1 concerning the belief of E_2 will be maximum (i.e. $a_1^*(a_2) = 1$) if E_1 is totally ignorant of the evidences A and B (because in that case $m_1(A) = m_1(B) = 0$ and $m_1(O) = 1$) and if E_1 and E_2 totally believe the same evidence (because $m_1(A) = m_2(A) = 1$ or $m_1(B) = m_2(B) = 1$). If $m_1(B) = 0$ and E_1 has some ignorance of A (i.e. $m_1(O) \in]0,1[$) then, his conviction of the belief of E_2 on A (i.e. $q_2(A)$) will be greater than $q_2(A)$ (for instance if $m_1(A) = m_2(A) = \frac{1}{2}$ then $m_1(O) = \frac{1}{2}$ and the conviction of E_1 will be $a_1^*(a_2) = 0.75$). If E_1 totally believes A ($m_1(A) = 1$, $m_1(B) = m_1(O) = 0$) and E_2 totally believes B ($m_2(B) = 1$, $m_1(A) = 0$) then,the conviction of E_1 of the belief of E_2 will be 0. If E_2 is totally ignorant (i.e. $m_1(A) = m_2(B) = 0$) then the conviction of E_1 in the belief of E_2 will be low if his belief is strong (i.e. his ignorance measured by $m_1(O)$ is low).

EXAMPLE

Several sensors, in different situations, have a belief of an event A. This knowledge induces a belief of each sensor in the belief of the other sensors when they are in the same situation.

In figure 6 we give 4 situations which allow four sensors to get a belief in the belief of sensor number 5; in this figure, if we denote $a_i = [y_i = q_i]$ the belief assertion associated to sensor i and F_i the focal element of the belief function q_i, we have in situation (a) $F_1 = F_5 = \{A\}$ hence $m_1(A) = m_5(A) = 1$ therefore, it results from (1) $X_1 = a_1^*(a_5) = 1$; in situation (b), $F_2 = \{A\}$, F_5 does not contain A and so, $a_2^*(a_5) = 0$; in situation (c), $F_3 = \{A, \text{l}A\}$ and $F_5 = \{A\}$, $m_3(A) = 0.7$, $m_3(\text{l}A) = 0.3$, therefore $a_3^*(a_5) = m_3(A)\, m_5(A) + m_3(\text{l}A)\, m_5(\text{l}A) + m_3(A)\, m_5(A) + m_3(O) = 0.7$; in situation (d) $F_4 = \{A, O\}$, $F_5 = \{A\}$, $m_4(A) = 0.7$, $m_4(O) = 0.3$, $m_5(A) = 1$, therefore $a_4^*(a_5) = m_4(A)\, m_5(A) = m_4(O) = 1$. If a large majority of sensors (for instance, at least 75%) have a belief on a given sensor lower than a given threshold α, this sensor may be rejected for the recognition of A. In this example, if the threshold is $d = 1/2$ the sensor 5 is not rejected; if $\alpha = 0.8$ it is rejected; notice that if a sensor i is completely ignorant ($m_i(O) = 1$ and therefore $\forall A$, $m_i(A) = 0$) it will

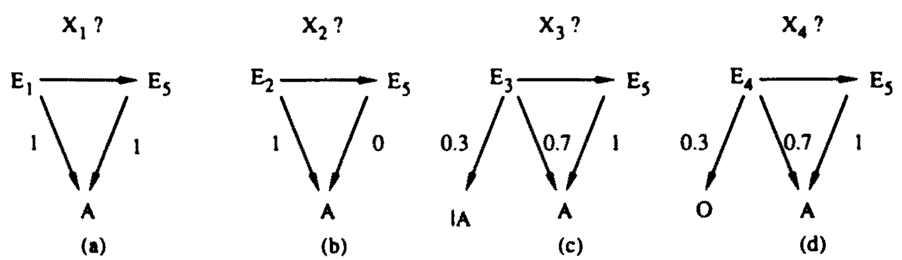

Figure 6. $X_i = a_i^*(a_5)$ is the belief of E_i in the belief of E_5, computed according to (1).

believe in any sensor whatever this sensor belief; hence, we may reject the judgement of sensors who are much too ignorant.

Instead of using a majority rule, it is also possible to use Dempster's rule (at second level) applied to the belief of belief, concerning a set of sensors, of a given sensor; in that way the sensor represented by a_5 is rejected if $\oplus_{i=1,4}\, a_i\,(a_5) < \alpha$. The belief in A, if no sensor is rejected, is measured by the classical Dempster rule (at level 1): $\oplus_{i=1,5}\, a_i\,(A)$.

There is an analogous theorem if a_1 is a plausibilist assertion and $a_1^{\bullet}\,(a_2)$ may be interpreted as the mutual "non-discordance" between what E_1 and E_2 believe. To illustrate that, going back to the preceding example we can see that if a_1 is a plausibilist object then: $a_1^{\bullet}\,(a_2) = g_{pl}\,(q_1, q_2) = \sum_{V \in F1} m_1(V)\; pl_2(V) = m_1(A)\;(m_2(A) + m_2(O)) + m_1(B)\;(m_2(B) + m_2(O)) + m_1(O)\; pl_2(O) = m_1(A)\; m_2(A) + m_1(B)\; m_2(B) + m_1(O) + m_2(O) - m_1(O)\; m_2(O)$. Hence, this corresponds to intuition as we can see (contrary to the case of conviction) that the non-discordance between what E_1 and E_2 believe remains high when E_2 is totally ignorant (i.e. $m_2(A) = m_2\,(B) = 0$) even if the belief of E_1 is strong (i.e. $m_1(O) = 0$).

Another kind of interpretation of $a_1^{\bullet}\,(a_2)$ may be obtained in terms of "fit"; if we consider the class C_i (of fruits produced by a village, for instance) described by the belief object a_i, we may say, when a_1 is a belief object, that $a_1^{\bullet}\,(a_2)$ measures how much C_2 "fits" C_1; when a_1 is a plausibilist object, we may say that $a_1^{\bullet}\,(a_2)$ measures the "non-disagreement" between C_1 and C_2. For instance, if y expresses the color and if the fruits of both villages have the same color, denoted A, (i.e. $m_1(A) = m_2\,(A) = 1, m_1(B) = m_2\,(B) = 0, m_1(O) = m_2\,(O) = 0$) then $a_1^{\bullet}\,(a_2) = 1$ measures how much C_2 "fits" C_1 and also the "non-disagreement", about color, between C_1 and C_2. If the color of the fruits of the second village is totally ignored (i.e. $m_2(A) = m_2\,(B) = 0$, $m_2\,(O) = 1$) and the color of the fruits of the first village is A (i.e. $m_1(A) = 1$, $m_1\,(O) = 0$) then, when a_1 is a belief object, we have $a_1^{\bullet}\,(a_2) = 0$ which measures how much C_2 fits C_1; when a_1 is a plausibilist object, we get $a_1^{\bullet}\,(a_2) = 1$ which measures the non-disagreement between C_1 and C_2.

10. Data analysis of symbolic objects

10.1. THE FOUR APPROACHES

Several studies have recently been carried out in this field: for histograms of symbolic objects, see De Carvalho [5]; for generating rules by decision graphs on im objects in the case of possibilist objects with typicalities as modes see Lebbe and Vignes [13]; for generating overlapping clusters by pyramids on symbolic objects see Brito and Diday [1].

More generally, four kinds of data analysis may roughly be defined depending on the input and output: (a) numerical analysis of classical data tables; (b) symbolic analysis of classical data tables (for instance obtaining a factor analysis or a clustering automatically interpreted by symbolic objects); (c) numerical analysis of symbolic objects (for instance by defining distances between objects); (d) symbolic analysis of symbolic objects, i.e. the input and output of the methods are symbolic objects.

To illustrate these four approaches, on a simple example, a similarity between symbolic objects defined as follows will be used:

Let $a_\ell = \wedge_i [y_i = q_i^\ell] \in a_x$ be the set of im assertions. We denote a_ℓ^* mapping $a_x \to [0,1]$ such that $a_\ell^*(a_k) = f_x(\{g_x(q_i^\ell, q_i^k)\}_i)$; then, we set: (1) $s(a_\ell, a_k) = \frac{1}{2}(a_\ell^*(a_k) + a_k^*)(a_\ell))/\sqrt{a_\ell^*(a_\ell)a_k^*)}$; in the case where g_x is symmetric (which happens when we have probabilist, possibilist and plausibilist assertions), s may be written: $s(a_\ell, a_k) = a_\ell^*(a_k)/\sqrt{a_\ell^*(a_\ell)a_k^*)(a_k)} = a_k^*(a_\ell)\sqrt{a_\ell^*(a_\ell a_k^*)(a_k)}$.

EXAMPLES

Let a_1, a_2 be two probabilist objects such that

$$a1 = [y = 0.7v_1, 0.3v_2], a_2 = [y = 0.3v_1, 0.7v_2];$$

we get:

$$s(a1, a2) \frac{a_1^*(a_2)}{\sqrt{a_1^*(a_1)a_2^*(a_2)}} \frac{0.7 \times 0.3 + 0.3 \times 0.7}{\sqrt{(0.7^2 + 0.3^2)(0.3^2 + 0.7^2)}} = 0.724.$$

From this example, it results that probabilist objects do not satisfy condition (c) given in section 8.2, since if we define $a = [y = 1v_1, Ov_2] = [y = v_1]$ we get $a_1 \cup_{pr} a_2 (a) = 0.79$ and $a_1 \cup_{pr} a_1 (a) = 0.91$; hence, $a_1 \cup_{pr} a_2$ may not be considered more general than $a_1 \cup_{pr} a_1$, even if the pair (a_1, a_2) may be considered more similar than the pair (a_1, a_1), since $s(a_1, a_1) = 1$ and $s(a_1, a_2) = 0.724$.

Let a_1, a_2 be two possibilist objects such that $a_1 = [y = 1\ v_1, xv_2]$ and $a_2 = [y = xv_1, 1\ v_2]$. Then,

$$s(a_1, a_2) = \frac{(\text{Max}(\min(1, x), \min(x, 1))}{\sqrt{(\text{Max}(\min(1, 1), \min(x, x))}} = x;$$

hence the lower x, the more a_1 and a_2 are dissimilar. Hence, $a_1 \cup_p a_2 = [y = 1v_1, 1v_2]$ is the full object since $\forall a$, $a_1 \cup_p a_2^*(a) = 1$ and therefore, contrarily to the probabilist case, in this example the possibilist case satisfies condition (c) in section 8.2).

We illustrate these four approaches by applying three data analysis methods: principal components, hierarchical and pyramidal clustering.

Let T be the following data table where the set of individual objects is $\Omega = \{\omega_1, \ldots, \omega_5\}$ which are five companies described by two variables, y_1: the employment rate and y_2: the profit. This table is represented in figure 7.

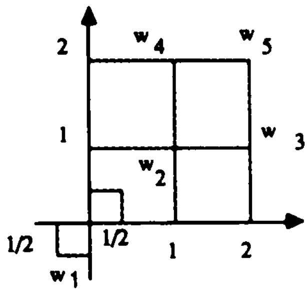

Figure 7. Graphical representation of table T.

10.2. NUMERICAL ANALYSIS OF CLASSICAL DATA TABLE

Principal component analysis of table T: From the covariance matrix $V = \begin{pmatrix} 0.9 & 0.7 \\ 0.7 & 0.9 \end{pmatrix}$ we deduce the eigenvalues: $\lambda_1 = 1.6$ and $\lambda_2 = 0.2$ and the eigenvectors $u_1^T = 1/\sqrt{2}\,(1,1)$, $u_2^T = 1/\sqrt{2}\,(1,-1)$. Finally we get the principal component representation given in figure 8, where the projection of w_j on the axis i is given by $F_i(\omega_j) = u_i^T.x_j$, where $x_j^T = (y_1(\omega_1) - Y_1, y_2(\omega_2) - Y_2)$ and $Y_i = 1$, is the mean of y_i; for instance, $F_1(\omega_1) = 1/\sqrt{2}\,(1\ \ 1)\begin{pmatrix} -3/2 \\ -3/2 \end{pmatrix}$.

Hierarchical and pyramidal clustering of table T:

We make the classical "complete link hierarchy" based on the city-block distance defined by

$$d(\omega_\ell, \omega_k) = \sum_{j=1} |y_j(\omega_\ell) - y_j(\omega_k)|.$$

The algorithm is the following: starting from 5 classes $C_i = \{\omega_i\}$ where $\omega_i \in \Omega$, we merge at each step the two classes with smallest $\delta(C_i, C_j)$:

$\delta\,(C_i, C_j) = \mathrm{Max}\,\{d(\omega_i, \omega_j)/\omega_i \in C_i, \omega_j \in C_j\}$. When two classes are merged their elements are suppressed from the set to be classified and the process continues until only one class remains.

Table T

	ω_1	ω_2	ω_3	ω_4	ω_5
y_1	$-1/2$	$1/2$	2	1	2
y_2	$-1/2$	$1/2$	1	2	2

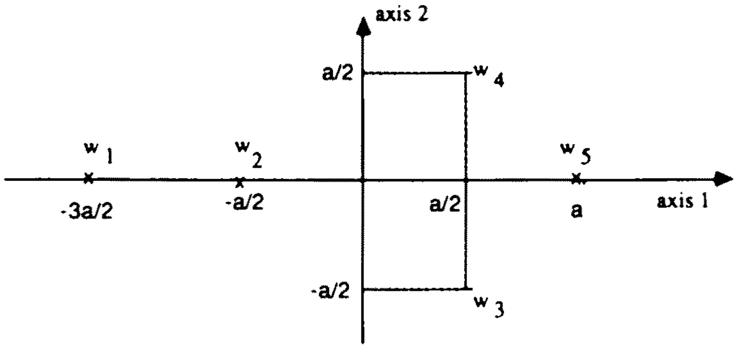

Figure 8. Principal component analysis of table T whith a = 2.

To obtain a pyramid, we may use a similar algorithm where classes may be merged twice (instead of only once in the case of hierarchies) if they respect a common order (for more details see for instance Brito and Diday [1]).

By using these algorithms we get the hierarchy and the pyramid given in figure 9.

Remark: if we associate a dissimilarity σ induced by the hierarchy and the pyramid by setting: $\sigma(\omega_i, \omega_j)$ = {height of the lower level which contains ω_i and ω_j}, then, it is easy to see that σ is closer to the initial distance d in the case of the pyramid than in the case of the hierarchy; more precisely, $d - \sigma = \Sigma\ d(\omega_i, \omega_j) - \sigma(\omega_i, \omega_j)$ is equal to 3 for the pyramid and to 11 for the hierarchy.

10.3. SYMBOLIC ANALYSIS OF A CLASSICAL DATA TABLE

The correlations between $(\omega_1, \ldots, \omega_5)$ and the first axis of the principal component analysis are respectively (−1, −0.707, 0.707, 0.707, 1); if we associate to each side of the first axis the objects whose correlation is higher than 0.707 or lower than −0.707, we obtain two classes of objects; the first class, $C_1 = \{\omega_1, \omega_2\}$, explains the

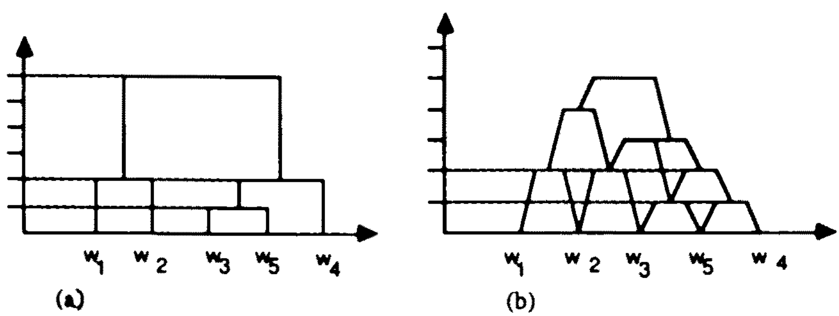

Figure 9. (a) The hierarchy of table T. (b) The pyramid of table T.

left side of the axis and the second one $C_2 = \{w_3, w_4, w_5\}$ explains the right side. By using these classes, we get two kinds of symbolic interpretation of the first axis; by using assertions, we may say that the left side is explained by: $a_1 = [y_1 = -1/2, 1/2] \wedge [y_2 = -1/2, 1/2]$; the right side is explained by $a_2 = [y_1 = 1,2] \wedge [y_2 = 1,2]$. If the input provides a taxonomy saying that the rate of employment and the profit are low when they are lower than 1/2 and high when they are higher than 1, we may use the assertions a_1 and a_2 to get the following explanation of the first axis: it is explained by two opposite assertions which characterize two classes of companies:

$$a_1 = [\text{Rate of employment} = \text{low}] \wedge [\text{Profit} = \text{low}]$$
$$a_2 = [\text{Rate of employment} = \text{high}] \wedge [\text{Profit} = \text{high}]$$

Of course, in real examples things become much more complicated; for instance, to get more accuracy when the two classes contain numerous objects, each side of the axis may be explained by a disjunction of assertions obtained by a symbolic interpretation of a clustering done on each class. We may also enrich the interpretation by adding certain properties; for instance, we may add to a_1 the following rules: [if $y_1 = -1/2$ then $y_2 = -1/2$] $\wedge$ [if $y_1 = 1/2$ then $y_2 = 1/2$] and to a_2 the rule [if $y_1 = 1$ then $y_2 = 2$].

We may also give an interpretation of the first axis by a horde object h: $h = a_1(u_1) \wedge a_2(u_2) = [\text{Rate of employment } (u_1) = \text{low}] \wedge [\text{Profit } (u_1) = \text{low}] \wedge [\text{Rate of employment } (u_2) = \text{high}] \wedge [\text{Profit } (u_2) = \text{high}]$ whose extension is composed of couples of companies (ω_i, ω_j) the first element of the couple, ω_i, being of low rate of employment and profit and the second one, ω_j, of high rate of employment and profit. If an external variable gives the age of the companies the horde object h may become: $h = a_1(u_1) \wedge a_2(u_2) \wedge [\text{age}(u_1) < \text{age}(u_2)]$.

A symbolic analysis of a classical data table may also be obtained by an automatic interpretation of a clustering by symbolic objects: for instance, it is possible to associate to each level of the hierarchy a complete symbolic object (see section 8.2); more precisely, if we denote $h_1 = \{\omega_3, \omega_5\}$ then we may associate to h_1, the assertion $a_1 = [y_1 = 2] \wedge [y_2 = 1,2]$; a_1 is complete, because: (i) it is defined by the intension of h_1, in other words, by the conjunction of all the events $e_i = [y_i = V_i]$ whose extension contains h_1 and (ii) its extension is h_1; in the same way $h_2 = \{\omega_1, \omega_2\}$, $h_3 = \{\omega_3, \omega_4, \omega_5\}$ and $h_4 = \Omega$ may be respectively associated to the complete assertions $a_2 = [y_1 = -1/2, 1/2] \wedge [y_2 = -1/2, 1/2]$, $a_3 = [y_1 = 1,2] \wedge [y_2 = 1,2]$, $a_4 = [y_1 = 0_1] \wedge [y_2 = 0_2]$ where 0_1 and 0_2 are the sets of all the values taken by y_1 and y_2 in the data table T. Using the fact that each level is represented by a complete assertion we deduce from any level $h_\ell = h_i \cup h_k$ the rule $a_\ell \rightarrow a_i \vee a_k$. Hence, from the hierarchy we obtain the two following rules: R_1: $a_4 \rightarrow a_2 \vee a_3$ and R_2: $a_3 \rightarrow a_1 \vee \omega_4^s$ where $\omega_4^s = [y_1 = 1] \wedge [y_2 = 2]$ is the symbolic object associated to ω_4. All the bottom-up rules, such as $a_1 \rightarrow a_3$, are true because the a_i and b_i are complete objects. Finally we have induced from the hierarchy given in a) a graph (see figure 10(a)) whose nodes are assertions and rules are expressed between them by

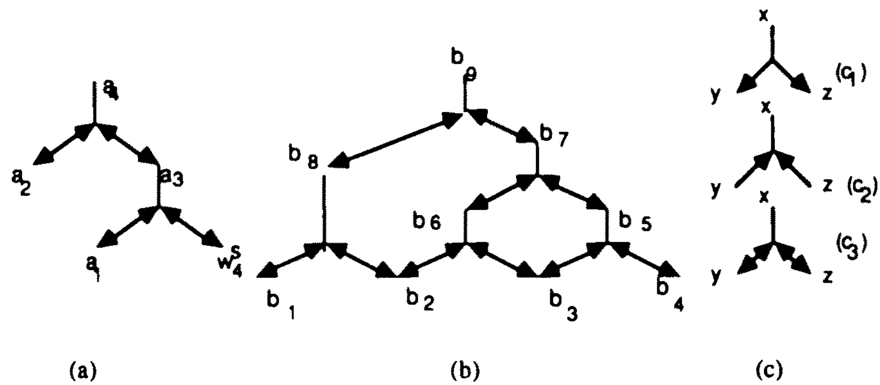

Figure 10. Induced graph of rules between assertions (a) from the hierarchy, (b) from the pyramid, where double headed arcs are explained by (c).

directions. In figure 10(c), (c_1) expresses the rule r_1: $x \rightarrow y \vee z$; (c_2) expresses the rule r_2 $(y \rightarrow x) \wedge (z \rightarrow x)$ and (c_3) expresses the rule $r_1 \wedge r_2$

The same kind of symbolic interpretation may be obtained by starting from the pyramid given in figure 10; hence, we obtain the graph given in figure 10(b); in this way, we obtain more assertions and more rules between them. If we denote $h_1 = \{\omega_1, \omega_2\}$, $h_2 = \{\omega_2, \omega_3\}$, $h_3 = \{\omega_3, \omega_5\}$, $h_4 = \{\omega_4, \omega_5\}$ $h_5 = \{\omega_3, \omega_4, \omega_5\}$, $h_6 = \{\omega_2, \omega_3, \omega_5\}$, $h_7 = \Omega \omega_1$, $h_8 = \{\omega_1, \omega_2, \omega_3\}$, $h_9 = \Omega$, the associated complete assertions are:

$b_1 = [y_1 = -1/2, 1/2] \wedge [y_2 = -1/2, 1/2]$, $b_2 = [y_1 = \frac{1}{2}, 2] \wedge [y_2 = \frac{1}{2}, 1]$,
$b_3 = [y_1 = 2] \wedge [y_2 = 1,2]$, $b_4 = [y_1 = 1,2] \wedge [y_2 = 2]$, $b_5 = [y_1 = 1,2] \wedge [y_2 = 1,2]$,
$b_6 = [y_1 = 1/2,2] \wedge [y_2 = 1/2,1,2]$, $b_7 = [y_1 = 1/2,1,2] \wedge [y_2 = 1/2,1,2]$,
$b_8 = [y_1 = -1/2, 1/2,2] \wedge [y_2 = -1/2, 1/2, 1]$, $b_9 = [y_1 = 0_1] \wedge [y_2 = 0_2]$.

Hence we can induce the following rules:

r_1: $b_9 \rightarrow b_8 \vee b_7$ r_2: $b_7 \rightarrow b_6 \vee b_5$ r_3: $b_8 \rightarrow b_1 \vee b_2$
r_4: $b_6 \rightarrow b_2 \vee b_3$ r_5: $b_5 \rightarrow b_3 \vee b_4$.

We have $b_1 = a_2$, $b_3 = a_1$, $b_5 = a_3$ and $b_9 = a_4$; hence, it is possible to deduce from the rules r_i given by the pyramid, the rules given by the hierarchy; to do so, we need to use the following property: if r: $b_i \rightarrow b_j \vee b_k \vee b_\ell$ and $\text{Ext}(b_j/\Omega) = \text{Ext}(b_\ell/\Omega)$, then r may be simplified to $b_i \rightarrow b_j \vee b_\ell$. Hence, for instance, from r_1, r_2 and r_3 we get $b_9 \rightarrow b_1 \vee (b_2 \vee b_6) \vee b_5$ and then $b_9 \rightarrow b_1 \vee b_5$ which is R_1: $a_4 \rightarrow a_2 \vee a_3$, obtained from the hierarchy (see figure 10a).

10.4. NUMERICAL ANALYSIS OF SYMBOLIC OBJECTS

The given set of symbolic objects is supposed to be the set of the five first symbolic objects defined by the pyramid: $\{b_1, b_2, b_3, b_4, b_5\} = B$. A simple way to make a bridge with classical data analysis methods is to compute a measure of similarity between the objects of B; having this measure it is then possible to use multidimensional scaling, clustering etc. To do so, we may compute the similarity s which has been defined by (1); as B is a set of symbolic objects, we have to use the mappings f_b and g_b defined in section 5.3. We have, for instance, $s_b(b_1,b_2) = b_1^*(b_2)/\sqrt{b_1^*(b_1)b_2^*(b_2)}$ with $b_1 = [y_1 = q_1^1] \wedge_b [y_2 = q_2^1]$ where q_1^1 and q_2^1 are characteristic mappings such that: $q_1^1(-1/2) = q_1^1(1/2) = 1$ and $q_2^1(-1/2) = q_2^1(1/2) = 1$ and $q_1^1(v) = q_2^1(v) = 0$ elsewhere. We have $b_2 = [y_1 = q_1^2] \wedge_b [y_2 = q_2^2]$ and $q_1^2(v) = 1$ if $v \in \{1/2\ 1/2\}$, $q_1^2(v) = 0$ elsewhere, $q_2^2(v) = 1$ if $v \in \{1/2, 1\}$ and $q_2^2(v) = 0$ elsewhere. As we have (see section 7), $b^*_1(b_2) = f_b(\{g_b(q_i^1, q_i^2)\}_i) =$

	b_1	b_2	b_3	b_4	b_5
b_1	1	1/2	0	0	0
b_2		1	$\frac{\sqrt{2}}{2}$	0	1/2
b_3			1	1	$\frac{\sqrt{2}}{2}$
b_4				1	$\frac{\sqrt{2}}{2}$
b_5					1

(a)

	b_1	b_2	b_3	b_4	b_5
b_1	0	0.5	1	1	1
b_2		0	0.3	1	0.5
b_3			0	0	0.3
b_4				0	0.3
b_5					0

(b)

	b_1	b_2	b_3	b_4	b_5
b_1	0	1.2	1.7	1.7	1.7
b_2		0	1	1.7	1.2
b_3			0	0.7	1
b_4				0	1
b_5					0

(c)

Figure 11.

Min $(\langle q_1^1,q_1^2\rangle, \langle q_2^1,q_2^2\rangle) = \mathrm{Min}(\Sigma\{q_1^1(v)q_1^2(v)/v \in O_1\}, \Sigma\{q_2^1(v)q_2^2(v)/v \in O_2\}) = \mathrm{Min}\,(q_1^1(1/2)\,q_1^2(1/2), q_2^1(1/2)q_2^2(1/2)) = \mathrm{Min}(1,1) = 1$. We have $b_1^*(b_1) = \mathrm{Min}\,(\langle q_1^1,q_1^1\rangle, \langle q_2^1,q_2^1\rangle) = \mathrm{Min}\,(2,2) = 2$ and also $b_2^*\,(b_2) = 2$; hence, $s_b(b_1,b_2) = 1/\sqrt{2\times 2} = 1/2$.

By computing in the same way all the similarities $s_b\,(b^{(i)}, b^{(j)})$ we finally get the symmetric table of similarities given in figure 11 (a).

The similarity s_b is transformed into a dissimilarity $d = 1 - s_b$ given in figure 11(b). If we choose $c = \mathrm{Max}\ d(b_i,b_j) - M$ where M is the sum of the two couples (b_i,b_j) of smallest dissimilarity $d(b_i,b_j)$, then $c \geqslant \mathrm{Max}\ (d(b_i,b_j) - d(b_i,b_k) - d(b_k,b_j))$ and D such that $D\,(b_i,b_j) = d(b_i,b_j)$, $D\,(b_i,b_i) = 0$, is a distance, because $\forall i,j,k\ d(b_i,b_j) + c \leqslant d(b_i,b_k) + c + d(b_k,b_j) + c$. It is easy to see that $M = 0 + 0.3$ and $c = 1 - 0.3$; it is then possible to change d into a distance D such that $D(b_i,b_j) = d(b_i,b_j) + 0.7$, which is given in figure 11(c). It is then possible to apply many existing methods of classical data analysis by using s, d or D as input.

10.5 SYMBOLIC ANALYSIS OF SYMBOLIC OBJECTS

As input we have the following set of probabilist objects:

$B = \{b_1,\ldots,b_5\}$ such that $b_i = [y_1 = q_1^i] \wedge_{pr} [y_2 = q_2^i]$ where q_i^j is a measure of probability from $P\,(O_i) \rightarrow [0,1]$ where $O_i = \{-1/2, 1/2, 1,2\}$ and $P\,(O_i)$ is the power set of O_i. If we set $b_i = \wedge[y_i = (q_i^1\,(v_1))\,v_1, (q_i^1\,(v_2))\,v_2,\ldots]$, then the b_i are defined as follows, where the value v_l associated to $q_i^1\,(v_l) = 0$ does not appear:

$$b_1 = \left[y_1 = \left(\tfrac{1}{2}\right) - \tfrac{1}{2}, \left(\tfrac{1}{2}\right) \tfrac{1}{2}\right] \wedge_{pr} \left[y_2 = \left(\tfrac{1}{2}\right) - \tfrac{1}{2}, \left(\tfrac{1}{2}\right) \tfrac{1}{2}\right]$$

$$b_2 = \left[y_1 = \left(\tfrac{1}{2}\right) - \tfrac{1}{2}, \left(\tfrac{1}{2}\right) 2\right] \wedge_{pr} \left[y_2 = \left(\tfrac{1}{2}\right) \tfrac{1}{2}, \left(\tfrac{1}{2}\right) 1\right]$$

$$b_3 = [y_1 = (1)2] \wedge_{pr} \left[y_2 = \left(\tfrac{1}{2}\right) 1, \left(\tfrac{1}{2}\right) 2\right]$$

$$b_4 = \left[y_1 = \left(\tfrac{1}{2}\right) 1, \left(\tfrac{1}{2}\right) 2\right] \wedge_{pr} [y_2 = (1)2]$$

$$b_5 = \left[y_1 = \left(\tfrac{1}{2}\right) 1, \left(\tfrac{1}{2}\right) 2\right] \wedge_{pr} \left[y_2 = \left(\tfrac{1}{2}\right) 1, \left(\tfrac{1}{2}\right) 2\right].$$

To treat this set of probabilist objects, we may compute, at first, the similarity $s_{pr}\,(b_i,b_j) = b_i^*(b_j)/\sqrt{b_i^*(b_i)b_j^*(b_j)}$ and then, to use for instance, principal component analysis or clustering methods interpreted by symbolic objects as has already been done in b).

For instance, for the couple (b_1,b_2), $b_1^*\,(b_2) = f_{pr}\,(\{g_{pr}\,(q_i^1,q_i^2)\}_i)$ is computed as follows:

$b_1^*\,(b_2) = \mathrm{Mean}\,(\langle q_1^1, q_1^2\rangle, \langle q_2^1, q_2^2\rangle)$; therefore:
$b_1^*\,(b_2) = \mathrm{Mean}\,(S\,\{q_1^1(v).\ q_1^2(v)\ v \in O_1\}, S\,\{q_2^1(v).\ q_2^2(v)\ v \in O_2\})$.
Hence $b_1^*\,(b_2) = \mathrm{Mean}\,(\ 1/2 \times 0 + 1/2 \times 1/2 + 0 \times 0 + 0 \times 1/2,\ 1/2 \times 0 + 1/2 \times$

$1/2 + 0 \times 1/2 + 0 \times 0) = \text{Mean}\,(1/4, 1/4) = (1/4 + 1/4)\,1/2 = 1/4.$
$b_1^*(b_1) = \text{Mean}\,(1/2, 1/2) = 1/2;\ b_2^*(b_2) = \text{Mean}\,(1/2, 1/2) = 1/2.$

Finally, setting $\alpha = \sqrt{3/2}$ we obtain the following similarities:

$$\{s_{pr}(b_i, b_j)\} = \begin{matrix} 1 & 1/2 & 0 & 0 & 0 \\ & 1 & \alpha/2 & \alpha/6 & 1/2 \\ & & 1 & 2/3 & 2\alpha/3 \\ & & & 1 & 2\alpha/3 \\ & & & & 1 \end{matrix} = \begin{matrix} 1 & 0.5 & 0 & 0 & 0 \\ & 1 & 0.6 & 0.2 & 0.5 \\ & & 1 & 0.7 & 0.8 \\ & & & 1 & 0.8 \\ & & & & 1 \end{matrix}$$

For treating B, another way is to obtain directly from B, clusters of symbolic objects represented by an "inheritance" hierarchy, where each node is expressed by a complete probabilist assertion a_{jk}, or an approximation of it such that if $a_{jk} = a_j \cup_{x,\alpha} a_k$ then $a_{jk} \geqslant_\alpha \text{Max}\,(a_j, a_k)$ where $\cup_{pr,\alpha}$ and $\geqslant_\alpha$ have been defined in section 9.1. To do so, we may use the following algorithm on a set of symbolic objects A:

First step: $a_{jk} = a_j \cup_{x,\alpha} a_k$ is computed $\forall a_j, a_k \in A$.
Second step: the a_{jk} of smaller extension constitute the first levels of the hierarchy, their height is the cardinality of their extension.
Third step: the retained a_{jk} at step 2 are added to A and a_j, a_k are suppressed from A; then, we go back to the first step until the cardinality of A becomes equal to 1.

In practice, how can we compute $a_{jk} = a_j \cup_{pr,\alpha} a_k$? By definition a_{jk} is the conjunction of the elementary events $a_{jk}^i = [y_i = q_i]$ such that $\text{Ext}(a_{jk}^i\ \Omega, \alpha)$ contains $\Omega_1 = \text{Ext}(a_j, \Omega, \alpha) \cup \text{Ext}(a_k\ \Omega, \alpha)$. Hence, for any $\omega \in \Omega_1$ such that $\omega^s = \wedge_i [y_i = r_i]$ we have $a_{jk}(\omega) \geqslant \alpha$; this condition is satisfied if we have $\forall i$, $g(q_i, r_i) \geqslant \alpha$ because $a_{jk}(\omega) = f(\{g(q_i, r_i)\}_i)$ and, by definition of f, it is the mean of numbers larger than α; hence, if we denote $x_j^i = q_i\ (v_j)$, we have the inequality: $g(q_i, r_i) = \Sigma\{x_j^i . r_i(v_j)\ v_j \in O_j\}\alpha$; hence, we have to solve a system of $\text{card}(\Omega_1)$ inequalities where the unknowns are the x_j^i. If this system has several solutions, for each i we denote them $[y_i = q_i^\ell]$; hence, we obtain $a_{jk} = \wedge_{ipr}\ (\wedge_{\ell pr} [y_i = q_i^\ell])$; by choosing $h_{pr} = \text{Min}$ (see section 3.1) the extension of a_{jk} at level α is $\Omega_2 = \{\omega\ a_{jk}(\omega) = f(\text{Min}\{g(q_i^\ell), r_i)\}\}_i \geqslant \alpha\}$.

To obtain the inheritance hierarchy on B given by the algorithm, the first step consists in computing the $a_{jk} = b_j \cup_{pr,\alpha} b_k$ whose extension is of minimum cardinality, we choose $\alpha = \frac{1}{2}$ and to compute for instance $a_{12} = b_1 \cup_{pr,\alpha} b_2$ we do the following: first we set $a_{jk} = a_{jk}^1 \wedge a_{jk}^2$ where $a_{jk}^\ell = [y_\ell = q_\ell]$ is such that $a_{12}^\ell(b_1) \geqslant 1/2$ and $a_{12}^\ell(b_2) \geqslant 1/2$. Then, for $x_j^\ell = q_\ell(v_j)$ where $\{v_1, \ldots, v_4\} = O_1 = O_2 = \{1/2, 1/2, 1, 2\}$, we have to solve the following inequalities, where the x_{jk}^ℓ are the unknowns, with the constraint

$$\sum_j q_\ell(v_j) = \sum_j x_j^\ell = 1:$$

$a_{12}^\ell(b_1) = g_{pr}(q_\ell, r_1^\ell) = \Sigma\{q_\ell(v_i)\ \ r_{12}^\ell)(v_i)/v_i \in O_\ell\}$; hence, we obtain: $a_{12}^1(b_1) = 1/2x_1^1 + 1/2x_2^1 \geqslant 1/2$; $a_{12}^1(b_2) = 1/2x_2^1 + 1/2x_4^1 \geqslant 1/2$ from which we deduce that $x_1^1 + x_2^1 = 1$, therefore (as $\Sigma\{x_i^1 i = 1, 4\} = 1$) we get $x_4^1 = 0$ and $x_2^1 = 1$, $x_i^1 = 0$ if $i \neq 2$. $a_{12}^2(b_1) = 1/2\,x_1^2 + 1/2x_2^2 \geqslant 1/2$; $a_{12}^1(b_2) = 1/2x_2^2 + 1/2\,x_3^2 \geqslant 1/2$, from which it results that $x_2^2 = 1$ and $x_i^2 = 0$ if $i \neq 2$. Finally we obtain:

$a_{12} = a_{12}^1 \wedge_{pr} a_{12}^2 = [y_1 = (1)1/2] \wedge_{pr} [y_2 = (1)1/2]$ (which is equivalent to the Boolean object $[y_1 = 1/2] \wedge_b [y_2 = 1/2]$).

Similarly, we get: $a_{13}^1(b_1) = 1/2x_1^1 + 1/2x_2^1 \geqslant 1/2$ and $a_{13}^1(b_3) = x_4^1 \geqslant 1/2$. This is a contradiction because the first equation implies $x_4^1 = 0$. Hence, the only symbolic object whose extension contains b_1 and b_3 is the full object Ω^s whose extension is Ω; $\Omega^s = \wedge_i[y_i = q_i]$ is defined in the case of probabilist objects by functions q_i: $P(O_i) \to \{1\}$ (which are not, of course, probabilities!), then it is easy to see that $\Omega^s(\omega) = 1$ $\forall\omega \in \Omega$. Similarly we get: $a_{14} = a_{15} = a_{24} = \Omega^s$; $a_{23} = [y_1 = (1)2] \wedge_{pr} [y_2 = (1)1]$; $a_{25} = a_{23}$; $a_{34} = [y_1 = (1)2] \wedge_{pr} [y_2 = 1(2)]$; a_{35} is computed as follows: $a_{35}^1(b_3) = x_4^1 \geqslant 1/2$ and $a_{35}^1(b_5) = 1/2\,x_3^1 + 1/2\,x_4^1 \geqslant 1/2$ implies $x_4^1 = 1$ and $x_i^1 = 0$ if $i \neq 4$; $a_{35}^2(b_3) = 1/2x_3^2 + 1/2\,x_4^2 \geqslant 1/2$ and $a_{35}^2(b_5) = 1/2\,x_3^2 + 1/2\,x_4^2 \geqslant 1/2$; we have three solutions (i) $x_3^2 = x_4^2 = 1/2$; (ii) $x_3^2 = 1$, $x_i^2 = 0$ for $i \neq 3$; (iii) $x_4^2 = 1$, $x_i^2 = 0$ for $i \neq 4$; therefore:

$$a_{35} = [y_1 = (1)2] \wedge_{pr} [y_2 = (1/2)1, (1/2)2] \wedge_{pr} [y_2 = (1)1] \wedge_{pr} [y_2 = (1)2].$$

In a similar way we get:

$$a_{45} = [y_1 = (1/2)1, (1/2)2] \wedge_{pr} [y_1 = (1)1] \wedge_{pr} [y_1 = (1)2] \wedge_{pr} [y_2 = (1)2].$$

In the following table we give in the cell of the ith row and jth column the extension of

$\text{Ext}\left(a_{ij}/B, \frac{1}{2}\right) =$

	1	2	3	4	5
1		b_1b_2	Ω	Ω	Ω
2			$b_2b_3b_5$	Ω	$b_2b_3b_5$
3				$b_3b_4b_5$	b_3b_5
4					b_4b_5
5					

Using this table it is easy to construct the inheritance hierarchy, by merging at each step the couple of least extension. Hence, the first couples are (b_1,b_2),(b_3,b_5),(b_4,b_5);

Table 2

Level	Representation	Extension
1	$a_{12} = [y_1 = (1)1/2] \wedge_{pr} [y_2 = (1)1/2]$	$\{b_1, b_2\}$
2	$a_{35} = [y_1 = (1)2] \wedge_{pr} [y_2 = ((1/2)1, (1/2)\ 2)] \wedge_{pr} [y_2 = (1)1] \wedge_{pr} [y_2 = (1)2]$	$\{b_3, b_4\}$
3	$a_{345} = [y_1 = (1)2] \wedge_{pr} [y_2 = (1)2]$	$\{b_3, b_4, b_5\}$
4	$a_{12345} = \sum_{i=1,2} [y_i = (1)\ -1/2,\ (1)1/2,\ (1)1,\ (1)2] = \Omega^s$	B

to get a hierarchy it is not possible to retain simultaneously (b_3, b_5) and (b_4, b_5) therefore if there are no external constraints on the clusters (for instance, constraints of geographical proximity) we have to choose one of them randomly; if we retain, for instance, (b_3, b_5) the first couples to be merged are finally (b_1, b_2) and (b_3, b_5); therefore, we obtain the two first levels of the hierarchy characterized by $a_{12} = b_1 \cup_{pr,1/2} b_2$ and $a_{35} = b_3 \cup_{pr,1/2} b_2$. Hence, it remains for b_4 to be merged with (b_1, b_2) or (b_3, b_5). It is then easy to see that $a^2_{124}(b_1) = 1/2\, x_1^2 + 1/2\, x_2^2 \geqslant 1/2$ and $a^2_{124}(b_4) = 1/2\, x_2^4 \geqslant 1/2$ which give no solution such that $\Sigma_{i=1,4} x_i^2 = 1$; therefore, $a_{124} = \Omega^s$ whose extension is B. We have already seen that $\text{Ext}(a_{34}/B, 1/2) = \{b_3, b_4, b_5\}$, therefore $a_{345} = a_{34}$; hence, the next couple to be merged will be $(b_4, (b_3, b_5))$ which gives a third level represented by $a_{345} = a_{34}$; the last level merges (b_1, b_2) with (b_3, b_4, b_5) and is represented by the full object Ω^s.

To summarize, we have finally obtained four levels whose representation and extension are given in table 2.

Using the fact that the height of each level is the cardinality of the extension of its associated probabilistic assertion, it is then easy to build the inheritance hierarchy associated to the set B of probabilist objects, represented in figure 12.

Notice that the same algorithm may be used with the probabilist, possibilist

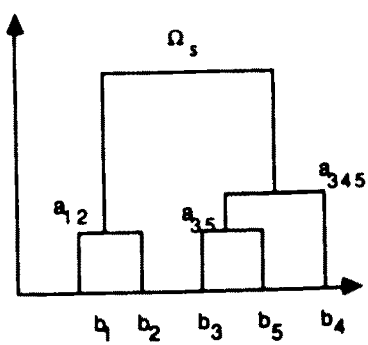

Figure 12. Inheritance hierarchy on probabilist objects

and belief union defined respectively in sections 5, 6, 7 instead of the symbolic union defined in section 9 which has been used here. The advantage of the symbolic union (see section 8.1) is that it defines the supremum of the lattice associated to the symbolic order. The advantage of the probabilist, possibilist and belief union is that they allow the use of theorems 1, 2, 3; in this case if the height of a level defined by $a_3 = a_1 \cup_x a_2$ is given by $a_3^*(a_1 \cup_x a_2)$, we get in case of probabilist objects $a_3^*(a_1 \cup_{pr} a_2) = a_3^*(a_1) + a_3^*(a_2) - a_3^*(a_1 \cap_{pr} a_2) \geqslant a_1^*(a_1) + a_2^*(a_2) - a_3^*(a_1 \cap_{pr} a_2)$; it results that the obtained hierarchy will have no inversions (as it may be shown that $a_3^*(a_3) \geqslant a_1^*(a_1)$ and $a_3^*(a_3) \geqslant a_2^*(a_2)$ and the more a_1 and a_2 are "independent" (i.e. $a_1 \cap_{pr} a_2$ close to 0) the more the height of a_3 will tend to be high.

We say that we have a rule between two probabilistic assertions a_1 and a_2 at level (α_1, α_2) denoted R: $a_1 \xrightarrow{(\alpha_1,\alpha_2)} a_2$ when $\mathrm{Ext}(a_1/B, \alpha_1) \subseteq \mathrm{Ext}(a_2/B, \alpha_2)$; in other words, the rule R is true if, when b is in the extension of a_1 at level α_1, then, it is in the extension of a_2 at level α_2 ; when $\alpha_1 = \alpha_2 = \alpha$ this rule is denoted $a_1 \xrightarrow{\alpha} a_2$. By using this notation, it is easy to induce from the inheritance hierarchy of figure 6, by going bottom-up, the rule: $a_{35} \xrightarrow{1/2} a_{345}$; it is also possible to induce top-down the following rule: $\Omega^s \xrightarrow{(1,1/2)} a_{12} \vee a_{345}$ which means that if b is in the extension of Ω^s at level 1, it is also in the extension of a_{12} or a_{345}, at the level 1/2; in the same way we get also $a_{345} \xrightarrow{1/2} b_4 \vee a_{35}$.

11. Conclusion

Considering a data base (Ω, Δ') where any individual object $\omega \in \Omega$ is described by $\delta \in \Delta' \subseteq \Delta$, we have built a knowledge base (W,A) where any symbolic object $a \in A \subseteq \mathfrak{a}_x$ describes a subset $W' \in W$ of Ω; these symbolic objects may be obtained from the meta-data given by a data analysis of (Ω, Δ') (for instance, from a symbolic interpretation of the axis of a factorial analysis or from a symbolic description of clusters obtained by a classical clustering technique); the set A of symbolic objects may also be obtain directly from the knowledge of an expert (for instance, from his description of a scenario of accident or of a species of mushrooms).

Having (W,A) we have given tools in order to be able to extract meta-knowledge from A, by extending data analysis methods on symbolic objects. These tools depend on the backgrownd knowledge of the domain of application; we have defined several local theories by giving axioms and operators coherent with Boolean, probabilities, possibilities and belief information. Many kinds of developments are needed in the future, by improving the basic choices given in this paper; more precisely, operators of union $\cup_x$ and intersection $\cap_x$ may be redefined, the mappings f_x and g_x may be changed depending on the kind of the semantic inherent to any curent application; for instance, in the case of probabilist objects instead of using the mean to compute $\wedge_{pr}$ by f_{pr} we may use the product and instead of using the scalar product to compute the fit between two probability distributions we may use many other classical similarities such as, for instance, Kullback, Kolmogorov etc. The advantage of the choices that we have made is that they are coherent on symbolic objects with the axioms defined by each theory on individual objects; for instance, theorem 2 shows that in case of probabilities $\mathfrak{a}^{*}$ defined on a_{pr} (the set of probabilist objects) satisfies properties which are analogous to the classical axioms of Kolmogorov. In order to obtain the same coherence with other choices of OP_{pr}, f_{pr} and g_{pr} we have to solve functional equations (given by the Kolmogorov axioms) and so, many research questions remain open, in this direction.

In practice it may happen that several semantics are used simultaneously (intensity together with probability, and possibility, for instance), an important challenge is then, to find the best way to define symbolic objects concerned different semantics; more precisely, how to define $\wedge_{xy}$ (eg. f_{xy}) in $e_x \wedge_{xy} e_y$ where e_x and e_y are two events representing two different kinds of semantic (for instance when e_x is a probabilist and e_y is a possibilist event).

If $A = \cup_x \{q/q \in A_x \subseteq \mathfrak{a}_x\}$ is called x-set, then in the case of possibilities ($x = \text{pos}$) A_{pos} is a fuzzy set in the original sense given by Zadeh [19]; in this case $\cup_{pos}$ is stable but not $\cap_{pos}$: in case of probabilist objects $\cup_{pr}$ and $\cap_{pr}$ are not stable. The advantage of belief objects is that $\cup_{bel}$ and $\cap_{bel}$ are both stable. In defining new kinds of operators we will have to try to satisfy stability. Several computer programs of symbolic data analysis have already been implemented independently, see for instance this issue: histograms of symbolic objects (De Carvalho [5]), symbolic pyramidal clustering (Brito) on decision tree on symbolic objects (Jacq), extracting rules from a special kind of symbolic objects (Sebag). More generally in the framework of the Esprit II program MLT ("Machine learning toolbox") an interface between Makey (Lebbe and Vignes [13]) and SICLA (Celeux et al. [3]) an interactive system of classification has been implemented and work on X-Windows under Hypernews.

The theory of Symbolic Data Analysis (SDA) that we have developped in this paper may be useful in the framework of vast domains of application as Data Base Systems, Pattern recognition, Image processing, Learning Machine etc.

In Data Base Systems, SDA gives tools to define new kind of units (probabilist and possibilist objects, for instance) and new kinds of queries, expressed by a modal assertion a_x, when the extension is composed of individual objects or by dual modal assertion a_x^{*} when the extension is computed on a set of assertions $A_x \subseteq \mathfrak{a}_x$.

In Pattern Recognition, SDA allows the representation and the analysis of complex patterns; in "Image Processing" SDA may be used for instance, in order to compare several sensors, for data fusion, or for image understanding by classification of high level objects (houses, trees, roads, ...) represented by symbolic objects.

In Machine Learning SDA makes it possible to extend learning algorithms (where input are usually individual objects) to symbolic objects; moreover, by defining symbolic objects on the set Ω of samples and not on the set of description Δ, SDA allows a bridge between statistics and Machine Learning.

Unlike most work carried out in Artificial Intelligence, symbolic data analysis constitutes a "critique of pure reasoning" by giving less importance to the reasoning and more importance to the statistical study of knowledge bases, considered as a set of "symbolic objects". A wide field of research is opened by extending classical statistics to statistics of intensions and more specially by extending problems, methods and algorithms of data analysis to symbolic objects.

Appendix*

PROOF OF THE THEOREMS 1 AND 2

Before giving the proof of both theorems let us remark that $a^*(\mathfrak{a}_x) = f_x(\{g_x(q_i, \{\cup_{jx} q_i^j / q_i^j \in Q_i^{\mathfrak{a}_x}\})\}_i)$; where, by definition, $\mathfrak{a}_x$ is the set of im assertions associated to x and $Q_i^{\mathfrak{a}_x} = \{q_i^\ell / a\ell = \wedge_i [y_i = q_i^\ell] \in \mathfrak{a}_x\} = Q_i^x$ the set of any $\cup_x, \wedge_x, c_x$ combination of elements $q_i^j \in Q_i$ associated to x. Hence, we have $a^*(\mathfrak{a}_x) = f_x(\{g_x(q_i, \cup_{jx} q_i^j / q_i^j \in Q_i^x\})\}_i)$. We set $1_a = \wedge_i [y_i = 1_{O_i}]$ where $1_{O_i}(v) = 1 \ \forall v \in O_i$. We denote $q_i^A = \{q / q \in Q_i^{Ax}\}$, where $A_x \subseteq \mathfrak{a}_x$ and Q_i^A is defined as in section 9 by $Q_i^A = \{q_i / a = \wedge_j [y_j = q_j] \in A_x\}$ which means that Q_i^A is the set of the mappings q_i which define the ith event $[y_i = q_i]$ of any $a \in A_x$.

We extend the operator $\cup_x$ on $\mathbb{R}$ by setting $\forall u_j \in \mathbb{R}$, $\cup_{x\, j=1,n} u_j = u_1 \cup_x u_2 \ldots \cup_x u_n$, where in the case of possibilities we have $u_1 \cup_p u_2 = \text{Max}(u_1, u_2)$ and in the case of probabilities $u_1 \cup_{pr} u_2 = u_1 + u_2 - u_1 u_2$.

We denote $I_i^A(v)$ the set of values taken by $q_i(v)$ when q_i varies in Q_i^{Ax} so $I_A(v) = \{q_i(v) / q_i \in Q_i^{Ax}\}$. See figure 13. Notice that as O_i, $I_A(v)$ is not necessarily countable. □

LEMMA 1

If for any sequence $\{u_n\}$ of rational numbers dense in $I_i^A(v)$, the sequence $U_n = \cup_{\ell=1,n} u_\ell$ converges towards the same limit U, then $q_i^A(v) = U$.

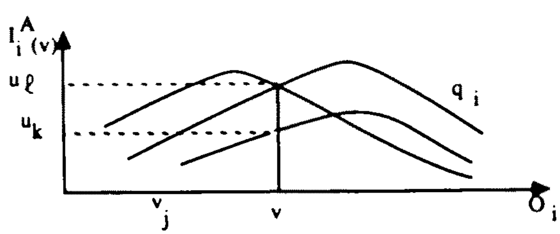

Figure 13. When q_i varies in Q_i^{Ax}, $q_i(v) = u_\ell$ is repeated each time that it exists a different $q_j \in Q_i^{Ax}$ such that $q_j(v) = q_i(v) = u_\ell$.

*A more complete and simpler proof may be found in a paper to be published by Diday, Rohmer and Emilion (Ceremade report, Dauphine University, Paris).

Proof

As the mappings q_i take their values in [0,1], I_i^A (v) is bounded; it is possible to decompose its boundaries by a partition of intervals of lenght 1/2k; we retain from these intervals only the one which contains at least one element of $I_i^A(v)$; we associate to each of these intervals a rational number r and we denote I_r^k (v) the interval of length $1/2^k$ which contains it; given k and v we denote $I_A^k(v)$ the set of these numbers r. At this step O_i is supposed to be a bounded subset of ℝ, it is possible to decompose the intervals defined by its boundaries, also by a partition of lenght $1/2^k$; we associate to each of these intervals which contains at least one element of O_i, a rational number from this interval. The set of these numbers is denoted O_i^k and the elements of O_i^k are denoted $v_1, v_2, \ldots, v_{nk}$, with $n_k = \text{card } O_i^k(\leqslant 2^k$ as some intervals may contains no elements of O_i).

To each $r_\ell \in I_A^k(v_\ell)$ we associate a set of mappings $q \in Q_i^A$ denoted $C_{r\ell}^k(v_\ell)$ and constructed as follows: we consider the set $I_A^k = I_A^k(v_1) \times \ldots \times I_A^k(v_{nk})$ and we suppose that $v_\ell \in \{v_1, \ldots, v_{nk}\}$; we associate to any $r = (r_1, \ldots, r_{nk}) \in I_A^k$ where r_ℓ is fixed a unique $q_r \in Q_i^A$ if it exists, such that for any $j \in \{1, \ldots, n_k\}$, $q_r(v_j)$ belongs to the interval $I_{r_j}^k(v_j)$; the set of these q_r is denoted $C_{r\ell}^k(v_\ell)$; so we have: $C_{r\ell}^k(v_\ell) = \{q_r \in Q_i^A / r = (r_1, \ldots, r_{nk}) \in I_A^k, r_\ell \text{ fixed}, q_r(v_j) \in I_{r_j}^k(v_j), q_r \text{ unique, for each r, when it exists}\}$.

We set

$$q_i^k(v) = \bigcup_{u \in I_A^k(v)}{}_x \bigcup_{q \in C_u^k(v)}{}_x u$$

which means that q_i^k (v) is the x-union of all the values u belonging to $I_A^k(v)$ repeated for each u by the number of times that there exists $q \in C_u^k$ (v).

Since, given k, the sets I_A^k (v) and C_u^k (v) are finite, it is possible to enumerate in a finite sequence, denoted u_ℓ^k, the $u \in I_A^k$ (v), including their C_u^k (v) repetition; thus, we may define a set S^k of these u_ℓ such that

$$S^k = \{u_\ell^k \in I_A^k(v) / \ell = 1, \sum_{u \in I_A^k(v)} \text{card}(C_u^k(v))\}$$

Hence, we get

$$q_i^k(v) = \bigcup_{u\ell \in S_k}{}_x u_\ell^k = \bigcup_{\ell = 1,n_k}{}_x u_\ell^k \text{ where } n_k = \sum_{u \in I_A^k(v)} \text{card } C_u^k(v).$$

When $k \to +\infty$, S^k becomes dense in I_i^A (v) since for any $u \in I_i^A$ (v), $u_\ell^k \in I_A^k$ (v) $\subseteq S^k$ such that $|u_\ell^k) - u| \leqslant 1/2^k$. Therefore, when $k \to \infty$ the sequence $\{u_\ell^k)\}$ becomes a sequence of rational numbers dense in I_i^A (v) and if the assumption of the lemma is satisfied, q_i^k (v) converges towards U. Notice that if I_i^A (v) and O_i are not finite, when $k \to \infty$, card I_A^k (v) $\to +\infty$ and card $C_u^k(v) \to +\infty$; if I_i^A (v) is finite and there exists j such that O_j is infinite, then I_A^k remains finite for any k and card $C_u^k(v) \to +\infty$ when $k \to +\infty$; hence, in all these cases $n_k \to +\infty$ when $k \to +\infty$. The only case where n_k remains finite when $k \to +\infty$ appears when $I_i^A(v)$ is finite and O_j is also finite for any j: in this case it is easy to see that $q_i^k(v)$ will converge always towards the same finite union:

$$U = \bigcup_{u \in I_i^A(v)}{}_x \bigcup_{q \in C_i^u(v)}{}_x u$$

where $C_i^u(v)$ is the finite set of $q \in Q_i^A$ such that $q(v) = u$.

As $k \to +\infty$, we have $I_A^k(v) \to I_i^A(v)$, since by construction $\forall u \in I_i^A(v)$ there exists $u_k \in I_A^k(v)$ such that $|u_k - u| \leqslant 1/2^k$ and therefore for any $u \in I_i^A(v)$ there exists a sequence $\{u_k\}$ with $u_k \in I_A^k(v)$ such that $u_k \to u$ when $k \to +\infty$.

When $k \to +\infty$, we may see in a similar way that $O_i^k \to O_i$ since by the construction of O_i^k, $\forall v \in O_i$, $v_k \in O_i^k$, such that $|v_k - v| \leqslant 1/2k$.

By the construction of $I_A^k(v_i)$, for any $q \in C_i^u(v)$ such that $v \in Q_i$, $u \in I_i^A(v)$ and $q(v) = u$, there exists for any $v_i \in O_i^k$, an element $r_i \in I_A^k(v_i)$ such that $q(v_i) \in I_{r_i}^k$; also by the construction of $C_u^k(v)$ there exists $q_k \in C_u^k(v)$ such that $q_k(v_i) \in I_{r_i}^k$; hence we get for any $v_i \in O_i^k$ $|q_k(v_i) - q(v_i)| \leqslant 1/2^k$ as by construction the length of the interval $I_{r_i}^k$ is $1/2k$; hence when $k \to +\infty$ $O_i^k \to O_i$ and $q_k \to q$, therefore $C_u^k(v) \to C_i^u(v)$.

Finally, as $k \to +\infty$, we have $q_i^k(v) \to U$, $I_A^k(v) \to I_i^A(v)$, $C_u^k(v) \to C_i^u(v)$, it follows that at the limit of the equality

$$q_i^k(v) = \bigcup_{u \in I_A^k(v)}{}_x \bigcup_{q \in C_u^k(v)}{}_x u,$$

we get:

$$U = \bigcup_{u \in I_i^A(v)}{}_x \bigcup_{q \in C_i^u(v)}{}_x u = \cup_x\{q/(v)/q \in Q_i^A\} = q_i^A(v).$$

□

We recall that 1_{O_i} and I_i^A are the mappings $O_i \to [0,1]$ such that $\forall v \in O_i$, $1_{O_i}(v) = 1$ and $I_i^A(v) = \{q_i(v) / q_i \in Q_i^{Ax}\}$; we have also $q_i^A = \{\cup_x q/q \in Q_i^A\}$.

LEMMA 2

If $\forall u_1, u_2 \in [0,1]$, $u_1 \cup_x u_2 \geqslant \text{Max}(u_1, u_2)$, then $\forall A_x \leqslant a_x$ and $\forall v \in O_i$, we have $q_i^A(v) = \text{Max}\{u / u \in I_i^A(v)\}$ and $q_i^{a_x} = 1_{O_i}$.

Proof

First we show that any sequence

$$U_n = \bigcup_{\ell = 1,n} u_n$$

where $\{u_n\}$ is a sequence of rational numbers dense in I_i^A, converges; this follows from the fact that $\forall n\ U_n = U_{n-1} \cup_x u_n \geq U_{n-1}$, since $u \cup_x v \geq \text{Max}(u,v)$, and so the sequence $\{U_n\}$ is increasing, as it is majored by 1. Second, the sequence U_n converges towards U, for, if $u < U$ was its limit we would obtain a contradiction because, since the sequence $\{u_n\}$ is dense in I_i^A (v), there would exist a k such that $u < u_k < 1$ and $U_k = U_{k-1} \cup_x u_k \geq u_k > u$, hence the sequence $\{U_n\}$ would never converge towards u, as it is increasing. Therefore, by applying lemma 1, we get q_i^A $(v) = U$; hence, in the case where $I_i^A(v) = [0,1]$ we have $U = 1$ and so $\forall v \in O_i$, q_i^A $(v) = 1$, therefore $q_i^A = 1_{O_i}$.

Hence, we have proved the theorem in the case where O_i is a bounded set of $\mathbb{R}$. Let $O_i =]a_n,b_n[$, where $\{a_n\}$ and $\{b_n\}$ are two sequences of $\mathbb{R}$ such that, when $n \to +\infty$, $a_n \to -\infty$ and $b_n \to +\infty$. We may say that the theorem remains true with $O_i =]-\infty, +\infty[$ since when $n \to +\infty$ $\forall$ a, b $\in \mathbb{R}$ there exists N_1, N_2 large enough such that the theorem remains true on $]a_{n1}, b_{n1}, [$with $a_{n1} < a$ and $b_{n2} > b$ for (n_1,n_2) such that $n_1 > N_1$ and $n_2 > N_2$. □

Proof of theorem 1 (possibilist objects)

(i) $a^{\bullet}(a_p) = 1$, $a^{\bullet}(\emptyset) = 0$.

It is easy to see that the assumptions of lemma 2 are satisfied as

(a) $\forall$ u,v $\in [0,1]$, $u \cup_p v = \text{Max}(u,v)$ by definition

(b) $\forall u \in [0,1]$, $1 \cup_p u = 1 = \text{Max}(1,u) = 1$; $0 \cup_p 0 = \text{Max}(0,0) = 0$; therefore $\forall_i$, $q_i^{a_p} = 1_{O_i}$. Therefore

$$a^{\bullet}(a_p) = f_p(\{g_p(q_i, 1_{O_i})\}_i) = \text{Max}(\{\sup_{v \in O_i}(\text{Min}(q_i(v), 1_{O_i}(v)))\}_i\}$$

$$= \text{Max}(\{\sup_{v \in O_i}(q_i(v))\}_i) = \text{Max}(\{1\}_i) \text{ as } q_i(O_i) = 1.$$

This implies the existence of $v \in O_i$ such that $q_i(v) = 1$. Therefore we get finally $a^{\bullet}(a_p) = 1$. By definition

$$a^{\bullet}(\emptyset) = f_p(\{g_p(q_i, \emptyset_i)\}_i) \text{ with } g_p(q_i, \emptyset_i) = \sup_{v \in O_i}(\text{Min}(q_i(v), \emptyset_i(v)))$$

$$= \sup_{v \in O_i}(\emptyset_i(v)) = 0 \text{ as } \emptyset_i(v) = 0 \forall v \in O_i.$$

Therefore $a^{\bullet}(\emptyset) = \text{Max}(\{0\}_i) = 0$.

(ii) $a^{\bullet}(A_1 \cup_p A_2) = \text{Max}(a^{\bullet}(A_1), a^{\bullet}(A_2))$. By definition we have: $A_1 \cup_p A_2 = \wedge[y_i = q_i^{A_1} \cup_p q_i^{A_2}]$ with $q_i^{A_k} = \{\cup_{jp} q_i^j / q_i^j \in Q_i^{A_k}\}$; Since the assumptions of lemma 2

are satisfied, q_i^A exists. Hence, we may write:

$$a^*(A_1 \cup_p A_2) = f_p(\{g_p(q_i, q_i^{A_1} \cup_p q_i^{A_2})\}_i\} \text{ where } g_p(q_i, q_i^{A_1} \cup_p q_i^{A_2})$$
$$= \operatorname*{Sup}_{v \in O_i} (\text{Min}\{q_i(v), \text{ Max}(q_i^{A_1}(v), \ q_i^{A_2}(v))\}).$$

Since Min{a,Max(b,c)} = Max{Min(a,b), Min(a,c)}, we have:

$$g_p(q_i, q_i^{A_1} \cup_p q_i^{A_2}) = \operatorname*{Sup}_{v \in O_i} (\text{Max}\{\text{Min}\{q_i(v), \ q_i^{A_1}(v)\}, \ \text{Min}\{q_i(v), q_i^{A_2}(v)\}).$$

But since

$$\operatorname*{Sup}_{v \in O_i} (\text{Max}\{a(v), b(v)\} = \text{Max}\{\operatorname*{Sup}_{v \in O_i}(a(v)), \operatorname*{Sup}_{v \in O_i}(b(v))\}$$

we get:

$$g_p(q_i^{A_1} \cup_p q_i^{A_2}) = \text{Max}\{\operatorname*{Sup}_{v \in O_i} (\text{Min}\{q_i(v), q_i^{A_1}(v)\}), \ \operatorname*{Sup}_{v \in O_i} (\text{Min}\{q_i(v), q_i^{A_2}(v)\})\}.$$

Hence, as $\text{Max}_i(\{\text{Max}(a_i, b_i)\} = \text{Max}(\text{Max}_i(\{a_i\}), \ \text{Max}_i(\{b_i\})\}$ and $a^*(A_1 \cup_p A_2) = \text{Max}_i\ (\{g_p(q_i, q_i^{A_1} \cup q_i^{A_2})$, by definition, we get:

$$a^*(A_1 \cup_p A_2) = \text{Max}\{\operatorname*{Max}_i \ \operatorname*{Sup}_{v \in O_i}(\text{Min}\{q_i(v), q_i^{A_1}(v)), \operatorname*{Max}_i \ \operatorname*{Sup}_{v \in O_i}(\text{Min}\{q_i(v), q_i^{A_2}(v))\}$$

and finally:

$$a^*(A_1 \cup_p A_2) = \text{Max}(a^*(A_1), a^*(A_2)).$$

□

Proof of theorem 2 (probabilist objects)

It is easy to see that the assumptions of lemma 2 are satisfied in case of probabilist union as:

(a) $\forall\, u_1, u_2 \in]0,1[$, we have: $u_1 \cup_{pr} u_2 = u_1 + u_2 - u_1 u_2 \geqslant u_1 + u_2(1 - u_1) \geqslant u_1$, and $u_2 + u_1(1 - u_2) \geqslant u_2$, so that $u_1 \cup_{pr} u_2 \geqslant \text{Max}(u_1, u_2)$.

(b) $\forall\, u \in [0,1]$, $1\ U_{pr}\ u = 1 + u - u = 1$;

(c) $0\ U_x\ 0 = 0$.

Hence it follows from this lemma that $q_i^{a_{pr}} = 1_{O_i}$ and $\forall A_{pr} \subseteq \mathbf{a}_{pr}$, $q_i^{A_{pr}}$ exists.

Now we may prove theorem 2:

(i) $a^{\bullet}(\mathfrak{a}_{pr}) = 1$; $a^{\bullet}(\phi) = 0$.

From $q_i^{a_{pr}} = 1_{Oi}$ we get: $a^{\bullet}(\mathfrak{a}_{pr}) = f_{pr}(\{g_{pr}(q_i, 1_{Oi})\}_i)$; were $g_{pr}(q_i, 1_{Oi}) = S\ \{q_i(v)\ 1_{Oi}\ (v) / v \in O_i\} = S\ \{q_i(v)\ /\ v \in O_i\} = 1$. Therefore: $a^{\bullet}(\mathfrak{a}_{pr}) = f_{pr}(\{1\}_i) = 1$.

By definition we have $a^{\bullet}(\phi) = f_{pr}(\{g_{pr}(q_i, \phi_i)\}_i)$, where $\forall\ v \in O_i,\ \phi_i(v) = 0$; hence, $g_{pr}(q_i, \phi_i) = \Sigma\{q_i(v)\phi_i(v)/v \in O_i\} = 0$; therefore $a^{\bullet}(\phi) = \text{Mean}\ (\{0\}_i) = 0$.

(ii) $\forall\ A_1, A_2 \subseteq \mathfrak{a}_x\ a^{\bullet}(A_1 \cup_{pr} A_2) = a^{*}(A_1) + a^{\bullet}(A_2) - a^{\bullet}(A_1 \cap_{pr} A_2)$.

As $q_i^{A\ell}$ exists, we may write, by definition:

$a^{\bullet}(A_1 \cup_{pr} A_2) = f_{pr}(\{g_{pr}(q_i, q_i^{A_1} \cup_{pr} q_i^{A_2})\}_i)$ with $g_{pr}(q_i, q_i^{A_1} \cup_{pr} q_i^{A_2} = \langle q_i,\ \ q_i^{A_1} + q_i^{A_2} - q_i^{A_1} \cap_{pr} q_i^{A_2}\rangle = \langle q_i, q_i^{A_1}\rangle + \langle q_i, q_i^{A_2}\rangle - \langle q_i, q_i^{A_1} \cap_{pr} q_i^{A_2}\rangle$

As f_{pr} is the mean, it results that:

$a^{\bullet}(A_1 \cup_{pr} A_2) = \text{Mean}(\{\langle q_i, q_i^{A_1}\rangle + \langle q_i, q_i^{A_2}\rangle - \langle q_i, q_i^{A_1} \cap_{pr} q_i^{A_2}\}_i) = \text{Mean}\ (\{\langle q_i, q_i^{A_1}\rangle\}_i) + \text{Mean}\ (\{q_i, q_i^{A_2}\}_i) - \text{Mean}(\{\langle q_i, q_i^{A_1} \cap_{pr} q_i^{A_2}\}_i) = a^{\bullet}(A_1) + a^{\bullet}(A_2) - a^{\bullet}(A_1 \cap_{pr} A_2)$.

□

References

[1] P. Brito and E. Diday, Pyramidal representation of symbolic objects, in: NATO ASI Series, Vol. F 61, *Knowledge Data and Computer-Assisted Decisions*, eds. M. Schader and W. Gaul (Springer, 1990).

[2] P. Brito, Symbolic objects: order structure and pyramidal clustering, in this volume.

[3] G. Celeux, E. Diday, G. Govaert, Y. Lechevallier and H. Ralambondrainy, *Classification Automatique: Environement Statistique et Informatique* (Dunod, 1989).

[4] G. Choquet, Théorie des capacités, Ann. Inst. Fourrier 5 (1953) 131–295.

[5] F.A.T. De Carvalho, Histogramme en analyse des données symboliques, Dissertation Univ. Paris 9 Dauphine (1991).

[6] A.P. Dempster, Upper and lower probabilites induced by a multivalued mapping, Ann. Math. Statist. 38 (1967) 325–339.

[7] E. Diday, Knowledge representation and symbolic data analysis, in: NATO ASI Series, Vol. F 61, *Knowledge Data and Computer-Assisted Decisions*, eds. M. Schader and W. Gaul (Springer, 1990).

[8] E. Diday, Des objets de l'analyse des données à ceux de l'analyse des connaissances, in: *Induction Symbolique-Numérique à Partir de Données*, eds. Y. Kodratoff and E. Diday (CEPADUES, Toulouse, 1991).

[9] E. Diday, From data to Knowledge: new objects for a statistical analysis, in: *New Techniques and Technologies for Statistics Conference*, GMD, Bonn (1992).

[10] D. Dubois and H. Prade, *Possibility Theory* (Plenum, New York, 1988).

[11] D. Dubois and H. Prade, A set-theoretic view of belief functions, Int. J. General Syst. 12 (1986) 193–226.

[12] S.L. Lauritzen and D. J. Spiegelhalter, Local computation with probabilities on graphical structures and their application to expert system, in: *Readings in Uncertain Reasoning*, eds. G. Shafer and J. Pearl (Morgan Kaufmann, 1990).

[13] J. Lebbe, R. Vignes and S. Darmoni, Symbolic numeric approach for biological knowledge representation: a medical example with creation of identification graphs, in: *Proc. Conf. on Data Analysis, Learning Symbolic and Numerical Knowledge*, Antibes, ed. E. Diday (Nova Science Publ., New York, 1990).

[14] R.S. Michalski, E. Diday and R.E. Stepp, Recent advances in data analysis: clustering objects into classes characterized by conjonctive concepts, *Progress in Pattern Recognition*, vol. 1, eds L. Kanal and A. Rosenfeld (1982).

[15] J. Pearl, *Probabilist Reasoning in Intelligent Systems* (Morgan Kaufmann, San Mateo, 1988).

[16] G. Schafer, Perspectives on the theory and practice of belief functions, Int. J. Approx. Reasoning 4 (1990).

[17] G. Schafer, *A Mathematical Theory of Evidence* (Princeton University Press, 1976).

[18] B. Shweizer and A. Sklar, Statistical matric spaces, Pacific J. Math. 10 (1960) 313–334.

[19] L.A. Zadeh, Quantitative fuzzy semantics, Inf. Sci. (1971) 159–176.