INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

An Introduction to Symbolic Data Analysis

Edwin DIDAY

N° 1936
Août 1993

PROGRAMME 5

Traitement du signal,
automatique et
productique

Rapport
de recherche

1993

AN INTRODUCTION TO SYMBOLIC DATA ANALYSIS*

UNE INTRODUCTION A L'ANALYSE DES DONNEES SYMBOLIQUES*

INRIA-Rocquencourt
Domaine de Voluceau
78153 Le Chesnay Cedex

Abstract

The main aim of the symbolic approach in Data Analysis is to extend problems, methods and algorithms used on standard data to more complex data called "symbolic objects", in order to distinguish them from objects (described by numerical or categorical variables) treated by standard Data Analysis methods."Symbolic objects" extend classical objects of data analysis in two ways : first, in the case of individuals, by giving the possibility of introducing structured information in their definition; second, in the case of sets or classes, by being intentionally defined.In both cases,in order to represent uncertain knowledge, it may be useful to use probabilities , possibilities (in case of vagueness and imprecision for instance), belief (in case of probabilities only known on parts and to express ignorance); that is why,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. Some mixture decomposition problems on these objects are settled.We show that in some cases,fractals are well adapted to representing duality between symbolic objects. Sets of symbolic objects are represented by categories of different kinds (hierarchies,pyramids and lattices). 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. Finally, important steps of a symbolic data analysis are described and illustrated by an example concerning road accidents.

Key-words : Knowledge Analysis, Symbolic Data Analysis, Metadata, Metaknowledge, Probability, Possibility, Evidence theory, Cognition.

Résumé

L'objectif principal de l'approche symbolique en Analyse des Données est d'étendre les problèmes, méthodes et algorithmes de l'Analyse des Données classique a des objets plus aptes à representer des connaissances. On présente d'abord les "objets symboliques" (sortes "d'atomes de connaissances") et ce qui les distingue des objets classiques de l'analyse des données usuelles. Ces objets, qui constituent les individus de l'analyse des données symboliques, permettent de représenter (en extension), des individus complexes ou (en intension), des classes d'individus, par des conjonctions de propriétés où des descripteurs peuvent prendre des valeurs multiples et pondérées (selon différentes sémantiques) et sont parfois reliés entre eux par des relations d'ordre logique. Dans ces deux cas, on s'intéresse a des objets de types probabilistes, possibilstes ou crédibilistes afin d'exprimer des connaissances incertaines. On introduit des outils pour manipuler ces objets : union, intersection, généralisation, extension etc.; on construit ainsi, un espace d'objets symboliques dual où les individus sont des objets définis en intension ; dans cet espace, on énonce trois théorèmes étendant les probabiltés, possibilités et credibilités à ce type d'objet et l'on pose des problèmes de décomposition de mélanges en lois de lois. On s'intéresse ensuite, à la représentation graphique de ces objets par différentes catégories (hiérarchies, pyramides, treillis etc., d'objets symboliques). En utilisant la dualité, on peut construire des suites d'objets symboliques (devenant individus dans l'espace dual suivant) ; ces suites définissent des fractals dans certains cas que nous préciserons. On décrit différents types d'analyse des données symboliques ainsi que les principales étapes d'une telle analyse. On illustre enfin par une application concernant la construction et l'étude de scénarios d'accidents de la route.

**Tutorial at IFCS'93..*

OUTLINE

1. Introduction

1.1. Metadata and symbolic objects

Metadata are data on the data, they are obtained at least from the two following ways : i) by aggregation of observed data, in order to get, for instance, data on a district from data on villages ; by this way, we obtain means, intervals of variation, means squares, correlations etc., which are metadata caracterising each district ; ii) from the knowledge of experts, for instance if an expert wishes 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 metadata 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 :

$a_i = [\text{weight} = [300,400]] \wedge [\text{color} = \{\text{red, white}\}] \wedge [\text{if } [\text{color} = \text{white}] \text{ then } [\text{weight} \leq 350]]$,

where a_i, associated to represents the ith village, is a mapping defined on the set of fruits Ω such that for a given fruit $w \in \Omega$, a_i (w) = true if the weight of w 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 statistics 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 1990,1991) is to provide tools for answering this problem.

In some fields, a boolean representation of the knowledge (a_i (w) = 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 $a_i = [\text{color} = q_i]$ 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 $a_i = [\text{color} = \{\text{red, white}\}]$ we have $q_i(\text{red}) = q_i(\text{white}) = 1$ and $q_i = 0$, for the other colors; in the probabilist case, if $a_i = [\text{color} = 0.9 \text{ red}, 0.1 \text{ white}]$ we have $q_i(\text{red}) = 0.9$, $q_i(\text{white}) = 0.1$.

If an expert says that the fruits are red we may represent this information by a symbolic object $a_i = [\text{color} = q_i]$ where q_i is a "possibilist" function in the sense of Dubois and Prade (1986) ;

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 in the sense of Schafer (1990) computed by using a "probability assignment" function denoted m_i such that m_i(red) = 0.6, m_i(white) = 0.3 and m_i (O) = 0.1, which expresses ignorance as O is the set of possible colors. Depending on the kind of the 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 (w) ; if there is doubt about the color of a given fruit w, for instance, if the expert says that "the color of w, is red or pink" then, w may be described by a charateristic function r and represented by a symbolic object w^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 w^s = [color = r] to compute a_i (w) we introduce a comparison function g such that a_i (w)= g (q_i,r) measures the fit between q_i and r. What is the meaning of a_i (w) ? May we say that a_i (w) measures a kind of probability, possibility or belief that w belongs to the class of fruits described by a_i when q_i and r, depending on the background knowledge, are charateristic, probability, possibility or belief functions respectively ? To answer this question we needed to extend a_x (where x represents a kind of background knowledge) to a^*_x defined on $\mathcal{a}_x$ a set of symbolic objects and to define set operators $OP_x = \{\cup_x, \cap_x, c_x\}$ in $\mathcal{a}_x$ adapted to x. If we say that classical 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; the 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 , events are generally identified as parts of the sample space Ω. In computer science, object oriented languages consider more complex 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 ; 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 first aim is to define complex objects called "boolean objects", inspired by those of oriented object languages in such a way that data analysis becomes generalized into symbolic data analysis. Such 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, instanciated structured information (see the case of "horde" in § 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 than 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 [wheight $\leq$ 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) and "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 have been neglected) and more generally in AI (where the results here obtained , in theorem 1,2,3, concerning 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, especially to statisticians ; secondly, because they cannot be used easily in the case of probabilist, possibilist and belief objects where uncertainty is present.

1.2. "Science of objects" attempting a historical review

In computer science we know the growing popularity of object oriented languages based on intensionally defined sets or classes called "objects" or "frames" (Minsky (1975)) and using "messages", "methods" and inheritance properties ; this new kinds of computer language started at Palo Alto (1971) with SMALLTALK, inspired by SIMULA (1967) and now many widely used languages based on objects exist as for instance SMALLTALK, CLOS,TELLOS,C++ etc.

In fact man has always been concerned by the problem of representation, description and identification of objects in the world. In prehistorical times, we may imagine the picture of a mammoth for example shows that man was already able to distinguish between a given mammoth (that has just been eaten for instance), from the class of mammoth represented by this picture. In Genesis God asks Adam to give a name to "the animals of the fields and the birds of the sky". In the fifth century B.C., Plato (428-348 BC) considers the existence of "universals" in an ideal world representing concrete individuals to which they apply ; his pupil

Aristotle introduced the notion of "Science of objects" in "De Partibus Animalium" (Louis (1956)) and described objects intensionally by their properties pointing out clearly the duality between objects and their description. About 600 years after Aristotle, Porphyre (Bochenski (1970)) introduced successive division of classes of objects by using "genus" and "species" which have technical meanings in logic and have been used in taxonomy (Sneath and Sokhal (1973)).

In the 17 th century the Port Royal logicians, Arnaud and Nicole, had this to say :

"Now in these universal ideas there are two things which it is important to keep quite distinct : comprehension and extension. I call the comprehension of an idea the attributes which it contains and which cannot be taken away from it without destroying it ; thus the comprehension of the idea of a triangle includes, to a superficial extent, figure, three lines, three angles, the equality of these three angles to two right angles, etc. I call the extension of an idea the subjects to which it applies, which are also called the inferiors of a universal term, that being called superior to them. Thus the idea of triangle in general extends to all different kinds of triangles.

Chomski (1965) pointing out the importance of the notion of "idea" in the sense of Arnault and Nicole said *"the comprehension (i.e. intension or meaning) of an idea is the fundamental notion in semantic interpretation and in so far as the deep structure of language is regarded as the direct reflection of mental process, it is the fundamental notion in the analysis of thought"*.

As recalled by Sutcliffe (1992), Frege (1848-1925) in his Grundgesetze der Arithmetik (see Geach and Black, 1952) considered that a concept is a mapping which identifies the membership relation between an individual and a class and he said that :

"A definition of a concept (of a possible predicate) must be complete : it must unambiguously determine, with regards to any object, whether or not is falls under the concept (whether or not the predicate is truly assertible of it). Thus there must not be any object as regards which the definition leaves in doubt whether it falls under the concept or not ; though for us, men with our defective knowledge, the question may not always be decidable".

Heidegger (1936) trying to answer the question "what is a thing ?" defined two kinds of things :

i) things in the sense of being "within reach" : a stone, a piece of wood, etc.
ii) things which "unify" things of the first kind.

Things which unify have an intension (intent, comprehension) and an extension formed by things of the first kind. Wille (1981) says as "in traditional philosophy (see H. Wagner (1973)) that things for which their "intent" describes all the properties valid for the individual objects of their extension are called "concept".

All those notions called "universals" (Plato), objects (Aristotle), "genus" or "species" (Porphyry), "ideas" (Arnaud and Nicole), "things which unify" (Heidegger), "concepts" (Frege, Wagner) "frames" or "objects" (object oriented computer languages) represent classes of individuals (instances, subjects).

This representation may be done in numerous ways ; in the aristotelian tradition it is defined by a logical conjunction of properties. The aristotelian logic representation has inspired many naturalists in building natural systems such as Cesalpino (1519-1603), Linnaus (1707, 1778) ; A.- L. De Jussieu (1774) closer from the aristotelian tradition announce his famous "Principle of Subordination of the characters" (Leguyader 1988) which means, roughly, that a class is characterized by a conjunction of properties inherited by subclasses.

On the other hand another tendency is well represented by Adanson (1727-1806), a French naturalist who was very much ahead of his time, working on a natural classification of organisms. He considered that classes are defined by a high ressemblance degree, by the individuals belonging to it ; these ressemblances are estimated by examining the characteristics of the individuals ; by doing so he defined the bases (200 years ago) of modern Numerical Taxonomy (Sneath and Sokhal (1973) Benzecri (1974), Gower (1974), Ward (1963), Dale and Anderson (1973)). These two tendencies (Aristotle versus Adanson) may be characterized by the fact that in the aristotelean tradition classes are "monothetic", whereas from the Adanson point of view classes are "polythetic" (Jevons 1877), Kaplan and Schlott (1951), Beckner (1959), Sneath (1962)) ; a class is monothetic, if there is a set of sufficient and necessary properties for membership in the class thus defined ; in contrast, it is polythetic if no property is nessary or sufficient to make an individual member of the class.

A third tendancy, coming from psychology and cognitive science, is to consider that classes must be represented by prototypes ; in contrast, with the Aristotle tradition, where all the members of a class are identical, Rosh (1978) says that classes "tend to become defined in terms of prototypes or prototypical instances that contain the attributes most representative of items inside the class" ; Smith an Medin (1981) said *"that the view of concepts we had inherited from Aristotle was severely lacking and needed to be replaced by a theory based on prototypes"*. D. Dubois (1992) add that recent research on the semantic of memory have led to the introduction of three key concepts :

a) basic level : "fruits" have a higher level than "apples",

b) typicality : some examples (instances) of a concept are more typical than others, unlike the classical conception where all instances are equivalent,

c) prototypes : which are the most representative examples. As a consequence of this point of view, Descles (1986, 1991) says that a concept has to be defined not only by its intension and extension but also by a prototype (whose extension is the set of individuals which incarnate the concept). How to obtain the classes and their representation ?

Briefly, we may say that there are three tendencies :

The first proposed by A.- L. de Jussieu (1774) and Lamark (1778) is in the aristotelean tradition and consists in defining top down the classes by a good choice of properties which characterize them from the most general to the most specific, in this way we obtain a decision tree where each node is characterized by a conjunction of properties.

In this tradition and by starting from examples, among many others the following authors have continued this kind of approach : Belson (1959), Morgan and Sonquist (AID programs (1963)), Lance and Williams (1966), Benzecri (1973), Sneath and Sokhal (1973), Breiman and al (1984), Qinlan (1986) ; by starting from the description of classes we have to mention : Pankurst (1970); Payne (1975) Gower (1975), Virion (1988), Lebbe and Vignes (1991), H. Ralambondrainy (1991), Ganascia (1991), Sebag et al (1991).

The second, put forward by Adanson (1757) who gave the first sequential agglomerative hierarchical clustering (SAHN) algorithm ; this well known bottom up algorithm starting by classes reduced to individuals, merges at each time the most similar classes. This tendance is now well represented by Numerical Taxonomy (Ward (1963), Lerman (1970), Sneath and Sokhal (1973) Benzecri (1974), Jambu (1978) Diday et al (1979), Diday et al (1984), M. Roux (1985), Celeux et al (1989)). The classes obtained in this way contain "Similar" objects. It is then possible to generalize them in term of conjunction of properties as suggested by Sokhal and Sneath (1973).

Whereas, the first tendance gives monothetic classes by a top down process, the second gives polythetic classes by a bottom up process ; other kinds of processes are possible, for instance, in Brito and Diday (1990), Brito (1992) an ascending process building a Pyramid (which is a generalisation of hierarchical trees allowing overlapping clusters) of monothetic classes is described.

The third tendance consists in looking directly for classes and their representation ; for instance, the "Dynamic Clustering Method" (Diday (1971), Diday et al (1979), Diday and Simon (1976)) define a general framework and algorithms which consists of discovering simultaneously classes and their representation in such a way that they "fit" together as well as possible ; this approach (see Diday et al (1979)), has been used with several kinds of inter-class structure (partitions, hierarchies, ...) and representation modes (seeds, probability laws, factorial axis, regressions etc...) in Diday (1976) a logical representation of clusters (as in conceptual clustering) is proposed. As regards to "Conceptual Clustering" algorithm based in the Dynamical Clustering Method or inspired by it, mention should be made of Diday, Govaert, Lechevallier and Sidi (1980), Michalski, Diday, Stepp (1982), Michalski, Stepp (1983) ; conceptual clustering has inspired other authors such as Langley and Sage (1984), Lebowitz (1983), Fisher D.H. (1987), Fisher and Langley (1986) for review.

1.3. An intuitive introduction to symbolic objects

If we are able to recognize usual individual objects of the real world, for instance chairs, we may imagine that it is because there is something in our mind which may be represented by a mapping denoted "a" defined on a set of individual objects denoted "Ω", which satisfies at least the following properties : for an individual object $w \in \Omega$ considered to be a chair, $a(w) =$ true and if $a(w) =$ false then w is not considered to be a chair.

In a latent situation if someone asks : "what is a chair ?" we are able to give a description by a set of properties necessarly satisfied by something that is a chair, which represents our knowledge on the notion of chair ; if the mathematical representation of this description is denoted "d" we may suppose that there is in our mind a "way" to build the mapping a, knowing d ; this way is represented by a mapping "h_Ω" such that $h_\Omega(d) = a$. More precisely, if D represents the set of "admissible" descriptions of any individual object of Ω and A the set of "admissible" mappings a, this way may be represented by a mapping denoted : $h' : DxD \rightarrow \{True, false\}$ which associates to my description of the notion of chair (that I have in my mind) denoted $d_M \in D$ and to this chair $w \in \Omega$ (on which I am sitting) whose description is denoted $d_w \in D$ the value $h'(d_M, d_w) = a(w) =$ true where $a \in A$; hence, fixing a description d_M, h' defines the mapping a when w varies, in other words $h'(d_M, .) = a(.)$; therefore, h_Ω is the mapping $D \rightarrow A$ such that $h_\Omega(d) = a$ iff $h'(d,.) = a(.)$. Roughly speaking, a mapping denoted "a" obtained from a pair (h_Ω, d) by $h_\Omega(d) = a$ is a symbolic object. Notice that we may have $a = h(d) = h(d_1)$ with $d \neq d_1$ which expresses the fact that two different descriptions may be sufficient to recognize the same object. In terms of concept we may say, following Frege that a is the mapping $\Omega \rightarrow$ [true, false] associated to the concept "the chairs", associated to the class of all known individual chairs, d is an intension of this concept and $\Omega' = a^{-1}$ (true) is its extension. Often, in the following, the name of the class, the name of the corresponding concept and the name of the mapping a, will be the same. Instead of considering a concept as "the chairs" whose extension is a set of individual things, we may be interested by a concept whose extension is a set of concepts ; for instance, the extension of the concept "the furniture" is "the chairs", "the tables" etc. We will associate (see section 9) to this kind of more general concepts a mapping denoted $a^* : P(\Omega) \rightarrow$ [true, false] where $P(\Omega)$ is the power set of Ω and a description d^* of a^* is an intension of intensions. This kind of mapping will be used to analyse a set of symbolic objects which is one of the aims of Symbolic Data Analysis and the main aim of Knowledge Analysis. As mentioned by Duquenne (1986), Carnap (1947) says :

"The purpose of this paper is to defend the thesis that the analysis of intension, for a natural language, is a scientific procedure, methodologically just a sound as the analysis of extension". In this paper, we follow the same idea, in order to analyze, organize and extract metaknowledge (rules, for instance) from a knowledge base considered as a set of symbolic objects. Hence, our aim is not as usual in AI to model reasonings but to discover regularities ; as noticed by Kant (1785) "Any ra tional knowledge is either material and concerns some objects, or formal and concern understanding and reasoning in themselves and universal rules of general thinking, regardless of objects". This work concern the first kind of rationnal knowledge.

2. Intensionally defined classes

This section, which main aim is to position precisely : descriptions, symbolic objects and complexes (defined for instance, in Michalski et al (1982)), may be dropped in a first lecture.

2.1 Description and symbolic objects in the boolean case

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

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 2.4 we compare these different kinds of intension ; the extension of a in Ω is a subset of Ω denoted $Ext(a/\Omega)$ and defined by $Ext(a/\Omega) = \{\omega \in \Omega / a(\omega) = true\}$; the extension of b is a subset of Δ defined by $Ext(b/\Delta) = \{\delta \in \Delta / b(\delta) = true\}$; the extension of $d \in D$ in X is denoted Ext(d/X) ; by definition,we set $Ext(d/\Omega) = Ext(a/\Omega)$ and $Ext(d/\Delta) = Ext(b/\Delta)$.

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

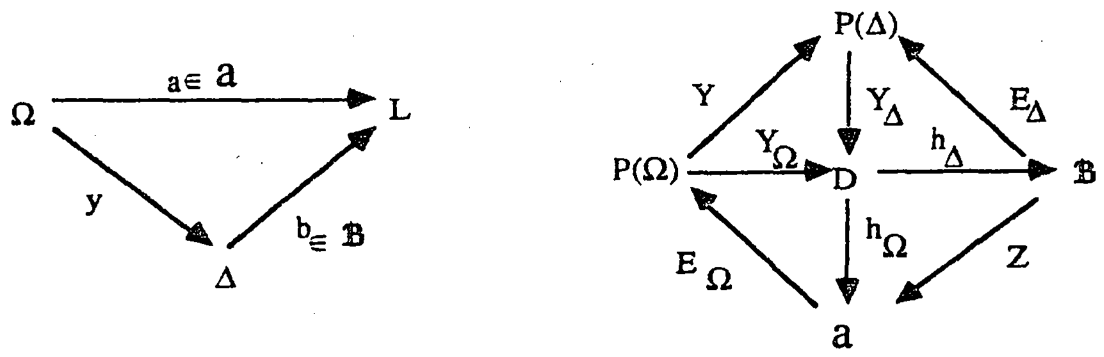

Figure 1 : Any element of D, $\mathcal{B}$ or $\mathbf{a}$ may be considered as a symbolic object.

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, $\mathcal{B}$ or $\mathbf{a}$. Notice that in probability theory, probabilities are usually defined on the set $(\Delta, P(\Delta))$.

A symbolic object is a mapping defined by a set of properties concerning a subset of Ω. Hence, any element of $\mathcal{B}$ or $\mathbf{a}$ 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 $\Delta = O_1 \times .... \times O_p$ where O_i is a domain containing a set of possible values (the color of fruits, for instance) and $D = P(O_1) \times .... \times P(O_p)$; it results in the finite case, that card $P(\Delta) = \text{card } 2^{\Pi_i \text{ card } O_i}$ and card $(D) = 2^{\Sigma_i \text{ card } O_i}$; hence D, which is included in $P(\Delta)$, and contains only the monothetic classes is generally much smaller than $P(\Delta)$.

In this case, if $d = (V_1,....V_p)$ 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 when $w = (x_1,....,x_p)$, $b(w) =$ true iff the statements $x_i \in V_i$ are true ; if moreover $h_\Omega(d) = a$ we have $a(w) = \wedge_i [y(w) \in V_i]_\Omega$ which may be written $a(.) = \wedge_i [y(.) \in V_i]_\Omega$ which is simplified in $a = \wedge_i [y = V_i]_\Omega$ and sometimes when there is no ambiguity in $a = \wedge_i [y = V_i]$; on the next section we compare this kind of symbolic objects to l-complexes introduced by Michalski (for instance in Michalski and al (1981)).

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 $\Delta = O_1 \times O_2$; W 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 description d defined by 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 $V_2 = \{red, white\}$; $Y(\Omega')$ is the set of descriptions Δ' of the fruits of the village and $Y_\Delta(\Delta') = d = V_1 \times V_2$.

$E_\Delta(b) = Ext(b/\Delta)$ is the set $\Delta'' \in P(\Delta)$ of the descriptions δ such that $b(\delta) =$ true and so, such that $\delta \in d = V_1 \times V_2$; hence, $Y(\Omega') = \Delta' \subseteq \Delta''$; $E_\Omega(a) = Ext(a/\Omega)$ is the set $\Omega'' \in P(\Omega)$ of individual objects $w \in \Omega$ such that $a(w) =$ true and so, such that $y_1(w) \in V_1$ and $y_2(w) \in V_2$, 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 add to the description $a = [\,y = [300,400]] \wedge [color = \{red, white\}]$ the constraint [if [color = white] then [weight $\leq$ 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, $\mathfrak{a}$, $\mathfrak{B}$, 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 of 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 $\mathfrak{B}$ and $\mathfrak{a}$.

Two natural choices for Δ are the following : the first denoted Δ_1 is the set of realisable 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" (observable or not observable) 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 l-complexes (Michalski et al. (1981)) which elements are logical expressions of the kind $c = \wedge_i [X_i = V_i]$ where $V_i \subseteq O_i$ and the statement $[X_i = V_i]$ means "value of X_i is one of the elements of V_i"; from the definition of $\mathcal{B}$ it results that it is not in bijection with C if $\Delta_1 = \Delta$ and in bijection if $\Delta = \Delta_2$. The comparison between the different sets of intensions, is given in figure 2 where the sign $\Leftrightarrow$ means the existence of a bijection.

Δ	y is a bijection	y is not a bijection
Δ_1	$\mathcal{A} <=> \mathcal{B} <\neq> C$	$\mathcal{A} <\neq> \mathcal{B} <\neq> C$
Δ_2	$\mathcal{A} <=> \mathcal{B} <=> C$	$\mathcal{A} <\neq> \mathcal{B} <=> C$

Figure 2 : Comparison between the sets of intensions ;
in any case $\mathcal{B} <=> D$; C is the Michalski set of l-complexes

Example :

Case a) $\Delta = \Delta_1$ and y is bijective.

Let be $\Omega = \Omega_1 = \{w_1, w_2, w_3\}$ the set of observed individuals, characterized by a mapping $y = (y_1, y_2) : \Omega \rightarrow (O_1, O_2)$ where $O_1 = O_2 = \{1, 2\}$ such that $y(w_1) = (y_1(w), y_2(w)) = d_1 = (1,1)$, $y(w_2) = d_2 = (2,2)$ and $y(w_3) = d_3 = (1,2)$; as y is bijective, it results that $\Delta = \Delta_1 = \{d_1, d_2, d_3\}$; $D_1 = Y_\Omega(P(\Omega_1)) = Y_\Delta(P(\Delta_1))$ where $Y_\Omega(\Omega') = \cup\{y(w)/w \in \Omega'\}$ and $Y_\Delta(\Delta') = \cup\{d/d \in \Delta'\}$; hence $D_1 = \{d_1, d_2, d_3, d_1 \cup d_2, d_1 \cup d_3, d_2 \cup d_3, d_1 \cup d_2 \cup d_3, \emptyset\}$ where, for instance $d_1 \cup d_2 = \{d_1, d_2\} = \{(\{1,2\}, \{1,2\}) \setminus (1,2), (2,1)\}$ and $d_1 \cup d_2 \cup d_3 = \{(\{1,2\}, \{1,2\}) \setminus (2,1)\} = \{d_1, d_2, d_3\}$.

By definition $\mathcal{B} = \mathcal{B}_1 = h_\Delta(D_1) = \{b_i\}_{i=1,7}$ where $\forall \delta_i \in D_1$, $b_i = h_\Delta(\delta_i)$; for instance, $h_\Delta(d_1) = b_1 = [y_1 = 1]_\Delta \wedge [y_2 = 1]_\Delta$ and $h_\Delta(d_1 \cup d_2 \cup d_3) = [y_1 = \{1,2\}]_\Delta \wedge [y_2 = \{1,2\}]_\Delta \wedge$ [if $y_1 = 2$ then $y_2 \neq 1]_\Delta$.

By definition $\mathcal{A} = \mathcal{A}_1 = h_\Omega(D_1) = \{a_i\}_{i=1,7}$ where for instance $h_\Omega(d_1) = [y_1 = 1]_\Omega \wedge [y_2 = 1]_\Omega$ and $h_\Omega(d_1 \cup d_2 \cup d_2) = [y_1 = \{1,2\}]_\Omega \wedge [y_2 = \{1,2\}_\Omega \wedge$ [if $y_1 = 2$ then $y_2 \neq 1]_\Omega$.

Let be $D = P(O_1) \times P(O_2)$, by definition of the set C of ℓ-complexes, it results that C is in bijection with D ; notice that D_1 is included in Δ_1, which is not the case of D. Finally there is a bijection between $\mathcal{B}_1$ and D_1, $\mathcal{a}_1$ and D_1, D and C but not between D and D_1. This, may be summarized by $\mathcal{a} \Leftrightarrow \mathcal{B} <\neq> C$.

In the following cases, $O = O_1 \times O_2$, $y = (y_1, y_2)$, the w_i, d_i, Y_Ω and Y_Δ are defined by the same way.

Case b) $\Delta = \Delta_2$ and y is bijective.
In this case $\Omega = \{w_1, w_3, w_3, w_4\}$, the set of possible individuals ; $\Delta = \Delta_2 = y(\Omega) = \{d_1, d_2, d_3, d_4\}$ where $d_4 = (2,1)$; hence y is bijective and $D = Y_\Omega(P(\Omega)) = Y_\Delta(P(\Delta_2)) = P(O_1) \times P(O_2)$; $\mathcal{B} = \mathcal{B}_2 = h_\Delta(D) = \{b_i\}_{i=1,16}$ and $\mathcal{a} = \mathcal{a}_2 = h_\Omega(D) = \{a_1, \ldots, a_{16}\}$. Finally in this case we have $\mathcal{B} \Leftrightarrow \mathcal{a} \Leftrightarrow D \Leftrightarrow C$.

Case c) $\Delta = \Delta_1$ and y is not bijective.
In this case, we choose $\Omega = \Omega_1 = \{w_1\}$the only observed individual, $\Delta = \Delta_1 = \{d_1, d_2, d_3\}$, hence $y(\Omega) = d_1$ is strictly included in Δ and so y is not bijective ; hence, $Y_\Omega(P(\Omega)) \neq Y_\Delta(P(\Delta_1))$, we choose $D = D_1 = Y_\Omega(P(\Delta_1)) = D_1$ which has been already defined in the case a) ; $\mathcal{B} = \mathcal{B}_1 = h_\Delta(D_1) = \{b_i\}_{i=1,7}$ as in case a) ; $\mathcal{a} = h_\Omega(D_1) = \{a_1, a_2\}$, where $h_\Omega(d_1) = h_\Omega(d_1 \cup d_2) = h_\Omega(d_1 \cup d_3) = h_\Omega(d_1 \cup d_2 \cup d_3) = a_1$ and $h_\Omega(d_2) = h_\Omega(d_3) = h_\Omega(d_2 \cup d_3) = a_2$ as $a_1(w_1) =$ true and $a_2(w_1) =$ false. Therefore, we have in their case : $D = D_1 <\neq> C$, $\mathcal{B} <\neq> D_1$ and $\mathcal{B} <\neq> \mathcal{a}$; which may be summarized by $\mathcal{a} <\neq> \mathcal{B} <\neq> C$.

Case d) $\Delta = \Delta_2$ and y is not bijective.
We choose $\Omega = \Omega_1 = \{w_1\}$, $\Delta = \Delta_2$ as in case b) ; $D = Y_\Delta(P(\Delta_2)) \neq Y_\Omega(P(\Omega))$; $\mathcal{B} = h_\Delta(D) = \{b_i\}_{i=1,16}$ $\mathcal{a} = h_\Omega(D) = \{a_1, a_2\}$ as in case c) ; finally it results that $\mathcal{a} \Leftrightarrow \mathcal{B} \Leftrightarrow D \Leftrightarrow C$.

2.5 Complete symbolic objects and lattices on $\mathcal{a}$, $\mathcal{B}$ and $\mathcal{C}$

When we associate to an element $\Omega' \in P(\Omega)$ a description $Y_\Omega(\Omega') = d \in D$ the extension of d in Ω which is $E_\Omega(h_\Omega(d))$ contains Ω' as it is the set of $w \in \Omega$ such that $y(w) \in d$; 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 denote $\mathcal{a}_c$ (resp. $\mathcal{B}_c$) the set of complete symbolic objects included in $\mathcal{a}$ (resp. $\mathcal{B}$). We define a partial order on a set of symbolic objects by stating that a symbolic object s_1 is lower than a symbolic object s_2 iff the extension of s_1 is contained in the extension of s_2. If 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 88), and $\mathfrak{B}_c$ (see Brito 93) constitute a lattice.

Example :

Let be $\Omega = \{w_1,w_2\}$ described by $y : \Omega \to O = \{1,2\}$ such that $y(w_1) = 1$, $y(w_2) = 2$; therefore $y(\Omega) = \Delta = \{\delta_1,\delta_2\}$ where $\delta_1 = 1$ and $\delta_2 = 2$; it results also that $D = Y_\Omega(P(\Omega)) = Y_\Delta(P(\Delta)) = P(\Delta) = P(O) = \{\{1\},\{2\},\{1,2\},\varnothing\}$.

We define the following symbolic objects of $\mathfrak{a}$: $a_1 = [y=1]_\Omega$, $a_2=[y=2]_\Omega$, $a_3=[y=\{1,2\}]_\Omega$ and $a_4=[y=\varnothing]_\Omega$; we choose $\Omega = \{w_1\}$, therefore $a_1=a_3$ and $a_2=a_4$.

We define also the set $B = h_\Delta(D) = \{b_i\}_{i=1,4}$ where $b_1 = [y=1]_\Delta$, $b_2 = [y=2]_\Delta$, $b_3 = [y=\{1,2\}]_\Delta$, $b_4 = [y=\varnothing]_\Delta$.

We are in the case of figure 2 where y is not a bijection and $\Delta = \Delta_2$; hence in this case $\mathfrak{a} <\neq> \mathfrak{B}$ $\mathfrak{B} <=> \mathcal{C}$; the set of complexes $\mathcal{C}$ is defined by $\mathcal{C} = \{c_i\}_{i=1,4}$ with $c_1 = [X=1]$, $c_2=[X=2]$, $c_3=[X=\{1,2\}]$ and $c_4=[X=\varnothing]$.

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 $a_2^c = [y=\varnothing]$. In figure 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 $C = \{c_1,c_2,c_3,c_4\}$.

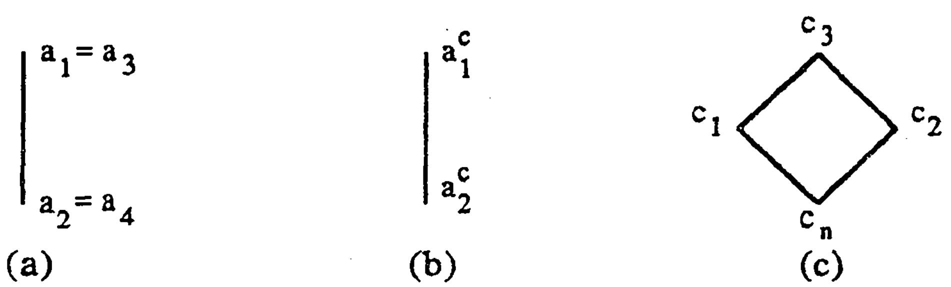

Figure 3 : (a), (b), (c) represent respectively the lattice of $\mathfrak{a}$, $\mathfrak{a}_c$ and C.
In (a) we represent the order $a_1=a_4 < a_1=a_3$; in (b) the order $a_2^c < a_1^c$ and in (c) $c_4 < c_1$, $c_4 < c_2$, $c_1 < c_3$, $c_2 < c_3$.

2.6. Choice of the knowledge base for a symbolic data analysis

We have seen in 2.1 that a knowledge base is a pair (W,X) where W is a set of subsets of Ω and $X = \mathfrak{a}$ or $\mathfrak{B}$ or C (which is in bijection with $D = P(O_1) \times ... \times P(O_p)$); so, a natural question is to ask in which case we have to use $\mathfrak{a}$, $\mathfrak{B}$ or C, in practice.

If we wish to take account only of the set of descriptions $D=\Delta_1$ then, the best choice to make is X=B ; this happens for instance, when the descriptions of subsets Ω' of Ω (i.e. $\Omega' \in W$) have constraints and don't depend on any given sample of Ω ; this kind of knowledge base is used when we wish to study species in biology, scenary of accidents in transportation, teams in a company (each species, scenary or team, is then an element of W), independently from any sample set.

If we wish to study a set W described by $D = \Delta_2$ without constraints and independently from Ω the best choice is $X=C$. If we wish to take into account the statistical informations contained in Ω the best choice to make is $\mathfrak{a}$; moreover $\mathfrak{a}$ allows the possibility to compute more simple lattices (see the previous example in section 2.5) and distances between symbolic objects when the descriptions vary ; this case may happen for instance when several sensors give different measures on the same set Ω, or when Ω is described by variables, the value of which vary with time.

More precisely, 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 $h_\Omega : D_i = P(O_i) \rightarrow \mathfrak{a}_i$ when i varies are comparable by using a dissimilarity (for instance $s(a_1,a_2) = \Sigma\ \{|a_1(w)-a_2(w_2)|/w \in \Omega\}$ whereas the mappings $c_i \in C_i$ defined by $h_{\Delta_i} : \Delta_i \rightarrow C_i$ are not comparable when i varies.

Example :

Let be $\Omega = \{w_1,w_2,w_3,w_4\}$ a set described by two ordinal mappings $y_1 : \Omega_1 \rightarrow O_1 = \{1,2\}$ and $y_2 : \Omega \rightarrow O_2 = \{1,2,3\}$ as given in figure 4. Let be $a_1 = [y_1=1]_\Omega$ and $a_2 = [y_2=1]_\Omega$:

	y_1
w_1	1
w_2	2
w_3	2
w_4	1

	y_2
w_1	1
w_2	2
w_3	2
w_4	3

Δ^1	O_1
δ^1_1	1
δ^1_2	2

Δ^2	O_2
δ^2_1	1
δ^2_2	2
δ^2_2	3

Figure 4

Considering that O_1 and O_2 are ordered sets, we may compute $s(a_1,a_2) = \sum_{w \in \Omega} |a_1(w)-a_2(w)| =$ 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 of the statistical informations 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, P. Brito focuses on the knowledge base $(W,\mathcal{B})$ when y is not bijective and $\Delta=\Delta_2$; F. De Carvalho (1992) focuses on the knowledge base $(W,\mathfrak{a})$ when y is bijective and $\Delta=\Delta_1$. In their dissertation J. Lebbe (1991) and R. Vignes (1991) focus on (W,D) with $\Delta=\Delta_1$ and y not bijective.

3. Boolean symbolic objects

In this section, descriptions are cartesian products ; so, we have $\Delta = O_1 \times ... \times O_p = O$ and $D = P(O_1) \times ... \times P(O_p)$; let be y_i a mapping $\Omega \rightarrow O_i$ which associate to $w \in \Omega$ its value $y_i(w)$ in the domain O_i ; $y = (y_1,...,y_p)$ is a mapping $\Omega \rightarrow \Delta$ such that $y(w) = (y_1(w),...,y_p(w))$. 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^i_Ω the mapping $D_i \rightarrow \mathfrak{a}$ such that $h^i_\Omega(V_i) = e_i$ where e_i is the mapping $\Omega \rightarrow$ {true, false} such that $e_i(w)$ = true iff $y_i(w) \in V_i$. 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 term we may write $e_i(w) = [y_i(w) \in V_i]_\Omega$ where $[y_i(w) \in V_i]_\Omega$ is the logical proposition which is true iff $y_i(w) \in V$; to express the symbolic object e_i, in order to simplify notations, instead of writing $\{\forall w, e_i(w) = [y_i(w) \in V]_\Omega\}$ 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 $e_i(w)$ = true iff the color of w is red or white. When $y_i(w)$ is meaningless (e.g. the kind of computer used by a company without computers) $V_i = \phi$ and when it has a meaning but it is not known $V_i = O_i$. The extension of e_i in Ω denoted by $ext(e_i/\Omega)$ is the set of elements $w \in \Omega$ such that $e_i(w)$ = true.

3.2. Assertions

An assertion is a conjunction of events ; more precisely, it is defined by the mapping h_Ω : $D = D_1 x ... x D_p \rightarrow a$ such that if $V = (V_1, ..., V_p)$ where $V_i \subseteq O_i$ then $h_\Omega(V) = a$ such that $a(w) =$ true iff $y(w) \in V$.
In logical term we may write $a(w) = \wedge_i [y_i(w) \in V_i] = \wedge_i e_i(w)$; in conformity with the notation for an event, an assertion a is denoted $a = \wedge_i [y_i = V_i]$. For instance, if $a = [\text{color}=\{\text{red, white}\}] \wedge [\text{height}=[0, 15]]$, $a(w) =$ true iff w is red or white and its height is between 0 and 15. The extension of an assertion denoted $\text{ext}(a/\Omega)$ is the set of elements of Ω such that $\forall i$, $y_i(w) \in V_i$.

3.3. Hordes and synthesis objects

A "horde" is a symbolic object which is used when we need to describe a structure composed by 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 $\Omega^p \rightarrow \{\text{true, false}\}$, such that $h_\Omega(V) = H$ where $V = (V_1, ..., V_p)$ and $H(u) =$ true where $u = (u_1, ..., u_p)$, 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 $u_1 = u_2 = ... = u_p$ a horde becomes an assertion. The extension of H in Ω^p is $\text{Ext}(H/\Omega^p) = \{w \in \Omega^p / H(w) = \text{true}\}$.

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 neighbour.

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 (1991).

Example :

Ω is a set of mushrooms, described by their color and their lenght ; they are represented by two variables $col_t : \Omega \rightarrow O_c$ and $\ell_t : \Omega \rightarrow O_\ell$ 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 lenght, such that $O_{col} = \{1, 2\}$ and $O_\ell = \{1, 2\}$. At time t_1 anf t_2 we obtain the tables (a) and (b) given below for a set of two mushrooms $\Omega_1 = \{w_1, w_2\}$; the table (c) represents the values taken by the elements of the set of decrivable object Ω at a given time.

Ω	col $_{t_1}$	ℓ_{t_1}
w_1	1	1
w_2	2	1

Table (a)

Ω	col $_{t_2}$	ℓ_{t_2}
w_1	2	1
w_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)

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

$a_{t_1} = [\text{col}_{t_1} = 1] \wedge [\ell_{t_1} = 1, 2]$; $a_{t_2} = [\text{col}_{t_2} = 1] \wedge [\ell_{t_2} = 1, 2]$;

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

By definition, a_{t_1} and a_{t_2} are mappings $\Omega \rightarrow \{\text{true, false}\}$ such that

$a_{t_1}(w_1) = [\text{col}_{t_1}(w_1) \in \{1\}] \wedge [\ell_{t_1}(w_1) \in \{1, 2\}] = \text{true}$; similarly way we obtain

$a_{t_1}(w_2) = \text{false}$, $a_{t_2}(w_1) = \text{false}$, $a_{t_2}(w_2) = \text{true}$, $c(x_1) = c(x_2) = \text{true}$

$c(x_2) = c(x_1) = \text{false}$.

It results that $\text{ext}(a_{t_1}/\Omega) = \{w_1\}$; $\text{ext}(a_{t_2}/\Omega) = \{w_2\}$ and $\text{ext}(c/O) = \{x_1, x_3\}$.

We may also define three hordes as follows :

$h_1 = [\text{col}_{t_1}(u_1) = 1] \wedge [\ell_{t_2}(u_2) = 1, 2]$,

$h_2 = [\text{col}_{t_2}(u_1) = 1] \wedge [\ell_{t_2}(u_2) = 1, 2]$ where $u_i \in \Omega$;

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

Therefore it is easy to see that $\text{Ext}(h_1/\Omega) = \{(w_1, w_1), (w_1, w_2)\}$, $\text{Ext}(h_2/\Omega) = \{(w_2, w_1), (w_2, w_2)\}$; $\text{Ext}(c, O) = \{(x_1, x_1), (x_1, x_2), (x_1, x_3), (x_1, x_4), (x_3, x_1), (x_3, x_2), (x_3, x_3), (x_4, x_4)\}$.

4. Modal objects

4.1. Internal and external modal objects and their extension

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* [weight = [300,500]] and f_2 = [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 (1990). 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 = {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 (or sometimes, more generally from $(O_i, P(O_i))$ where $P(O_i)$ is the power set of O_i) in M^x, for instance O_i = {red, yellow, green}, $M^x = [0,1]$ and q_i^j (red) = 0.1 ; q_i^j (yellow) = 0.3 ; q_i^j (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 §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 w is red then $y_i(w) = r$ 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 $Q_i^j \subseteq Q_i$

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

$q_i^1 \cap_x q_i^2 = q_i^1 q_i^2$ where $q_i^1 q_i^2 (v) = q_i^1 (v)\ q_i^2 (v)$; $c_x (q_i) = 1-q_i$

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 (1960) and recalled by Dubois and Prade (1988). In 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 $g_x (q_i^1 , q_i^2) = <q_i^1 , q_i^2>$ a scalar product, for instance $<q_i^1, q_i^2> = \Sigma \{q_i^1 (v), q_i^2 (v)/v \in O_i\}$

• f_x is an "aggregation" mapping from $P(L^x)$, the power set, of L^x in L^x .

The aim of f_x is to allow the computation of "x-conjunctions" denoted by $\wedge_x$; hence we have $f_x (\{L_1,L_2\}) = L_1 \wedge_x L_2$ where $L_1, L_2 \in [0,1]$; coherent choices with the conjunction $\wedge$ of classical logique are for instance $f_x (\{L_1,L_2\}) = \text{Min} (L_1,L_2)$ or $f_x (\{L_1,L_2\}) = L_1 . L_2$, as in these cases when $L_1,L_2 \in \{0,1\}$ we get $f_x(\{1,1\}) = 1$ and $f_x(\{1,0\}) = f_x(\{0,0\}) = 0$; it is also possible to define a "x-disjunction" denoted v_x, in this case we have $f'_x(\{L_1,L_2\}) = L_1 v_x L_2$ where coherent choices with the classical disjunction $\vee$ are for instance $f'_x(\{L_1,L_2\}) = \text{Max}(L_1, L_2)$ or $f'_x (\{L_1,L_2\}) = L_1 + L_2 - L_1 . L_2$.

By chosing $L_1 \oplus_x L_2 = \frac{1}{2} (L_1 \wedge_x L_2 + L_1 v_x L_2)$ we define a mean operator (between f_x and f'_x) denoted $\oplus_x$, which takes the values $1 \oplus_x 1 = 1$, $1 \oplus_x 0 = 0 \oplus_x 1 = \frac{1}{2}$ $0 \oplus_x 0 = 0$; this operator will be chosen in the case of probabilist objects and denoted $\wedge_x$ in order to avoid new denotations and so we will have $f_x (L_1,L_2) = \text{Mean} (L_1,L_2)$. The mean choice may also be considered as a numerical operator which fits as well as possible L_1 and L_2, in other words it is obtained by optimizing a criterion as we have $f_x (L_1,L_2) = \arg\min_x ((L_1 - x)^2 + ((L_2 - x)^2)$; other fitting criterion may be chosen as the median for instance, which is given by $f_x (L_1,L_2) = \arg\min_x (| L_1 - x | + | L_2 - x |)$.

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 $D = P(Q_1^x) x ... x\ P(Q_p^x)$ and $h(d) = a$ where $d = (\{q_1^j\}_j, ..., \{q_p^j\}_j)$ and a is an im assertion defined as follows :

Definition of an im assertion

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

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

By convention, in all this paper an event $[y_i = \{q_i^j\}_j]$ may also be denoted $[y_i = q_i^1, q_i^2 ...]$; Notice also that it results from the definition that $[y_i = \{q_i^j\}_j]$ is equivalent to the event $[y_i = \cup_{x\,j} q_i^j]$; in other words, by using the preceding denotation, 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 $a_j = \wedge_{x\,i} [y_i = q_i^1]$ is defined by $a_1 \cup_x a_2 = \wedge_{x\,i} [y_i = q_i^1 \cup_x q_i^2]$; more generally we have $\cup_{x\,j}\ a_j = \wedge_{x\,i} [y_i = \cup_{x\,j}\ q_i^j]$; hence, it results with our convention that $\cup_{x\,j}\ a_j = \wedge_{x\,i} [y_i = \{q_i^j\}_j]$. The intersection of assertions is defined similarly : $\cap_{x\,j}\ a_j = \wedge_{x\,i} [y_i = \cap_{x\,j}\ q_i^j]$. The operators OP_x extended on a_x will be studied in greater depth in § 9.

We may also combine im assertions a_i by standard logic operators by setting $a_x = \wedge_i a_i$ where $a_x : \Omega \to \{\text{true, false}\}$ is such that $a_\alpha(w) = \wedge_i [a_i(w) \geq \alpha]$ be true iff the proposition $[a_i(w) \geq \alpha]$ is true for any i ; by the same way standard disjunctions and negation may be defined : $a_\alpha = \vee_i a_i$ means that $a_\alpha(w) = \text{true}$ iff $\exists$ i : $a_i(w) \geq \alpha$ and $a_\alpha = \Psi\ a_i$ means that $a_\alpha(w) = \text{true}$ iff $a_i(w) < \alpha$. By this way, a_α is a boolean assertion as it is a mapping $\Omega \to [\text{true,false}]$; another way consists of setting : $a(w) = \text{Max}\{\alpha / a(w) \geq \alpha\}$; by this way it results that $\wedge_i a_i(w) = \text{Min}_i\ a_i(w)$ and $\vee_i a_i(w) = \text{Max}_i\ a_i(w)$.

Definition of the extension of an im object

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

There are many ways of choosing of the threshold α ; for instance $\alpha_{min} = \underset{i}{\mathrm{Min}}\ \underset{v}{\mathrm{Min}}\ q_i(v)$ or $\alpha_{max} = \underset{i}{\mathrm{Max}}\ \underset{v}{\mathrm{Max}}\ q_i(v)$; in the second case, the elements of the extension may be called "prototypes" (see the Example : given in 15). Notice that it is possible to compute the extension of a standard logic combination of modal assertions ; for instance, let be $e_i = [y_i = q_i]$ where $q_i \in Q_i^x$ and $a_\alpha = (e_1 \wedge_{x_1} e_2) \vee (\neg(e_3 \wedge_{x_2} e_4)) \wedge e_5$; by setting $a_1 = e_1 \wedge_{x_1} e_2$ and $a_2 = e_3 \wedge_{x_2} e_4$ we get $a_\alpha = a_1 \vee \neg a_2 \wedge e_5$then, we may say that the extension of a at level α in Ω is $\mathrm{Ext}(a/\Omega,\alpha) = \{w \in \Omega / b_1(w) \geq \alpha$ or $b_2(\alpha) < \alpha$ and $e_5(\alpha) \geq \alpha\}$.

Many other kinds of operators may be based on given tables coming from the background knowledge; an example of such operators is given in 4.4.

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(w) 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 $a = \wedge_{x_i} [y_i = \{q_i^j\}_j]$; 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 $w \in \Omega$ by $\varphi(w) = w^s = \wedge_i\ [y_i = y_i\ (w)]$ where $y_i\ (w) = \{r_i^j\}_j$ 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 $w \in \Omega$ is a manufactered 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 be obtained from the description of a typical rugby ball of "Star's team" given by several experts and w^s is obtained from the description given by the same experts on my own ball be defined by :

a = [y_1 = *a little* flat, *quite* rounded] $\wedge_i$ [y_2 = *a little* heavy]

w^s = [y_1 = *quite* rounded] $\wedge_i$ [y_2 = *very* heavy, *quite* heavy].

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

The problem is to know if it is acceptable to say that w 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

$r_2^1 \cup_i r_2^2\ (v)$ = 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 $f(\{L_j\})$ = Min L_j and $L_1 < L_2 < L_3$, we obtain

$a(w) = f_i\ (g_i\ (q_1^1, r_1^1)\ ,\ g_i\ (q_2^1, r_2^1 \cup_i r_2^2)) = f_i$ (acceptable, not acceptable) = not acceptable.

In this case, a determination in a sense close to that given by Descles (1991) is to write the sentence :

Name (w^s) is Name ($a(w)$) Name (a) = "my own ball" is "a not acceptable" rugby ball of the star team".

More precisely, if we focus on a given feature y_i we have the following kind of sentence :

The name (y_i) of Name (w^s) is Name ($a(w)$) Name (y_i) for a Name (a) ; for instance if we settle Name (y_1) = "pattern" we get :

The "pattern" of "my own ball" is "an acceptable" "pattern" for a "rugby ball of the Star team".

By setting Name (y_2) = "wheight" we obtain a similar sentence :

The "weight" of "my own ball" is "a not acceptable" "weight" for a "rugby ball of the star team".

4.5. The case of conjonction of events concerning the same description

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 $a_i = \wedge x_\ell [y_i = q_i^\ell]$; in this case, it is necessary to introduce a third mapping h from $P(L^x)$ in L^x such that $a_i(w) = h(\{g(q_i^\ell, r_i)\}_\ell)$; hence, more generally, if $a = \wedge x_i\, a_i = \wedge x_i \wedge x_\ell [y_i = q_i^\ell]$ then $a(w) = f_x(\{a_i(w)\}_i) = f_x(\{h_x(\{g_x(q_i^\ell, r_i)\}_\ell)\}_i)$; for instance, in § 10.5 the choice h_{pr} = min has been made. The following example may be omitted in a first lecture, its aim is to built an assertion a_i formed by the conjonction of the events for which extension at level $\frac{1}{2}$ contains a given $w \in \Omega$.

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 $w \in \Omega$ is described by $w^s = [y_i = r]$ is such that $r(v_1) = r(v_2) = \frac{1}{2}$; let $a = \wedge x_i\, e_i$; the set of im events $e_i = [y = q_i]$ such that $a(w) \geq \frac{1}{2}$ is defined by the set of probability measures q_i which satisfy the inequality $e_i(w) = f_x(g_x(q_i, r)) \geq \frac{1}{2}$; if f_x is the mean and g_x is the scalar product we get $e_i(w) = \text{Mean}(\{<q_i, r>\}) = <q_i, r>$ as there is only one variable. Hence q_i has to satisfy the following inequality :

$e_i(w) = <q_i, r> = q_i(v_1)\, r(v_1) + q_i(v_2)\, r(v_2) \geq \frac{1}{2}$ which is equivalent to $\frac{1}{2}\, q_i(v_1) + \frac{1}{2}\, q_i(v_2) \geq \frac{1}{2}$, 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 $a = \wedge x_\ell \{e_i^\ell / e_i^\ell(w) \geq \frac{1}{2}\}$ then $a(w) = h_x(\{e_i^\ell(w)\}_\ell)$; if h_x = Min then $a(w) = \text{Min}(\{e_i^\ell(w)\}_\ell) = \frac{1}{2}$.

5. Possibilist objects

5.1. The possibilist approach

Here we follow Dubois and Prade (1988) in giving the main idea of this approach.

Definition of a measure of possibility and of necessity

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) $\forall A, B \subseteq \Omega \qquad \Pi(A \cup B) = Max(\Pi(A), \Pi(B))$

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

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

The following properties may then be shown :

$N(\phi) = 0$; $N(A \cap B) = \mathrm{Min}(N(A), N(B))$; $\Pi(\cup_i A_i) = \mathrm{Max}_i(\Pi(A_i))$;

$N(\cap_i A_i) = \mathrm{Min}_i(N(A_i))$; $\Pi(A) \leq \Pi(B)$ if $A \subseteq B$; $\mathrm{Max}(\Pi(A), \Pi(\bar{A})) = 1$;

$\mathrm{Min}(N(A), N(\bar{A})) = 0$; $\Pi(A) \geq N(A)$; $N(A) > 0$ implies $\Pi(A) = 1$;

$\Pi(A) < 1$ implies $N(A) = 0$; $\Pi(A) + \Pi(\bar{A}) \geq 1$ and $N(A) + N(\bar{A}) \leq 1$.

Example :

We define $\Pi_E(A)$ (resp. $N_E(A)$) as the possibility (resp. the necessity) that $w \in A$ when $w \in E$. 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 valuated 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 $\Pi(\{200\} \cup_p \{250\}) = \mathrm{Max}(\{200\}, \{250\}) = 0.8$.

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.*

. $OP_p : \forall i, q_i^1, q_i^2 \in Q_i \quad q_i^1 \cup_p q_i^2 = Max(q_i^1, q_i^2)\,;\, q_i^1 \cap_p q_i^2 = Min(q_i^1, q_i^2);$

$c_p(q) = 1 - q$ *denoted also* $\bar{q}$.

. $g_p : g_p(q_i^1, q_i^2) = sup\{Min(q_i^1(v), q_i^2(v)) \,/\, v \in O_i\}$

. $f_p : \forall L \subseteq [0, 1], f_p(L) = Max(\ell \,/\, \ell \in L)$

Notice that OP_p is defined as in fuzzy sets and g_p has also been proposed by Zadeh (1971). 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 $a_n(w) = 1 - f_p(\{g_p(\bar{q}_i, r_i)\}_i)$ and then

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

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

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 , $g_n(q_i, r_i) = inf\{Max(q_i(v), \bar{r}_i(v)) \,/\, v \in O_i\}$ and $f_n = Min$.

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 w is defined by w^s = [height = close to 16].

The question is to find the possibility and necessity of w knowing e_p, in the case where e_p and w^s may be written : e_p = [height = q_1, q_2] and w^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 w by

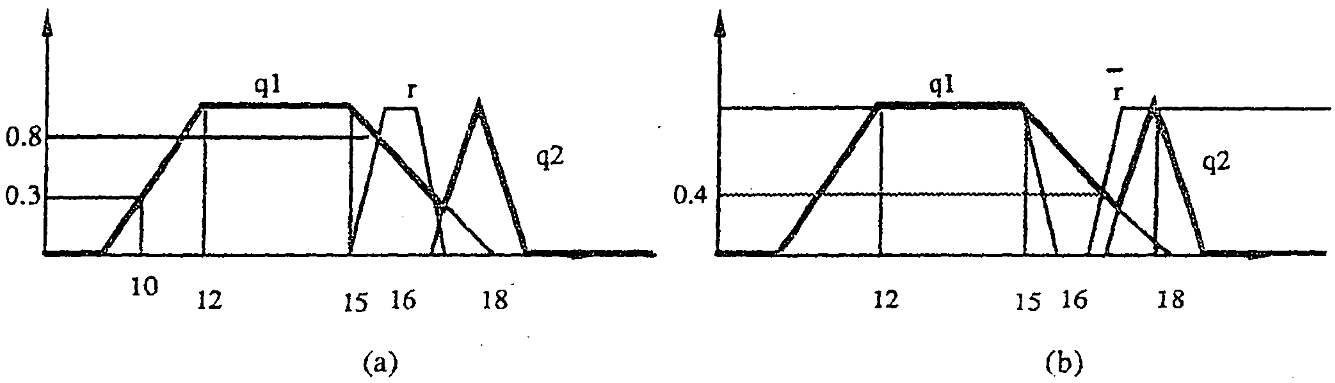

Figure 5
(a) $q_1 \cup q_2 = \text{Max}(q_1, q_2)$
(b) $\bar{r}_i = 1 - r_i$

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

The necessity of w is given by :
$e_n(w) = g_n(q_1 \cup_p q_2, (r_1) = \inf\{\text{Max}(q_1 \cup_p q_2(v), \bar{r}_1(v)) / v \in O\} = 0.4.$

This kind of information may be described by a kind of "determination" (Desclés (1991)) of e_p or e_n given by the following sentence :

Name (w^s) is a Name (e_x (w)) Name (e_x).

More precisely, let e_p and e_n be descriptions of typical houses of my village and w^s the house that Mr Dupont wishes to build ; so, we may write Name (e_p) = Name (e_n) = "house of my village" and Name (w^s) = "The house of Mr Dupont" ; if we also say that Name ($e_p(w)$) = "possible" when $e_p(w) \geq 0.8$ and Name (e_n (w)) = "not necessary" when $e_n(w) \leq 0.5$, we get : "The house of Mr Dupont" is a "possible" "house of my village" ; in case of necessity, we obtain : "The house of Mr Dupont" is a "not necessary" "house of my village".

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 $a = \hat{\uparrow}_i [y_i = V_i]$ is an im object $a_b = \hat{\uparrow}_i [y_i = q_i]$ 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 $w^s = \hat{\uparrow}_i [y_i = r_i]$ where r_i is the characteristic mapping of $y_i(w) \subseteq O_i$, (there is doubt if $y_i(w)$ is not reduced to a single element), it is then easy to show that in the possibilist choice $y_i(w) \cap V_i \neq \phi \Leftrightarrow a_b(w) = 1$ and in the necessitist choice $y_i(w) \subseteq V_i \Leftrightarrow a_b(w) = 1$.

If we denote $|a|_\Omega$ the set of elements of Ω such that $a(w) = \mathrm{true}$, we have $|a|_\Omega = \mathrm{Ext}(a_b / \Omega, \alpha)$ $\forall \alpha \in]0,1]$, for both choices.

6. Probabilist objects

6.1. The probabilist approach

First we recall the well known axioms of Kolmogorov :
If $C(O_i)$ is a σ-algebra on O_i (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 $(O_i, C(O_i))$ if

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

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 probability distributions defined on $(O_i, C(O_i))$. We suppose that the $w^s = \hat{\uparrow}_i [y_i = y_i(w)]$ are such that $y_i(w) \in Q_i$. We recall that Q_i^X has been defined in 4.1.

6.2. A formal definition of probabilist objects

There are many ways of defining probabilist objects (see Diday [1993]) where two alternatives are given) ; in the one defined here, any element of Ω represents an entity which is of higher level unit (for instance : a dice, a species of animals, a scenario of accidents etc.) then the classical samples used statistics ; from these entities it is possible to induce probability measures on $(O_i, C(O_i))$ associated to each variable y_i. Informally, if $w \in \Omega$, its associated "probabilist object" (below, the exact definition is given) will be denoted $w^s = \hat{}_{i\,pr}\ [y_i = r_i]$ where r_i is the probability density of a for y_i ; this probability is obtained objectively from a classical sample or subjectively from the knowledge of an expert. A probabilist assertion represents a subset $\Omega' \subseteq \Omega$; informally, it is denoted $a = \hat{}_{i\,pr}\ [y_i = q_i]$ where q_i is a mapping : $O_i \to [0,1]$ such that $q_i(v)$ is the probability density that v occurs when simultaneously, an element associated to each entity of Ω' is trialed independently. Notice that only when Ω' is reduced to a single element, q_i is a probability density as the following example shows.

Example :

Ω is a set of dices and w_i is a dice represented by the probabilist assertion $w^s = [y_1 = r_i]$ where $r_i : O = \{1, 2, \ldots, 6\} \to [0,1]$ is a probability density such that $r_1(v)$ is the probability that $v \in O$ occurs when the dice w is trialed. If $a = [y_1 = q_1]$ represents a class of five dice, $q_1 : O \to [0,1]$ is such that $q_1(v)$ is the probability that v occurs when the five dices are trialed simultaneously and independently. As $q_1(v) \in [0,1]$, it results that $\sum_{v \in O} q_1(v) \in]1,6]$ and therefore q_i is not a probability.

Let us give now the formal definition of a probabilist assertion concerning this paper.

Definition

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

$OP_{pr} : \forall q_i^1, q_i^2 \in Q_i \quad q_i^1 \cup_{pr} q_i^2 = q_i^1 + q_i^2 - q_i^1 q_i^2 ; q_i^1 \cap_{pr} q_i^2 = q_i^1 q_i^2$ *which is the mapping which associate to* $v \in O_i$, $q_i^1(v)\, q_i^2(v)$; $c_{pr}(q) = \bar{q} = 1 - q$.

$g_{pr} : \forall \{q_i, r_i\} \in Q_i^{pr} \times Q_i \quad g_{pr}(q_i, r_i) = < q_i, r_i > = \Sigma\, \{q_i(v)\, r_i)(v)\, q_i^2(v) / v \in O_i\})$.

$f_{pr} : f_{pr}(\{L_i\}) =$ *mean of the* L_i.

Coming back to the preceding example of dices, where $a = [y_1 = q_1]$ and w is represented by $w^s = [y_1 = r_1]$ we get from this definition $a(w) = g_{pr}(q_1, r_1) = \Sigma\, \{q_1(v)\, r_1(v) / v \in O_i\}$ which is the sum of the probabilities that v occurs simultaneously in a trial of the five dices represented by a.

As $q_i(v) \le 1$ and $r_1(v)$ is a density of probability we always have $a(w) \le 1$, because $g_{pr}(q_i, r_i) \le \sum_v r_i(v) = 1$; however, a (w) is generally not a probability but just a sum of probabilities computed on events which are compatible, (the only case where it is a probability, happens when $\forall v \in O_i, q_i(v) \in \{0,1\}$) ; nevertheless, as the probabilities used to obtain a (w) are always computed on the same events (the elements of the O_i), it results that when w varies in Ω the a(w) may be compared. We may say that a (w) measures how much w is in the extension of a. How can a (w) be called? as the word "prototype" is now widely used in cognitive science (see Rosh (1978) Descles (1986), Dubois (1992)),we may say that a (w) measures the degree to which w is a prototype for the class of entities that a represents ; hence, if a (w) is high we may say that w is a "probabilist prototype" and the contrary if a(w) is low ; we may also say that a(w) measures a "probabilist typicality" of w to the class of entities Ω' represented by a (see J. Lebbe and R. Vignes (1991) for the use of typicalities) ; more simply, we may also say that a(w) is a measure of ressemblance of w to Ω'.

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 set of variables are given by the expert to describe a symbolic object, propagation technics (see Pearl (1988)), Lauritzen & Spiegelhalter (1988) may be used which induce other mappings g_{pr} and f_{pr}. Notice also that we could use a mixture decomposition law, in order to compute the union by setting : $q_i^1 \cup_{pr} q_i^2 = p_1 q_i^1 + p_2 q_i^2 - p_1 p_2 q_i^1 q_i^2$ with $p_1, p_2 \in [0,1] : p_1 + p_2 = 1$, but in this case $q_i^1 \cup_{pr} q_i^2$ is not a generalisation as we lose the following property : $q_i^1 \cup_{pr} q_i^2 \geq q_i^j$.

In order 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 trialed independently. This comes from the fact that if (X_1, X_2) is a pair of random variable $\Omega \to O_i \times O_i$ where $O_i = \{1,2,...,6\}$ with probability (q_i^1, q_i^2), then the probability that the number j occurs in both dices trialed independently is $q_i^1 \cap_{pr} q_i^2 (j) = Pr((X_1, X_2) = (j, O_i) \cap (O_i,j)) = Pr((X_1, X_2) = (j,j)) = Pr(X_1 = j) Pr(X_2=j) = q_i^1(j)\, q_i^2(j)$; 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 also that if q_i^1 and $q_i^2 \in Q_i^{pr}$ then $\Sigma_v\, q_i^1 \cup_{pr} q_i^2 (v)$, $\Sigma_v\, q_i^1 \cap_{pr} q_i^2 (v)$ and

$\sum_v c\,(q_i^j)\,(v)$ belong in [0,card O_i] as $q_i^j\,(v) \in$ [0,card O_i] even if $q_i^1 \cup_{pr} q_i^2\,(v)$, $q_i^1 \cap_{pr} q_i^2\,(v)$ and $c\,(q_i^j)\,(v)$ are probabilities.

Hence, it results 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 $\cup_{pr}\, q_i^2(v)$ on O_i is larger than 1. Also, $q_i^1 \cup_{pr} q_i^2$ 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 $Q_i^x \times Q_i$ and not on $Q_i^x \times Q_i^x$ 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 (this will be done in section 9.1).

Example :

A stone w is described by its color y_1 (w) which may be red or blue and its roundness y_2(w) which may be round or flat.

Let $a = [y_1 = q_1^1, q_1^2] \wedge_{pr} [y_2 = q_2]$ (which is a typical description of a stone from my garden) and $w^s = [y_1 = r_1] \wedge_{pr} [y_2 = r_2]$ 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 (red) = 0.9 + 0.5 - 0.9 x 0.5 = 0.95 ;

q_1^3 (blue) = 0.1 +0.5 - 0.1 x 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

$a\,(w) = g_{pr}\,(q_1^3, r1) \wedge_{pr} g_{pr}\,(q_2, r_2)$

$= (0.95 \times 1 + 0.55 \times 0) \wedge_{pr} (0.2 \times 1 + 0.8 \times 0)$

$= 0.95 \wedge_{pr} 0.20 = \frac{1}{2}\,(0.95 + 0.20) = 0.57.$

As in the case of intensities or possibilities, it is also possible in the case of probabilist objects to describe the result by sentences which express "determinations" of the concept represented by a symbolic object. These sentences take the general form : "Name (w^s) is Name ($a(w^s)$) Name (a)". More precisely, by using the preceding example, if we set : Name (a) = "Stone from my garden", Names (w^s) = "the stone found by my son" and $a(w^s)$ = {a not probabilist prototype}, if a (w^s) <0.6} we get the sentence : "The stone found by my son" is "a not probabilist prototype" "stone from my garden".

7. Belief objects

7.1. The belief functions formalism

At the origine of this theory we may mention at least the work of Choquet (1953) on "Capacities of order 2" and Dempster (1967) on "upper and lower probabilities induced by a multivalued mapping". The basic notions of this formalism are in Schafer's book (1976) : "A mathematical theory of evidence" which is "still a standard reference for this theory" Schafer (1990). First a "probability assignment" function m from P (Ω) (the power set of Ω, supposed finite) in [0,1] is defined by : Σ {m (V)/V $\in$ P(Ω)} = 1 and m (ϕ) = 0 ; then a belief function Bel : P(Ω) $\rightarrow$ [0,1] is defined by :

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

A "body of evidence" is viewed as a pair ($\mathcal{F}$,m) where m is a probability assignment function and $\mathcal{F}$ = {V $\in$ P(Ω)/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(Ω) $\rightarrow$ [0,1] such that :

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

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

It may be proved (Schafer (1976)) that we have the following properties : Bel is a belief function iff :

i) Bel (Ω) = 1

ii) Bel (ϕ) = 0

iii) $\mathrm{Bel}\,(A_1 \cup \ldots \cup A_n) \geq \sum_i \mathrm{Bel}\,(A_i) - \sum_{i<j} \mathrm{Bel}\,(A_i \cap A_j) + \ldots =$

$$\sum_{\substack{I \subseteq \{1,\ldots,n\} \\ I \neq \phi}} (-1)^{|I|+1}\, \mathrm{Bel}\,\Big(\bigcap_{i \in I} A_i\Big),$$ where $|I|$ denotes the cardinality of I.

As a consequence of iii) we get :

$$Pl\,(A_1 \cap ... \cap A_n) \leq \sum_i Pl\,(A_i) - \sum_{i<j} Pl\,(A_i \cup A_j) + ...$$

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

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

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

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

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

$$Bel\,(A/B) = \frac{Bel\,(A \cup \bar{B}) - Bel(\bar{B})}{(1 - Bel\,(\bar{B}))}$$

We have the following link with probability and possibility theories : it may be shown that if $\mathcal{F}$ contains only singletons then Bel is a classical probability measure. Dempster (1967) said that Pl and Bel may be viewed as upper and lower probabilities. Schafer (1976) has shown that if $\mathcal{F}$ contains only a nested sequence of subsets $V_1 \subseteq V_2 \subseteq ... \subseteq V_n$ then we have :

$Bel\,(A \cap B) = Min\,(Bel(A), Bel(B))$ and $Pl\,(A \cup B) = Max\,(Pl\,(A), Pl\,(B))$ 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 $\leq$ bel $\leq$ pr $\leq$ pl $\leq$ pos.

The theory of evidence models several kinds of knowledge :

i) Probability : as said by J. Pearl (1990) : "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 $V_2 = \{speed > 110\}$ and 0.10 for unknown speeds, we may represent this information by a belief function q with body of evidence ($\mathcal{F}$,m) such that
$\mathcal{F} = \{V_1, V_2, IR^+\}$, $m(V_1) = 0.40$, $m(V_2) = 0.50$, $m(IR^+) = 0.10$.
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 vehicule 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 $\{\mathcal{F}_1, m_1\}$ such that
$\mathcal{F}_1 = [V_1, IR^+]$, $m_1(V_1) = 0.90$ and q_2 defined by $\{\mathcal{F}_2, m_2\}$ such that : $\mathcal{F}_2 = \mathcal{F}_1$ 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 (1986), we define the union and intersection of two bodies of evidence ($\mathcal{F}_1$,m_1) and ($\mathcal{F}_2$,m_2) as follows :

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

$m_1 \cap_{bel} m_2 (A) = \sum_{V_1 \cap V_2 = A} m_1(V_1)\, m_2(V_2)$ which is consistent with Dempster's rule if the term $m_1 \cap m_2(\phi)$ (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 ($\mathcal{F}_i^j$, m_i^j).

Definition

A belief assertion denoted $a_{bel} = \hat{\imath}_{bel}\,[y_i = \{q_i^j\}_j]$ 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} : $\forall i, q_i^1, q_i^2 \in Q_i \quad q_i^1 \cup_{bel} q_i^2 (V) = \sum_{A \subseteq V} m_i^1 \cap_{bel} m_i^2 (A)$;

$q_i^1 \cap_{bel} q_i^2 (V) = \sum_{A \subseteq V} m_i^1 \cup_{bel} m_i^2 (A)$; the complement is defined by

$c_{bel} (q_i^j) (V) = \bar{q_i} (V) = \sum_{A \subseteq V} \bar{m}_i^j (A)$ where $\bar{m}_i^j (A) = m_i^j (\bar{A})$.

g_{bel} : $g_{bel} (q_i^1, q_i^2) = \Sigma \{m_i^1 \cap_{bel} m_i^2 (V_2) / V_2 \subseteq V_1, (V_1, V_2) \in \mathcal{F}_1 x \mathcal{F}_2\}$

f_{bel} : *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_{bel} may be more general; we have chosen the mean in order to simplify. It is also possible to define a plausibilist object by $OP_{p\ell} : q_i^1 \cup_{p\ell} q_i^2 (V) = \sum_{A \cap V \neq \phi} m_i^1 \cap_{bel} m_i^2 (A)$; $q_i^1 \cap_{p\ell} q_i^2 (V) = \sum_{A \cap V \neq \phi} m_i^1 \cup_{bel} m_i^2 (A)$ and $c_{p\ell} (q_i) = \bar{q_i}$ is defined as in the belief case.

$g_{p\ell} : g_{p\ell} (q_i^1, q_i^2) = \Sigma \{m_i^1 (V_1)\ m_i^2 (V_2) / V_1 \cap V_2 \neq \phi, (V_1, V_2) \in \mathcal{F}_1 x \mathcal{F}_2\}$ and f_{bel} 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 $q_i^1 \cap_{bel} q_i^2 (V) = \sum_{A \subseteq V} m_i^1 \cup_{bel} m_i^2 (A) = \sum_{V_1 \cup V_2 = A \subseteq V} m_i^1 (V_1)\, m_i^2 (V_2) = \sum_{V_1 \subseteq V} m_i^1 (V_1) \sum_{V_2 \subseteq V} m_i^2 (V_2)$.

We have also, $g_{bel} (q_i^1, q_i^2) = \sum_{V_1 \in \mathcal{F}1} m_i^1 (V_1)\, q_i^2 (V_1)$;

$g_{p\ell} (q_i^1, q_i^2) = \sum_{V_2 \in \mathcal{F}2} m_i^2 (V_2)\ p\ell_i^1 (V_2) = \sum_{V_1 \in \mathcal{F}1} m_i^1 (V_1)\, p\ell_i^2 (V_2)$

where $p\ell_i^j (V_j) = \sum_{V \cap V_j \neq \phi} q_i^j (V)$; 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 $\mathcal{F}_i^1 = \mathcal{F}_i^2 = \{A, O\}$, 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 $\{\mathcal{F}_1, m_1\}$ such that $\mathcal{F}_1 = \{V_1, O\}$, 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 $\{\mathcal{F}_2, m_2\}$ such that $\mathcal{F}_2 = \{V_2, O\}$. If we wish to know how much a given accident defined by w^s= [speed = q_2], satisfies the scenario defined by a, we have to compute a(w); as a is a belief object, by definition we have : $a(w) = \sum_{V \in \mathcal{F}1} m_1(V)\, q_2(V) = m_1(V_1)\, q_2(V_1) + m_1(O)\, q_2(O)$, therefore

$a(w) = m_1(V_1)\, q_2(V_1) + m_1(O)$. Hence if $V_2 \subseteq V_1$, then $a(w) = m_1(V_1)\, m_2(V_2) + m_1(O)$ and the higher the witness' belief in V_2 the more w satisfies the scenario defined by a ; if $V_1 \subseteq V_2$ then $a(w) = m_1(O)$, as $q_2(V_1) = O$ and the greater the ignorance of the expert who has defined the scenario, the more w 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 $\leq_\alpha$ on the im objects by : $a_1 \leq_\alpha a_2$ iff $\forall w \in \Omega$, $\alpha \leq a_1(w) \leq a_2(w)$.

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 $\text{Ext}_\alpha(a_1 / \Omega, \alpha) \subseteq \text{Ext}_\alpha(a_2 / \Omega, \alpha)$).

We call intension at the level α of a subset $\Omega_1 \subset \Omega$ the symbolic object b defined by a given conjunction $\wedge_y$ of events whose extension at the level α (see 4.2) contains Ω_1.
The symbolic union $a_1 \cup_{x,\alpha} a_2$ (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 union and intersection

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 then each one ; 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 an union which 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, analoguous "natural" conditions may be defined, they express the inverse conditions, for instance : the intersection of two symbolic objects is less general then 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 : $q_1 \cup_{pr} q_2 = q_1 + q_2 - q_1 q_2 \geq q_k$ for $k = 1, 2$. When q_1 and q_2 are possibilist measures we have $q_1 \cup_p q_2 = \text{Max}(q_1, q_2) \geq q_k$ 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 $\forall w \in \Omega : w^s = \wedge_i [y_i = r_i]$ we have : $a_1 \cup_x a_2 (w) = f_x (\{g_x (q_i^1 \cup_x q_i^2, r_i)\}_i)$; hence,

$$a_1 \cup_{pr} a_2 (w) = \text{Mean} \{ \sum_{v \in O_i} q_i^1 \cup_{pr} q_i^2(v)\, r_i (v)\}_i \geq \text{Mean} \{ \sum_{v \in O_i} q_i^k (v)\, r_i (v)\}_i = a_k (w)\}$$

with $k = 1, 2$.

Similarly in case of possibilities we have :

$$a_1 \cup_p a_2 (w) = \underset{i}{\text{Max}} \{\underset{v \in O_i}{\text{Max}} \text{ Min} (q_i^1 \cup_{pr} q_i^2(v), r_i (v))\} \geq \underset{i}{\text{Max}} \{\underset{v \in O_i}{\text{Max}} \text{ Min} (q_i^k (v), r_i (v))\} =$$

$a_k (w)$ with $k = 1, 2$.

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

The 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 conjonction of several events : let be $a = [y = p]$; we have by definition $a \cup_{pr} a = [y=p \cup_{pr} p] = [y = 2p - p^2]$; hence $\forall w^s = [y = r]$, we have $a \cup_{pr} a(w) =$

$\sum_{v \in O} (2p - p^2)(v)\, r(v) \geq \sum_{v \in O} p(v) r(v)$ and so $a \cup_{pr} a(w) \geq a(w)$; therefore $a \cup_{pr} a \geq a$.

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 $a \cup_p a = [y = q \cup_p q] = [y = \text{Max}(q, q)] = a$.

The conditions c) and e) depend on the chosen similarity ; with the similarity proposed in § 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 $w^s = [y = r]$ where r is a probability measure, $\cup_{pr\, i}\, a_i(w) = 1$. 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 $w^s = [y = r]$ where p is a measure of possibility $\cup_{p\, i}\, a_i(w) = 1$; 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

As in the boolean case, see Diday (1989), Brito, Diday (1990), 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 doesn't describe the dogs in a complete way, since I am not saying that they are white. It results from this definition that the symbolic union or intersection of symbolic objects is complete.

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.

It may be shown, see Diday (1992) for instance, that given a level α, the set of boolean objects is a lattice for the symbolic order and that, in this case, the symbolic union and intersection define the supremum and infimum of any couple. In order to show that the set of im objects is a lattice for te symbolic order f_x, g_x and h_x (see § 3.1) have to be well chosen and we introduce a "full" and an "empty" symbolic object (which could be also called "top" and "bottom") because they are the most and the less general symbolic object denoted Ω^s and s such that $\forall$ $w \in \Omega$, $\Omega^s(w) = 1$ and $s(w) = 0$; 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 s contains no one (e.g. it is "empty"), at the level $\alpha : 0 < \alpha \leq 1$.

Let be S the set defined by $S = \{\Omega^s, \phi^s, \{\underset{i}{\wedge_y}\, a_i, \underset{i}{\vee_y}\, a_i \,/\, a_i \in \mathbf{a}_x\}$ where $\mathbf{a}_x$ is a set of im objects (see section 4.1) and $\wedge_y$ is a given kind of "conjunction". From the definition of a lattice the supremum and infimum of any pair $(s_1,s_2) \in S \times S$ associated to the symbolic order $\leq_\alpha$ defined on S is given by :

$\mathrm{supr}(s_1,s_2) = \inf\{s \in S \,/\, s \geq_\alpha s_1 \text{ and } s \geq_\alpha s_2\}$
$\mathrm{infi}(s_1,s_2) = \sup\{s \in S \,/\, s \leq_\alpha s_1 \text{ and } s \leq_\alpha s_2\}$.

Hence, by setting that $\mathrm{Max}_\alpha(s_1,s_2)$ is defined by the following equivalence class, associated to the equivalence relation R (see 8.1) :
$\mathrm{Max}_\alpha(s_1,s_2) = \{s : \Omega \to [0,1] \,/\, \alpha \leq s(w) = \mathrm{Max}(s_1(w), s_2(w))\}$,
it results that $\mathrm{supr}(s_1,s_2) = \mathrm{Max}_\alpha(s_1,s_2)$ and $\mathrm{infi}(s_1,s_2) = \mathrm{Min}_\alpha(s_1,s_2)$. Therefore we have the following result :

Proposition

A necessary and sufficient conditions for S to be a lattice associated to the symbolic order $\leq_\alpha$, is that $\mathrm{Max}_\alpha(s_1,s_2) \in S$ and $\mathrm{Min}_\alpha(s_1,s_2) \in S$.

From this proposition, it results that the following choice of $\wedge_y$ and $\vee_y$ (denoted $\wedge_\varrho$ and $\vee_\varrho$) insure that S is a lattice for the symbolic order : $s_1 \wedge_\varrho s_2 = \mathrm{Min}(s_1,s_2)$ and $s_1 \vee_\varrho s_2 = \mathrm{Max}(s_1,s_2)$. Notice, that the symbolic union and intersection of any pair $(s_1,s_2) \in S$ associated to the conjuction $\wedge_y$ define also the supremum and infimum of this pair, as we have :

$s_1 \cup_{x,\alpha} s_2 = \wedge_\varrho \{s \in S \,/\, s \geq_\alpha s_1, s \geq_\alpha s_2\} = \mathrm{Max}(s_1,s_2) = s_1 \vee_\varrho s_2$.
$s_1 \cap_{x,\alpha} s_2 = \wedge_\varrho \{s \in S \,/\, s \leq_\alpha s_1, s \leq_\alpha s_2\} = \mathrm{Min}(s_1,s_2) = s_1 \wedge_\varrho s_2$.

(it is easy to show, that we have also $\vee_\varrho \{s \in S \,/\, s \leq_\alpha s_1, s \leq_\alpha s_2\} = \mathrm{Min}(s_1,s_2)$).

In other words, the equivalence class of R which contains the symbolic object $s_1 \vee_\varrho s_2 \in S$ contains also the symbolic object $s_1 \cup_{x,\alpha} s_2 = \wedge_\varrho \{s \in S \,/\, s \geq_\alpha s_1, s \geq_\alpha s_2) \in S$ which is by definition a complete object associated to the conjunction $\wedge_\varrho$. If we denote S_c teh subset of complete objects of S, we may say that S_c associated to $\leq_\alpha$ is a lattice of complete symbolic objects, where the supremum and infimum of any pair of elements of S_2 is their symbolic union and symbolic intersection.

Notice that if we choose $\mathfrak{a}_x = \mathfrak{a}_{pos}$ (the set of possibilist objects), $\wedge_y$ and $\vee_y$ such that $s_1 \wedge_y s_2 = Min(s_1,s_2)$ and $s_1 \vee_y s_2 = s_1 \wedge_{pos} s_2 = Max(s_1,s_2)$, from the definition of possibilist objects ; it results from the preceding proposition, that $S = \{\Omega^s, \varnothing^s, \mathfrak{a}_{pos}\}$ is a lattice for the symbolic order.

In the case of boolean objects (where $0<\alpha\leq1$), it is easy to see that the necessary and sufficient conditions of the proposition, are satisfied by setting $\wedge_y = \wedge$, $\vee_y = \vee$ as $s_1 \wedge s_2 = Min(s_1,s_2)$ and $s_1 \vee s_2 = Max(s_1,s_2)$.

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

9.1. Dual assertions

In this paper, our aim is to extend an im assertion $a = \hat{\wedge}_{i}^{x} [y_i = q_i]$ (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. In order to do so, we define the x-union or x-intersection of subsets of $\mathfrak{a}_x$ by the following definition where $*_x \in \{\cup_x, \cap_x\}$:

$\forall A_x^1, A_x^2 \subseteq \mathfrak{a}_x,\ A_x^1 *_x A_x^2 = \{a_1 *_x a_2/(a_1\, a_2) \in A_x^1 \times A_x^2\}$

and we study 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 $a^*(A_x) = \Sigma\{a^*(a_j)/a_j \in A_x\}$).

More precisely :

Given $A_x \subseteq \mathfrak{a}_x$, we have $A_x = \{a/a \in A_x\}$ and to define $A = \cup_x \{a/a \in A_x\}$ we use the set $Q_i^{Ax} \subseteq Q_i^x$ such that $Q_i^{A_x} = \{q_i/a = \hat{\wedge}_j^x [y_j=q_j] \in A_x\}$; 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 : $\cup_x \{a/a \in A_x\} = \hat{\wedge}_i^x [y_i = q_i^A]$; hence, we have

$A = \hat{\wedge}_i^x [y_i = q_i^A]$.

We define a_ℓ^* a "dual" measure of $a_\ell = \hat{\wedge}_i^x [y_i = q_i^\ell]$ by $a_\ell^*(a_j) = f_x(\{g_x(q_i^\ell, q_i^j)\}_i)$; hence, given $A_x^k \subseteq \mathfrak{a}_x$, we denote $A_k = \cup_x \{a/a \in A_x^k\}$ and we get $a_\ell^*(A_k) = f_x(\{g_x(q_i^\ell, q_i^{A_k})\}_i)$;

more generally $a^*_\ell (A_1 *_x A_2) = f_x (\{g_x (q_i^\ell, q_i^{A1} *_x q_i^{A2})\}_i)$, where $*_x \in \{\cup_x, \cap_x\}$ and $q_i^{A_k} = *_x \{q_i / q_i \in Q_i^{A_k^x}\}$. In case of probabilist objects, g_{pr} has been only defined on $Q_i^{pr} \times Q_i$, we extend it on $Q_i^{pr} \times Q_i^{pr}$ by setting :

$g_{pr}(q_i^1, q_i^2) = <q_i^1, q_i^2>$ with $<q_i^1, q_i^2> = \sum_v p_i(v)\; q_i^1(v)\, q_i^2(v)$ where $\Sigma\{p_i(v)/v \in O\} = 1$.

Hence, g_{pr} is a mapping $Q_i^{pr} \times Q_i^{pr} \rightarrow [0,1]$.

9.2. Three theorems of meta-knowledge

The three following results (Diday 1992), 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) $a^*(\mathfrak{a}_p) = 1 \qquad a^*(\phi) = 0$

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

b) In the case of probabilist objects :

Theorem 2

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

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

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 $(\mathcal{F}_i^j, m_i^j)$ 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 $V_i^1 \in \mathcal{F}_i^1$, $V_i^2 \in \mathcal{F}_i^2$ are such that : $V_i^1 \cap V_i^2 \neq \phi$). The body of evidence of two subsets A_1, A_2 of $\mathfrak{a}_{bel}$ are said to be independent iff for $\forall$i and j = 1,2 such that $Q_i^j = \cup_{bel} \{q_i^j / q_i^j \in Q_i^{Aj}\}$, the body of evidence of Q_i^1 and Q_i^2 are independent.

Theorem 3

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

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

$$a^*\Big(\bigcup_{i \in \{1,\dots,n\}} {}_{bel}\, A_i\Big) \geq \sum_{I \subseteq \{1,\dots,n\}} (-1)^{|I|+1}\, a^*\Big(\bigcap_{i \in I} {}_{bel}\, A_i\Big).$$

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

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

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

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

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

When $q_i \in Q_i$, q_i is a measure of probability and $q_i(v)$ is the probability of occurence of the value $v \in O_i$ among the possible values that a given individual may take ; if this individual is a dice, $q_i(v)$ may be obtained after enough trials of the dice ; $q_i(v)$ may also be obtained in a more subjective way by asking several experts. What is the meaning of q_i when it belongs to Q_i^{pr} and not to Q_i ? This means that q_i has been obtained by a union, intersection or complementary operator, and so, that it is not a probability ; however, each value $q_i(v)$ remains a probability whose meaning depends on the way used to obtain q_i (for instance, we may have $q_i = (q_i^1 \cup_{pr} q_i^2) \cap_{pr} (q_i^3 \cup_{pr} c_{pr}(q_i^4))$. Hence, if $q_i^1, q_i^2 \in Q_i^x$ and $q_i^1, q_i^2 \notin Q_i$, $g_{pr}(q_i^1, q_i^2)$ expresses the sum of the probabilities for $v \in O_i$ that v occurs simultaneously in the way that $q_i^1(v)$ and $q_i^2(v)$ have been obtained ; for instance, if $q_i^1(v) = a_1 \cup_{pr} b_1$ and $q_i^2(v) = a_2 \cap_{pr} b_2$ where $a_j, b_j \in Q_i$ represents the measure of probability associated to two dices called A_j, B_j, $g(q_i^1, q_i^2) = \sum_{v \in O_i} p(v)\, q_i^1(v)\, q_i^2(v)$, where $q_i^1(v)\, q_i^2(v)$ is the probability that v occurs for dices A_1 or for dice B_1 when they are trialed independantly and simultaneously in dices A_2 and B_2 when they are also trialed independantly. If $p(v) = \dfrac{1}{\mathrm{card}O_i}$, $g(q_i^1, q_i^2)$ is the mean of the $q_i^1(v)\, q_i^2(v)$.

Roughly speaking we may say that $a_1^*(a_2)$ 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

$\forall i\ g(q_i^1, q_i^2) = \sum_v p(v)\ q_i^1(v)\ q_i^2(v)$ 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 element $v \in O_i$, 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 the 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 extrem 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 § 5.3) that $a_1^*(a_2) = \mathrm{Max}_i(\sup \{\mathrm{Min}(q_i^1(v), q_i^2(v)) / v \in O_i\} = 1$ iff $\forall i\ V_i^1 \cap V_i^2 \neq \phi$ which express 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 necessistist object we have in the boolean case :
$a_1^*(a_2) = \mathrm{Min}_i(\inf \{\mathrm{Max}(q_i(v), \bar{r}_i(v) / v \in O_i\}) = 1$ 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 at 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 $O \rightarrow [0,1]$ with body of evidence $(\mathcal{F}_i , m_i)$ and $\mathcal{F}_1 = \mathcal{F}_2 = \{A,B,O\}$ with $A \cap B = \phi$; then we have :

$$a^*_1\ (a_2) = g_{bel}\ (q_1,q_2) = \sum_{V \in \mathcal{F}1} m_1(V)\ q_2(V) = m_1(A)\ m_2(A) + m_1(B)\ m_2(B) + m_1(O). \quad (1)$$

Following a classical example given by Schafer (1990), 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 $1\text{-}p=0.1$), I say that her testimony alone justify 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 then 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 induce 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

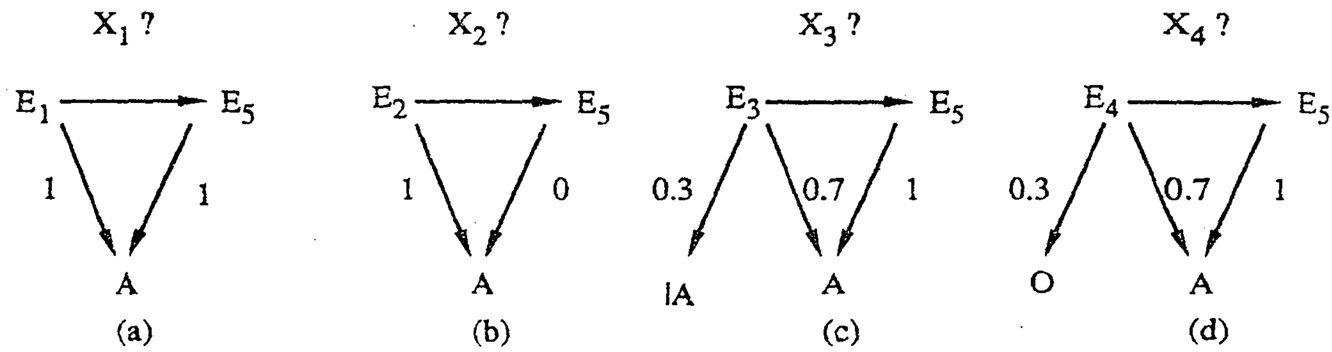

Figure 6 : $X_i = a_i^*(a_5)$ is the belief of E_i in the belief of E_5, computed according to (1).

and $\mathcal{F}_i$ the focal element of the belief function q_i, we have in situation (a) $\mathcal{F}_1 = \mathcal{F}_5 = \{A\}$ hence $m_1(A) = m_5(A) = 1$ therefore, it results from (1) $X_1 = a_1^*(a_5) = 1$; in situation (b) , $\mathcal{F}_2 = \{A\}$, $\mathcal{F}_5$ doesn't contain A and so, $a_2^*(a_5) = 0$; in situation (c), $\mathcal{F}_3 = \{A, \neg A\}$ and $\mathcal{F}_5 = \{A\}$, $m_3(A) = 0.7$, $m_3(\neg A) = 0.3$, therefore $a_3^*(a_5) = m_3(A)\, m_5(A) + m_3(\neg A)\, m_5(\neg A) + m_3(A)\, m_5(A) + m_3(O) = 0.7$; in situation (d) $\mathcal{F}_4 = \{A, O\}$, $\mathcal{F}_5 = \{A\}$, $m_4(A) = 0.7$, $m_4(O) = 0.3$, $m_5(A) = 1$, 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 $d = \frac{1}{2}$ 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 belief 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 rule (at second level) applied to the belief of belief, concerning a set of sensors, of a given sensor ; by that way the sensor represented by a_5 is rejected if $\underset{i=1,4}{\oplus} a_i (a_5) < \alpha$. The belief in A, if no sensor is rejected, is measured by the classical Dempster rule (at level 1) : $\underset{i=1,5}{\oplus} a_i (A)$.

There is an analogous theorem if a_1 is a plausibilist assertion and $a^*_1 (a_2)$ 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:
$a^*_1 (a_2) = g_{pl} (q_1,q_2) = \sum_{V \in \mathcal{F}_1} m_1(V)\, pl_2(V) = m_1(A)\,(m_2(A) + m_2(O)) + m_1(B)\,(m_2(B) + m_2(O)) + m_1(O) pl_2(O) = m_1(A)\, m_2(A) + m_1(B)\, m_2(B) + m_1(O) + m_2(O) - m_1(O)\, m_2(O)$. 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 $a^*_1 (a_2)$ 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 $a^*_1 (a_2)$ measures how much C_2 "fits" C_1; when a_1 is a plausibilist object,we may say that $a^*_1 (a_2)$ 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 $a^*_1 (a_2)=1$ 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 $a^*_1 (a_2)=0$ which measures how much C_2 fits C_1 ; when a_1 is a plausibilist object,we get $a^*_1 (a_2)=1$ 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 & al (1990) and (1991) ; for generating rules by decision graphs on im objects in the case of possibilist objects with typicalities as modes see Lebbe and Vignes (1991); for generating overlapping clusters by pyramids on symbolic objects see Brito, Diday (1990).

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 $a_\ell = \hat{\wedge}_i [y_i = q_i^\ell] \in \mathbf{a}_x$ be the set of im assertions. We denote a_ℓ^* a mapping $\mathbf{a}_x \longrightarrow [0,1]$ such that $a_\ell^*(a_k) = f_x(\{g_x(q_i^\ell, q_i^k)\}_i)$; then, we set :

(1) $s(a_\ell, a_k) = \frac{1}{2}(a_\ell^*(a_k) + a_k^*(a_\ell))/\sqrt{a_\ell^*(a_\ell)\, a_k^*(a_k)}$; 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

$a_1 = [y = 0.7\, v_1, 0.3\, v_2]$, $a_2 = [y = 0.3\, v_1, 0.7\, v_2]$; we get :

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

From this example, it results that probabilist objects do not satisfy the condition c) given in §.8.2, since if we define $a = [y = 1\, v_1, 0\, v_2] = [y = v_1]$ 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 $a_1 = [y = 1\, v_1, x v_2]$ and $a_2 = [y = x v_1, 1\, v_2]$. 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, $a_1 \cup_p a_2 = [y = 1\, v_1, 1\, v_2]$ 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 the condition c) § 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 $\Omega = \{w_1,...,w_5\}$ 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.

	w_1	w_2	w_3	w_4	w_5
y_1	-1/2	1/2	2	1	2
y_2	-1/2	1/2	1	2	2

Table T

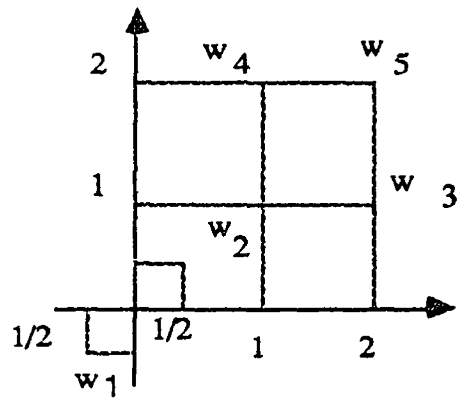

Figure 7 : Graphical representation of table T

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 $u_1^T = \frac{1}{\sqrt{2}}(1,1)$, $u_2^T = \frac{1}{\sqrt{2}}(1,-1)$. Finally we get the principal component representation given in figure 8, where the projection of w_j on the axis i is given by $F_i(w_j) = u_i^T.x_j$,where $x_j^T = (y_1(w_1) - Y_1, y_2(w_2) - Y_2)$ and $Y_i=1$, is the mean of y_i ; for instance, $F_1(w_1) = \frac{1}{\sqrt{2}}(1\ 1)\begin{pmatrix} -3/2 \\ -3/2 \end{pmatrix}$.

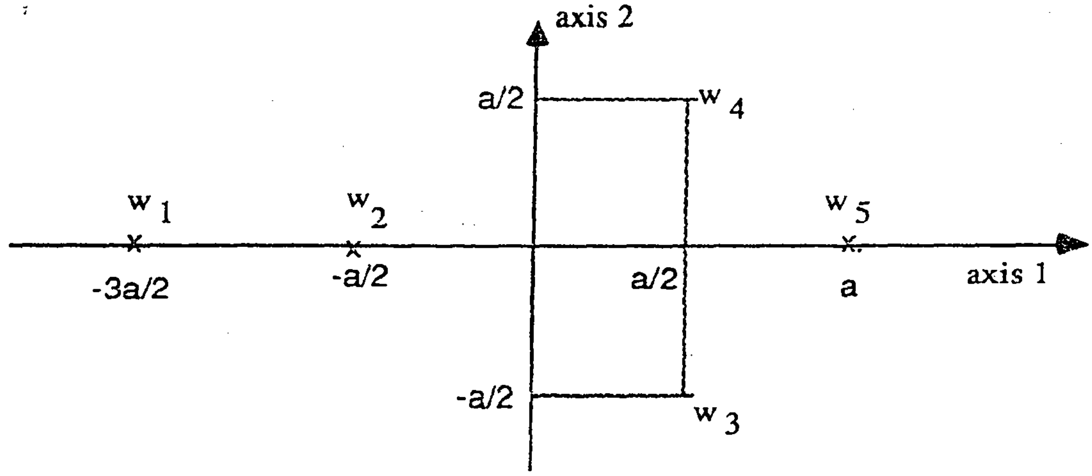

Figure 8 : Principal component analysis of table T whith a= √2.

Hierarchical and Pyramidal clustering of table T :

We make the classical "complete link hierarchy" based on the city-block distance defined by $d(w_\ell, w_k) = \sum_{j=1}^{2} |y_j(w_\ell) - y_j(w_k)|$.

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

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

To obtain a pyramid, we may use a similar algorithm where classes may be merged twice (instead only once in the case of hierarchies) if they respect a common order (for more details see for instance Brito, Diday (1990)).

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

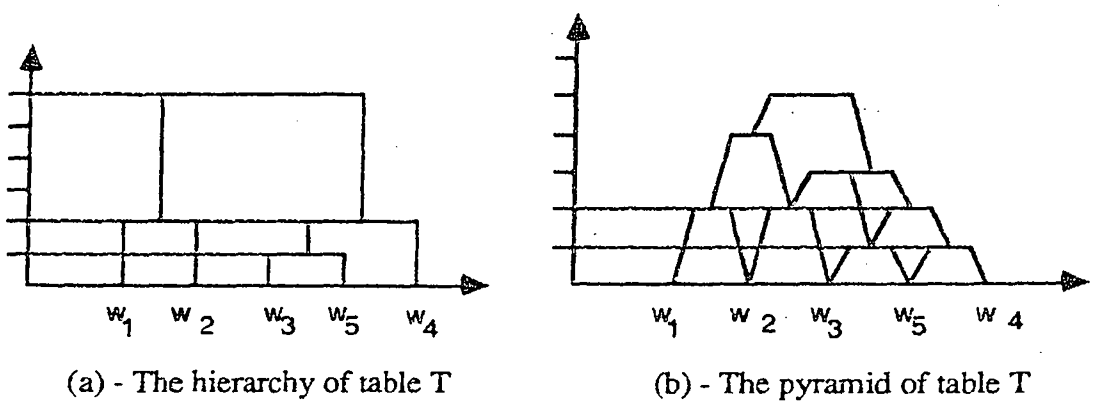

Figure 9

Remark : if we associate a dissimilarity σ induced by the hierarchy and the pyramid by setting : $\sigma(w_i, w_j)$ = {height of the lower level which contains w_i and w_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, $|d-\sigma| = \sum |d(w_i, w_j) - \sigma(w_i, w_j)|$ 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 $(w_1, \ldots, w_5)$ 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, $C_1 = \{w_1, w_2\}$, explains the 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 $\frac{1}{2}$ 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 = [Rate of employment = low] $\wedge$ [Profit = low]
a_2 = [Rate of employment = high] $\wedge$ [Profit = 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 $y_1 = -\frac{1}{2}$ then $y_2 = -\frac{1}{2}$] $\wedge$ [if $y_1 = \frac{1}{2}$ then $y_2 = \frac{1}{2}$] 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 : $h = a_1(u_1) \wedge a_2(u_2)$ = [Rate of employment (u_1) = low] $\wedge$ [Profit (u_1) = low] $\wedge$ [Rate of employment (u_2) = high] $\wedge$ [Profit (u_2) = high] whose extension is composed of couples of companies (w_i, w_j) the first element of the couple, w_i ,being of low rate of employment and profit and the second one, w_j, of high rate of employment and profit. If an external variable gives the age of the companies the horde object h may become : $h = a_1(u_1) \wedge a_2(u_2) \wedge [age(u_1) < age(u_2)]$.

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 § 8.2); more precisely, if we denote $h_1 = \{w_3, w_5\}$ 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 $h_2 = \{w_1, w_2\}$, $h_3 = \{w_3, w_4, w_5\}$ 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 set 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 $R_2 : a_3 \rightarrow a_1 \vee w_4^s$ where $w_4^s = [y_1=1] \wedge [y_2 = 2]$ is the symbolic object associated to w_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 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 $r_1 \wedge r_2$

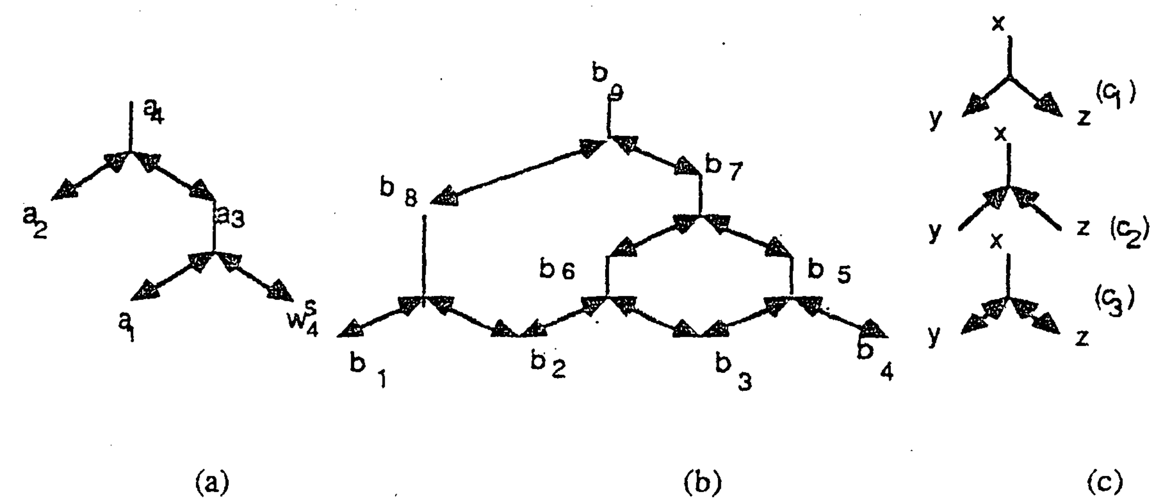

Figure 10 : Induced graph of rules between assertions (a) from the hierarchy, (b) from the pyramid, where double headed arcs are explained by (c)

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 $h_1 = \{w_1, w_2\}$, $h_2 = \{w_2, w_3\}$, $h_3 = \{w_3, w_5\}$, $h_4 = \{w_4, w_5\}$ $h_5 = \{w_3, w_4, w_5\}$, $h_6 = \{w_2, w_3, w_5\}$, $h_7 = \Omega \backslash w_1$, $h_8 = \{w_1, w_2, w_3\}$, $h_9 = \Omega$, 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 :

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

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 $\mathrm{Ext}(b_j / \Omega) = \mathrm{Ext}(b_\ell / \Omega)$, 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 § 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 $b_1 = [y_1 = q^1_1] \wedge_b [y_2 = q^1_2]$ where q^1_1 and q^1_2 are characteristic mappings such that : $q^1_1(-\frac{1}{2}) = q^1_1(\frac{1}{2}) = 1$ and $q^1_2(-\frac{1}{2}) = q^1_2(\frac{1}{2}) = 1$ and $q^1_1(v) = q^1_2(v) = 0$ elsewhere.
We have $b_2 = [y_1 = q^2_1] \wedge_b [y_2 = q^2_2]$ and $q^2_1(v) = 1$ if $v \in \{\frac{1}{2}\frac{1}{2}\}$, $q^2_1(v) = 0$ elsewhere, $q^2_2(v) = 1$ if $v \in \{\frac{1}{2}, 1\}$ and $q^2_2(v) = 0$ elsewhere. As we have (see §7), $b^*_1(b_2) = f_b(\{g_b(q^1_i, q^2_i)\}_i) = \mathrm{Min}(<q^1_1,q^2_1>, <q^1_2,q^2_2>) = \mathrm{Min}(\Sigma\{q^1_1(v)\, q^2_1(v) / v \in O_1\}, \Sigma\{q^1_2(v)\, q^2_2(v) / v \in O_2\}) = \mathrm{Min}(q^1_1(\frac{1}{2})\, q^2_1(\frac{1}{2}), q^1_2(\frac{1}{2}) q^2_2(\frac{1}{2})) = \mathrm{Min}(1,1) = 1$.
We have $b^*_1(b_1) = \mathrm{Min}(<q^1_1,q^1_1>, <q^1_2,q^1_2>) = \mathrm{Min}(2,2) = 2$ 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 11 (b) ; If we choose $c = \text{Max } d(b_i,b_j) - M$ where M is the sum of the two couples (b_i,b_j) of smallest dissimilarity $d(b_i,b_j)$, then $c \geq \text{Max } (d(b_i,b_j) - d(b_i,b_k) - d(b_k,b_j))$ and D such that $D(b_i,b_j) = d(b_i,b_j)$, $D(b_i,b_i) = 0$, is a distance, because $\forall\, i,j,k \;\; d(b_i,b_j) + c \leq d(b_i,b_k) + c + d(b_k,b_j) + c$. 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 11 (c).It is then possible to apply many existing methods of classical data analysis by using s, d or D as input.

	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

For instance, by using the following algorithm it is possible to obtain a mapping of B which respect 2 card (B) - 3 exact distances $D(b_i, b_j)$ let C_1 and C_2 be two criteria $C_1 : B \times B \rightarrow \mathbf{R}^+$ and $C_2 : B \times B \times B \rightarrow \mathbf{R}^+$ depending on some intensions, extensions and distances (in the example given below, we give two examples of each criteria); let T and N be two empty sets, at the begining; then, the algorithm is described at follows :

Step 1 Find a couple b_i, b_j which minimizes C_1 ;
put b_i and b_j in T and (b_i, b_j) in A.

Step 2 Find b_i^*, b_j^* in T and b^* in $B \setminus T$ such that:
$C_2(b^*, b_i^*, b_j^*) = \mathrm{Min}\ \{C_2(b,b_i,b_j) / b \in B \setminus T;\ b_i, b_j \in T\}$.
Put b^* in T and (b^*,b_i), (b^*, b_j^*) in A.

Step 3 While card(T) < card(B), go back to step 2.

If the user wishes to privilege a couple (b_i, b_j), this couple is chosen at the first step; at the second step, (b_i^*, b_j^*) is fixed and identical to (b_i, b_j). By this way, we get a more precise view from b_i, b_j on the other elements of B, "as if they was on a hill".

At the first step it is possible to map (b_i, b_j) in such a way that the distance $D(b_i, b_j)$ be exact. As D is a distance, at the next step, when a new element b enter in T, it is always possible to map it in such a way that the two distances $D(b, b_i)$ and $D(b, b_j)$ be exact. Finally it is easy to see that we get 2 card B-3 exact distances because we associate to the first couple only one exact distance instead of four.

Example :

In using the denotation $|a|_\Omega = \mathrm{Ext}(a/\Omega)$ we choose $C_1(b_i,b_j) = D(b_i,b_j).\ \mathrm{card}(|b_i \cup b_j|_\Omega) / \mathrm{card}(\Omega)$ and $C_2(b_i,b_j,b_k) = (D(b_i,b_j) + D(b_i,b_k))\ \mathrm{card}(|b_i \cup b_j \cup b_k|_\Omega) / \mathrm{card}\ \Omega$.

At the first step of the algorithm we obtain :

$C_1(b_3,b_4) = 0.7 \times 3/5 = 0.42 = \mathrm{Min}\ C_1(b_i,b_j)$, then $T = \{b_3,b_4\}$ and $A = (b_3,b_4)$ because $b_1 \cup b_2 = [y_1 = -\frac{1}{2}, \frac{1}{2}, 2] \wedge [y_2 = -\frac{1}{2}, \frac{1}{2}, 1]$ and $|b_1 \cup b_2|_\Omega = \{w_1,w_2,w_3\}$.

At the second step we get :
$C_2(b_5,b_3,b_4) = 2.3/5 = 1.2 = \mathrm{Min}\{C_2(b,b_3,b_4) / b \in B \setminus T\}$, then $T = \{b_3,b_4,b_5\}$ and $A = \{(b_3,b_4), (b_5,b_3), (b_5,b_4)\}$

At the third step :
$C_2(b_2,b_3,b_5) = 2.2 \times 4/5 = 1.8 = \mathrm{Min}\ \{C_2(b,b_i,b_j) \mid b \in B \setminus T,\ b_i,b_j \in T\}$ then
$T = \{b_3,b_4,b_5,b_2\}$ and $A = \{(b_3,b_4), (b_5,b_3), (b_5,b_4), (b_2,b_3), (b_2,b_5)\}$
$C_2(b_1,b_2,b_3) = C_2(b_1,b_2,b_5) = 2.9 \times 3/5 = 1.7 = \mathrm{Min}\{C_2(b,b_i,b_j) \mid b \in B \setminus T,\ b_i,b_j \in T\}$.

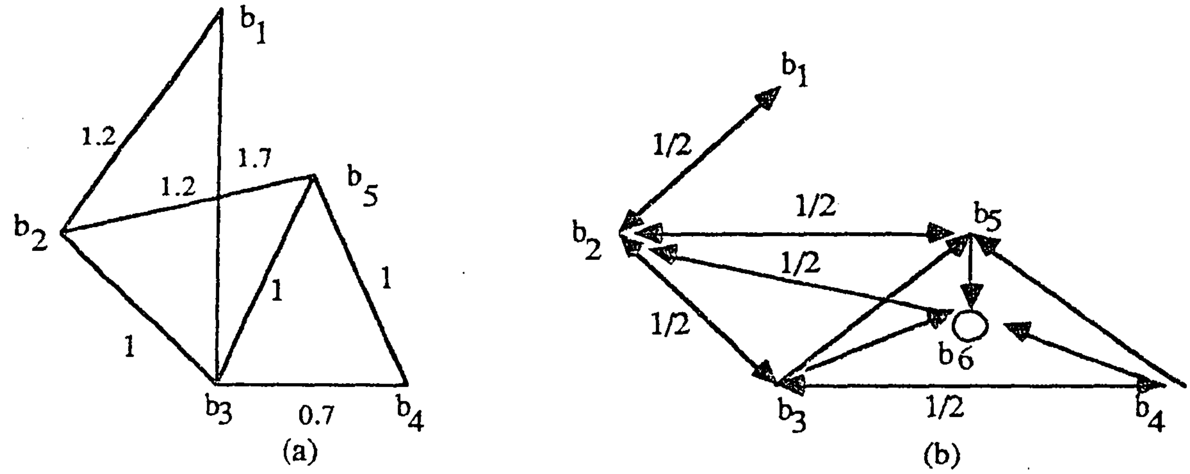

Figure 12 : (a) mapping of B with 2 card (B)-3 exact distances
(b) mapping of B with rules $a \overset{p(a/b)}{\longrightarrow} b$ if $p(a/b) \neq 0$.

We say that a implies b with a weight $p(a/b) = \mathrm{card}\ (|a|_\Omega \cap |b|_\Omega) / \mathrm{card}\ |b|_\Omega$; this rule is denoted $a \overset{p(a/b)}{\longrightarrow} b$ in figure 5b; when $p(a/b) = 1$ we get an exact implication denoted $a \rightarrow b$.

If we add the element $b_6 = b_3 \cup b_4$ we get the representation given on figure 5, by putting b_6 at the center of gravity of the elements b_3,b_4,b_5 that it generalizes (as $b_3 \cup b_4 \cup b_5 = b_6$).

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 $b_j= [y_1 = q_1^j] \wedge_{pr} [y_2 = q_2^j]$ where q_i^j is a measure of probability from $P(O_j) \rightarrow [0,1]$ where $O_j = \{-\frac{1}{2}, \frac{1}{2}, 1, 2\}$ and $P(O_j)$ is the power set of O_j. If we set $b_j = \wedge_i [y_i = (q_i^j(v_1))\, v_1,\ (q_i^j(v_2))\, v_2, \ldots]$, then the b_j are defined as follows, where the value v_ℓ associated to $q_i^j(v_\ell) = 0$ does not appear :

$b_1 = [y_1 = (\frac{1}{2})\ -\frac{1}{2}, (\frac{1}{2})\,\frac{1}{2}\,] \ {}^{\wedge}{}_{pr}\ [y_2 = (\frac{1}{2})\ -\frac{1}{2}\ , (\frac{1}{2})\ \frac{1}{2}\,]$

$b_2 = [y_1 = (\frac{1}{2})\ \frac{1}{2}, (\frac{1}{2})\,2\,] \ {}^{\wedge}{}_{pr}\ [y_2 = (\frac{1}{2})\ \frac{1}{2}\ , (\frac{1}{2})\ 1\,]$

$b_3 = [y_1 = (1)\,2] \ {}^{\wedge}{}_{pr}\ [y_2 = (\frac{1}{2})\ 1,\ (\frac{1}{2})\,2\,]$

$b_4 = [y_1 = (\frac{1}{2})\,1, (\frac{1}{2})\,2\,] \ {}^{\wedge}{}_{pr}\ [y_2 = (1)\,2\,]$

$b_5 = [y_1 = (\frac{1}{2})\ 1,\ (\frac{1}{2})\,2\,] \ {}^{\wedge}{}_{pr}\ [y_2 = (\frac{1}{2})\,1,\ (\frac{1}{2})\,2\,].$

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 :

$b_1^*\ (b_2) = \text{Mean}\ (< q_1^1,\ q_1^2 >, < q_2^1, q_2^2 >)$; therefore :

$b_1^*\ (b_2) = \text{Mean}\ (\Sigma\ \{q_1^1(v).\ q_1^2(v) / v \in O_1\},\ \Sigma\ \{q_2^1(v).\ q_2^2(v) / v \in O_2\}).$

Hence $b_1^*\ (b_2) = \text{Mean}\ (\frac{1}{2} \times 0 + \frac{1}{2} \times\ \frac{1}{2} + 0 \times 0 + 0 \times \frac{1}{2},\ \frac{1}{2} \times 0 + \ \frac{1}{2} \times\ \frac{1}{2} + 0 \times\ \frac{1}{2} + 0 \times 0) =$

$\text{Mean}\ (\frac{1}{4},\ \frac{1}{4}) = (\frac{1}{4} + \frac{1}{4})\ \frac{1}{2} = \frac{1}{4}.$

$b_1^*\ (b_1) = \text{Mean}\ (\frac{1}{2},\ \frac{1}{2}) = \frac{1}{2};\ b_2^*\ (b_2) = \text{Mean}\ \ (\frac{1}{2},\ \frac{1}{2}) = \frac{1}{2}.$

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

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

To analyse 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 $a_{jk} \geq_\alpha \text{Max}\ (a_j,a_k)$ where $\cup_{pr,\alpha}$ and $\geq_\alpha$ have been defined in §8.1. To do so, we may use the following algorithm of sequential agglomerative hierarchical clustering 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 $a^i_{jk} = [y_i=q_i]$ such that $Ext(a^i_{jk}/\Omega,\alpha)$ contains $\Omega_1 = Ext(a_j/\Omega,\alpha) \cup Ext(a_k/\Omega,\alpha)$.

Hence, for any $w \in \Omega_1$ such that $w^s = \hat{\wedge}_i[y_i= r_i]$ we have $a_{jk}(w) \geq \alpha$; this condition is satisfied if we have $\forall$ i , $g(q_i,r_i) \geq \alpha$ because $a_{jk}(w) = f(\{g(q_i,r_i)\}_i)$ and, by definition of f, it is the mean of numbers larger then α ; hence, if we denote $x^i_j= q_i (v_j)$, we have the inequality :

$g(q_i,r_i) = \Sigma\{x^\ell_j.r_i(v_j)/v_j \in O_j\} \geq \alpha$; hence, we have to solve a system of card(Ω_1) inequalities where the unknowns are the x^ℓ_j. If this system has several solutions, for each i we denote them $[y_i=q^\ell_i]$; hence, we obtain $a_{jk} = \hat{\wedge}_{i\,pr} (\hat{\wedge}_{\ell\,pr}[y_i=q^\ell_i])$; by choosing h_{pr}=Min (see §4.5) the extension of a_{jk} at level α is $\Omega_2=\{w/a_{jk}(w) = f(\{\underset{\ell}{Min}\{g(q^\ell_i,r_i)\}\}_i \geq \alpha\}$.

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 $a_{12} = a^1_{12} \wedge a^2_{12}$ where $a^\ell_{12} = [y_\ell=q_\ell]$ is such that $a^\ell_{12}(b_1) \geq \frac{1}{2}$ and $a^\ell_{12}(b_2) \geq \frac{1}{2}$. Then, for $x^\ell_j = q_\ell (v_j)$ where $\{v_1,...,v_4\} = O_1 = O_2 = \{-\frac{1}{2}, \frac{1}{2}, 1, 2\}$, we have to solve the following inequalities, where the x^ℓ_j are the unknowns, with the constraint $\sum_j q_\ell(v_j)= \sum_j x^\ell_j =1$ and $x^\ell_j \in \{0, \frac{1}{2}, 1\}$ in order to simplify :

$a^\ell_{12}(b_1) = g_{pr}(q_\ell,q^1_1) = \Sigma\,\{q_\ell(v_i)\, q^1_1(v_i) / v_i \in O_\ell\}$; hence, we obtain :

$a^1_{12}(b_1) = \frac{1}{2} x^1_1 + \frac{1}{2} x^1_2 \geq \frac{1}{2}$; $a^1_{12}(b_2) = \frac{1}{2} x^1_2 + \frac{1}{2} x^1_4 \geq \frac{1}{2}$

from which we deduce that $x^1_1 + x^1_2 = 1$, and $x^1_2 + x^1_4 = 1$ therefore (as $\Sigma\{x^1_i$ / i=1, 4} =1) we get $x^1_2 = 1$, $x^1_i = 0$ if $i \neq 2$.

$a^2_{12}(b_1) = \frac{1}{2} x^2_1 + \frac{1}{2} x^2_2 \geq \frac{1}{2}$, $a^2_{12}(b_2) = \frac{1}{2} x^2_2 + \frac{1}{2} x^2_3 \geq \frac{1}{2}$, from which it results that $x^2_2 = 1$ and $x^2_i = 0$ if $i \neq 2$.Finally we obtain :

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

Similarly, we get : $a^1_{13}(b_1) = \frac{1}{2} x^1_1 + \frac{1}{2} x^1_2 \geq \frac{1}{2}$ and $a^1_{13}(b_3) = x^1_4 \geq \frac{1}{2}$. This is a contradiction because the first equation implies $x^1_4 = 0$. 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 $q_i : P(O_i) \rightarrow \{1\}$ (which are not, of course, probabilities !), then it is easy to see that $\Omega^s(w) = 1$, $\forall\ w \in \Omega$. Similarly we get : $a_{14}=a_{15}=a_{24}=\Omega^s$; $a^1_{23}(b_2) = \frac{1}{2} x^1_2 + \frac{1}{2} x^1_4 \geq \frac{1}{2}$ and $a^1_{23}(b_3) = x_4 \geq \frac{1}{2}$ gives two solutions

i) $x^1_2 = x^1_4 = \frac{1}{2}$, $x^1_3 = x^1_1 = 0$

ii) $x^1_4 = 1$, $x^1_i = 0$ if $i \neq 4$

therefore : $a_{23} = [y_1 = (\frac{1}{2})\frac{1}{2}, (\frac{1}{2})2] \wedge_{pr} [y_1=(1)2] \wedge_{pr} [y_2=(1)1]$; $a_{25}=[y_1=(1)2] \wedge_{pr} [y_2=(1)1]$; $a_{34}=[y_1=(1)2] \wedge_{pr} [y_2=1(2)]$; a_{35} is computed as follows : $a^1_{35}(b_3) = x^1_4 \geq \frac{1}{2}$ and $a^1_{35}(b_5) = \frac{1}{2} x^1_3 + \frac{1}{2} x^1_4 \geq \frac{1}{2}$ implies $x^1_4 = 1$ and $x^1_i = 0$ if $i \neq 4$; $a^2_{35}(b_3) = \frac{1}{2} x^2_3 + \frac{1}{2} x^2_4 \geq \frac{1}{2}$ and $a^2_{35}(b_5) = \frac{1}{2} x^2_3 + \frac{1}{2} x^2_4 \geq \frac{1}{2}$; we have three solutions i) $x^2_3 = x^2_4 = \frac{1}{2}$; ii) $x^2_3 = 1$, $x^2_i = 0$ for $i \neq 3$; iii) $x^2_4 = 1$, $x^2_i = 0$ for $i \neq 4$; therefore :

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

In a similar way we get :

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

In the following table we give in the cell of the ith row and jth column the extension of $a_{ij} = b_i \cup_{pr,1/2} b_j$:

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

	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); 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 b_4 to be merged with (b_1,b_2) or (b_3,b_5). It is then easy to see that $a^2_{124}(b_1) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 \geq \frac{1}{2}$ and $a^2_{124}(b_4) = \frac{1}{2}$ $x_2^4 \geq \frac{1}{2}$ which give no solution such that

$\sum_{i=1,4} x_i^2 = 1$; therefore, $a_{124} = \Omega^s$ whose extension is B. We have already seen that $Ext(a_{34}/B,\frac{1}{2}) = \{b_3,b_4,b_5\}$, 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.

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

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.

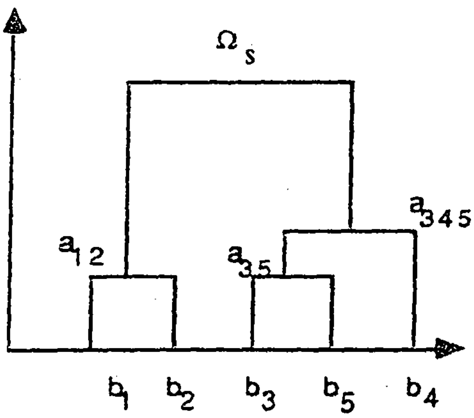

Figure 13 : Inheritance hierarchy on probabilist objects

10.6. Induction by probabilist, possibilist and belief union

Notice that the same algorithm may be used with the probabilist, possibilist and belief union defined respectively in §5 §6 §7 instead of the symbolic union defined in §9 which has been used here. The advantage of the symbolic union (see §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 $a_3^*(a_1 \cup_{pr} a_2) = a_3^*(a_1) + a_3^*(a_2) - a_3^*(a_1 \cap_{pr} a_2) \geq a_1^*(a_1) + a_2^*(a_2) - a_3^*(a_1 \cap_{pr} a_2)$; it results that the obtained hierarchy will have no inversions (as it may be shown that $a_3^*(a_3) \geq a_1^*(a_1)$) and $a_3^*(a_3) \geq a_2^*(a_2)$ 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 $Ext(a_1/B, \alpha_1) \subseteq Ext(a_2/B,\alpha_2)$; 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. Generalization of symbolic objects

11.1. Generalization of a symbolic object

In AI and more specially in "Machine Learning" there are many classical ways to do generalizations (see for instance Michalski, Carbonell and Mitchell 1983, 1986), we present here only four of them :

a) generalization by "dropping rule" : this consists of dropping an event in the description of a symbolic object ; for instance a = [color = red, green] $\wedge$ [height = [0,15]] is generalized by a_g = [color = red, green, grey].

b) generalization by dropping values in an event : for instance a = [color = red, green] is generalized by a_g = [color = red].

c) Generalization by taxonomy :

If we consider that red, green and black are dark colors a_g = [color = dark] generalizes a = [color = blue, black].

d) Generalization by changing contents in variables :

This may be applied to "structured objects" and more precisely to what we have called a "horde" (see 3.3) ; for instance, if we have an instanciated horde defined on a couple of human of the following kind :

h(Tom, Jane) = [height (Tom) = high] $\wedge$ [height (Jane) = low] $\wedge$ [weight (Tom) = low] whose extension is the couple (Tom, Jane) ; it may be generalized by :

h(v, Jane) = [height (v) = high] $\wedge$ [height (Jane) = low) $\wedge$ weight (v) = low] whose extension is the set of couples (v, Jane) where v is a human.

11.2. Generalization in the case of several symbolic objects

We may consider at least the following two cases

i) Generalization by the symbolic union.

In 10.5 we have defined an algorithm which induces from several symbolic objects $a_1, \ldots, a_n$ a new one denoted b, such that $b = \bigcup_{i}{}_{x,\alpha}\, a_i$ which means that b is the conjunction of the events whose extension at level α contains the extension of the a_i at the same level ; this definition implies $b \geq_\alpha a_i\ \forall i$, therefore b generalizes the a_i at the α level.

ii) Generalization by probabilist, possibilist and belief union.

In this case $A = \{a_1, \ldots, a_n\}$ is generalized by $b = \bigcup_{i}{}_{x}\, a_i$; in the case of probabilist, possibilist and belief objects the union insures that $b \geq a_i$ and therefore b is a generalization of the a_i.

12. Fitting a set of symbolic objects

12.1. Fitting without decomposition

In this case we look for a fitting $b^* \in \mathfrak{a}_x^*$ of the $a_i \in$ A in order to get an extension of b* in $\mathfrak{a}_x$ which contains the a_i at the highest level ; in other words, we may look for $b^* \in \mathfrak{a}_x^*$ in such a way that $W(b) = \operatorname{Min}_i p_i\, b^*(a_i)$ be maximum, where p_i = 1/card (Ext (b*/A, $b^*(a_i)$)) which means that the extension in A of b* at level $b^*(a_i)$ must be as small as possible in order to avoid the trivial solution $b^* = 1_{\mathfrak{a}_x}$ which extension is $\mathfrak{a}_x$.

Many other kinds of criterion to optimize my be defined ; for instance :

$$W_1(b) = \prod_i p_i b^*(a_i)$$
$$W_2(b) = \sum_i p_i\, b^*(a_i)$$
$$W_3(b) = \sum_i (b^*(a_i))^2$$

The first two cases are equivalent. In the third one if $b = [y = q]$ and $a_i = [y = q_i]$, in the case of probabilist objects we get $W_3(b) = \sum_i <q, q_i>^2$ and so b^* is the first factor of a kind of factorial analysis with the constraints for any i : $\forall\, v \in O_i$, $0 \leq q_i(v) \leq 1$ and $\Sigma\{(q_i(v))^2/v \in O_i\} = 1$ or $\Sigma\{q_i(v)/v \in O\}=1$ if the q_i are supposed to be probabilities. Hence, in this case we avoid the trivial solution without using the weights p_i.

12.2. Fitting with a set of symbolic objects by a generalization of a fitting decomposition

Given $A = \{a_1, \ldots, a_n\}$ a set of symbolic objects, the "decomposition problem" consists of looking for a "decomposition" $B = \{b_1, \ldots, b_k\}$ of k symbolic objects (generally k is much smaller than n) such that a generalization b of B "fit" as well as possible A. If the b_i are modal assertion of the following kind : $b_i = \hat{\wedge}_{i}x\ [y_i = t_i]$ where t_i is a mapping $O_i \rightarrow [0,1]$, the unknown are $\forall v \in O_i$ the $t_i(v)$; if O_i is large and in order to get an other kind of explanatory power it may be useful to introduce models and to use symbolic objects of the following kind : $b_i = \hat{\wedge}_{i}x\ [y_i = t_i(c_i, .)]$ where c_i are the parameters which define the mapping t_i and are the unknown ; for instance in the case on probabilist assertions $t_i(c_i, .)$ may be a gaussian probability density of parameters c_i.

In this section we consider the three following generalizations of B :

i) $b^* = \mathrm{Max}_i\, b_i^*$ (which means that $b^*(a) = \mathrm{Max}_i\, b_i^*(a)$

ii) $b^* = \vee_i x\ b_i^*$

iii) $b^* = \sum_i p_i\, b_i^*$ (where $p_i \geq 0$, $\Sigma\, p_i = 1$ and p_i is proportionnal to the size of the extension of b_i^* in A).

Notice that in case i) and in case ii) for probabilist, possibilist and belief objects b^* is more general then b_i^* (i.e. $\forall a\ \ b^*(a) > b_i^*(a)$) ; in case iii) b^* is more general than $p_i\, b_i^*$. Let us study in more detail these three cases.

i) $b^* = \underset{\ell=1,k}{\mathrm{Max}}\ b_\ell^*$; for instance, if each a_i represents a species of mushrooms, each b_ℓ represents a class P_ℓ of species such that b_ℓ and P_ℓ fit as well as possible ; this fit may be represented, for instance, by a mapping $D_i : \mathcal{a}^* \times P(A) \rightarrow \mathbf{R}^*$ such that, for instance :

$$D_1(b,P_\ell) = \Sigma\ \{b^*(a)/a \in P_\ell\}$$
$$D_2(b,P_\ell) = \Pi\{b^*(a)/a \in P_\ell\}$$
$$D_3(b,P_\ell) = \Sigma\{(b^*(a))^2/a \in P_\ell\}$$
$$D_4(b,P_\ell) = \mathrm{Min}\{b^*(a)/a \in P_\ell\}$$

Notice that D_1 and D_2 are equivalent as $b^*(a) \geq 0\ \forall a$ but D_2 is sometimes more practicable in order to optimize it. The decomposition problem may be set in by the following way :
find $B = (b_1,..b_k)$ and $P = (P_1, \ldots, P_k)$ which maximizes the criterion $W(B,P) = \sum_{j=1,k} D(b_j, P_j)$ which is equivalent when $D=D_1$, to $W(B,P) = \Sigma\{b^*(a)/a \in A\}$ if $P_j = \{a/b_j^*(a) \geq b_i^*(a)\}$ as in this case $D_1(b_j,P_j) = \Sigma\{b^*(a)/a \in P_j\}$.
If $D=D_3$, the optimization of W consists of finding k factorial axis which fit the best k classes of a partition to be also found ; it is a kind of "typological factorial analysis" in the framework of the Dynamical Clustering Method (see Diday, Simon (1977), Diday et al (1979)).
If $D=D_4$ the optimization of W consists of finding $B = (b_1, \ldots, b_k)$ such that in each class P_ℓ the extension of b_ℓ^* at level $\alpha = \mathrm{Min}\ \{b_\ell^*(a)/a \in P_\ell\}$ is as high as possible.
In order to avoid the trivial solution $b^* = 1\mathcal{a}_x$ several kinds of constraints may be added, for instance, by setting $b^*(a)=p_i\, b_1^*(a_i)$ where $p_i = 1/\mathrm{card\ Ext}\ (b^* / A, b^*(a_i))$ as in 12.1 ; if $b^* = \hat{\wedge}_x\ [y_i = q_i]$ several kinds of constraints on the q_i may also avoid the trivial solution, for instance $\Sigma\{q_i(v)/v \in O_i\} = 1$ or $\Sigma\{q_i^2(v)/v \in O_i\} = 1$; if moreover, we wish b^* to be a generalization of A we have to add the constraints $b(w) \geq a(w)\ \forall w \in \Omega$.

In order to optimize such criteria, the following algorithm based on the "Dynamic Clustering Method" may be used :
Starting step : a partition $P^{(0)} = (P_1^{(0)}, \ldots, P_k^{(0)})$ of A estimated by an expert or chosen at random.
Representation step : compute $B^{(1)} = \{b_1^{(1)}, \ldots, b_k^{(1)}\}$ such that $D(b_j^{(1)}, P_j) = \mathrm{Max}\ \{D(b, P_j^{(0)})/\ b \in \mathcal{a}$ and b satisfies some given constraints.
Assignement step : compute $P_j^{(1)} = \{a \in A/b_j^{(1)^*}(a) \geq b_i^{(1)^*}(a), j \neq i\}$.
By using alternatively these two process the criterion W defines a sequence $u_1 = W(B^{(n)}, P^{(n)})$ which increases at each step, until convergence.

ii) $b^* = \underset{i=1,k}{\cup_x} \overset{*}{b_i}$; for instance, if $b_i = [y = t_i]$ for $i = 1,2$ in the case of probabilist assertions $b = b_1 \cup_{pr} b_2 = [y = t_1 \cup_{pr} t_2]$; if $a_i = [y = r_i]$ we get by definition $b^*(a_i) = <t_1 \cup_{pr} t_2, r_i> = <t_1, r_i> + <t_2, r_i> - <t_1\, t_2, r_i> = \overset{*}{b_1}(a_i) + \overset{*}{b_2}(a_i) - \overset{*}{b_1} \cap_{pr} \overset{*}{b_2}(a_i)$; this means that $b^*(a)$ will be the highest if $\overset{*}{b_1}(a_i)$ (resp $\overset{*}{b_2}(a_i)$) is the highest and $\overset{*}{b_2}(a_i)$ (resp $\overset{*}{b_1}(a_i)$) is the lowest, as in this case we get $b^*(a_i) = 1 + 0 - 0 = 1$. Unlike to the preceding case i), in this case $b^*(a_i)$ doesn't take only account of $\overset{*}{b_\ell}(a) = \underset{i}{\mathrm{Max}}\ \overset{*}{b_i}(a)$ but also of the other values $\overset{*}{b_j}(a)$ with $j \neq \ell$; that is why, the case i) may be considered as an "extreme" solution of this case ii).
In this case the decomposition problem may be set by the following way : find $b \in \mathcal{a}$ such that the criterion $W(b,A) = \Pi\{b^*(a)/a \in A\}$ be maximized and $\Sigma\{\underset{i}{\Pi}\, b_i(a) / a \in A\}$ be minimized. In order to optimize such a criteria, the following algorithm may be used :
Initialisation : start from $b^{(0)} = \underset{i=1,k}{\cup_x} b_i^{(0)}$ where the $b_i^{(0)}$ are estimated by an expert, or obtained from the algorithm given in i) or chosen at random.
Decomposition at step n : find sequentially $b_1^{(n)}, \ldots, b_k^{(n)}$ which maximizes $\Pi\{b^{(n)*}(a)/a \in A\}$ and minimizes $\Sigma\, \{\underset{i}{\Pi}\, b_i^{(n)*}(a) / a \in A\}$; set $b^{(n+1)} = \underset{i=1,k}{\cup} b_i^{(n)}$ and compute by the same way $b_1^{(n+1)}, \ldots, b_k^{(n+1)}$.
This process induces a sequence $u_n = W(b^{(n)}, A)$ which increases at each step until convergence. Notice, that in order to avoid the trivial solution several kinds of constraints may be added to the b_i.

iii) $b^* = \sum_i p_i \overset{*}{b_i}$ where p_i is proportional to the extension of $\overset{*}{b_i}$ in A of b^*, we may say that $b^*(a)$ is high if $\overset{*}{b_i}(a)$ is high and the extension of $\overset{*}{b_i}$ in A is large.
The decomposition problem may be set in the following way where $D(b_i,P_i)=\Pi\{\overset{*}{b_i}(a)/a \in P_i\}$ find $b = (b_1, \ldots, b_k)$ and $P = (P_1, \ldots, P_k)$ which maximizes the criterion :

$$W(B,P) = \sum_{i=1,k} p_i\, D(b_i, P_i).$$

In order to optimize W many algorithm inspired from the Dynamical Clustering Method may be applied ; for instance, the following :

a) Start from $\{b_1^{(0)}, \ldots, b_k^{(0)}\} \subseteq A$ and $P^{(0)} = (P_1^{(0)}, \ldots, P_k^{(0)})$ given by an expert or chosen at random and settle $p_\ell^{(0)} = \dfrac{\text{card } P_\ell^{(0)}}{\text{card } A}$.

b) Assignement step, by two possible ways :

i) $P_\ell^{(1)} = \{a \in A / b_\ell^{(0)*}(a) \geq b_j^{(0)*}(a) \quad \ell \neq j\}$

ii) $P^{(1)}$ is build by assigning at random to one of the cluster $P_1^{(1)}, \ldots, P_k^{(1)}$ with probability $\overset{*}{b}_\ell(a) = p_\ell^{(0)}\, b_\ell^{(0)}(a) \,/\, W(B^{(0)}, P^{(0)})$.

This second choice (see Celeux, Diebolt (1985) or Celeux et al (1989)) reduces the initial position dependance.

c) Representation step, by the following way :

$$b_\ell^{(1)} = \operatorname{Arg} \underset{b \in a_x}{\operatorname{Max}} \prod_{a \in P_\ell^{(1)}} (p_\ell^{(1)} \,;\, b(a)/W(B,P^{(1)})) \text{ where } p_\ell^{(1)}\,) = \frac{\text{card } P_\ell^{(1)}}{\text{card } A}$$

Those two steps are repeated iteratively while $U_n = W(B^{(n)},P^{(n)})$ increases until convergence. Notice that in the special case where the assertions $a \in A$ may be assimilated to point of $\mathbb{R}^p$ and the b_i to probability distribution defined on $\mathbb{R}^p$, this problem and algorithm enters in the framework of the mixture decomposition of probability distribution problem by the maximum likelihood approach in classical statistics and the EM family of algorithms (see Dempster et al (1977),Everitt and Hand (1981), Celeux and Govaert (1992)) gives solution to it.

12.3. Induction by mixture decomposition in the case of probabilist objects

Let $A = \{a_1,\ldots,a_n\}$ and $A_1,\ldots,A_k$ a partition of A.

In the case of probabilist objects theorem 2 shows that $b^*(A_1 \cup_{pr} A_2) = b^*(A_1) + b^*(A_2) - b^*(A_1 \cup_{pr} A_2)$. More generally by using the Poincarré formula we get :

$$b^*(A_1 \cup_{pr} \ldots \cup_{pr} A_k) = \sum_{i=1,n} b^*(A_i) - \sum_{i<j} b^*(A_i \cap_{pr} A_j) + \ldots + (-1)^{n+1}\, b^*(A_1 \cap_{pr} \ldots A_k)\,.$$

From this there results the following proposition, where by definition $b^*(a/A) = b^*(a \cap_{pr} A)/b^*(A)$, $f(a) = b^*(a/A_1 \cup_{pr} \ldots \cup_{pr} A_n)$, $f_i(a) = b^*(a/A_i)$, $f_{i\ldots\ell}(a) = b^*(a/A_i \cap_{pr} \ldots A_\ell)$, $p_i = b^*(A_i)/b^*(A_1 \cup_{pr} \ldots A_n)$, $p_{i\ldots\ell} = b^*(A_i \cap_{pr} \ldots A_\ell) \,/\, b^*(A_1 \cup_{pr} \ldots A_n)$, $\underset{\ell}{\cap}_{pr}\, a = a \cap_{pr} \ldots a$ ℓ times.

Proposition :

$$f(a) = \Sigma\, p_i\, f_i(a) - \sum_{i<j} p_{ij}\, f_{ij}\, (\underset{2}{\cap}_{pr}\, a) \ldots + (-1)^{k+1}\, p_{1\ldots k}\, f_{1\ldots k}\, (\underset{k}{\cap}_{pr}\, a)$$

Proof :

We have :

$$b^*(a/A_1 \cup_{pr} ...A_k) = b^*(a \cap_{pr} (A_1 \cup_{pr} ... A_k))/b^*(A_1 \cup_{pr} ... A_k)$$

where, $b^*(a \cap_{pr} (A_1 \cup_{pr} ... A_k)) = b^*(a \cap_{pr} A_1 \cup_{pr} ... a \cap_{pr} A_k)$ from the Poincarré formula it results :

$$b^*(a \cap_{pr} (A_1 \cup_{pr} ... A_k)) = \sum_i b^*(a \cap_{pr} A_i) - \sum_{i<j} b^*(\underset{2}{\cap}_{pr}\, a \cap_{pr} A_i \cap_{pr} A_j) ...$$
$$+ (-1)^{k+1} b^*(\underset{k}{\cap}_{pr}\, a \cap_{pr} A_1 ... \cap_{pr} A_k) .$$

Hence :

$$b^*(a \cap_{pr} (A_1 \cup_{pr} ... A_k)) = \sum_i b^*(A_i)\, b^*(a/A_i) - \sum_{i<j} b^*(A_i \cap_{pr} A_j)\, b^*(\underset{2}{\cap}_{pr}\, a/A_i \cap_{pr} A_j)$$
$$+ ... + (-1)^{k+1} b^*(A_1 \cap_{pr} ... A_k)\, b^*(\underset{k}{\cap}_{pr}\, a/A_1 \cap_{pr} ... A_k) .$$

From which we get :

$$b^*(a/A_1 \cup_{pr} ... A_k) = \sum_i \frac{b^*(A_i)}{b^*(A_1 \cup_{pr} ... A_k)} b^*(a/A_i)$$
$$- \sum_{i<j} \frac{b^*(A_i \cap_{pr} A_j)}{b^*(A_1 \cup_{pr} ... A_k)} b^*(\underset{2}{\cap}_{pr}\, a/A_i \cap_{pr} A_j)\ \ b^*(\underset{2}{\cap}_{pr}\, a/A_i \cap_{pr} A_j)$$
$$+ ... + (-1)^{k+1} \sum_{i<j} \frac{b^*(A_i \cap_{pr} ... A_k)}{b^*(A_1 \cup_{pr} ... A_k)} b^*(\underset{k}{\cap}_{pr}\, a/A_1 \cap_{pr} ... A_k)$$

Therefore

$$f(a) = \sum_i p_i f_i(a) - \sum_{i<j} p_{ij} f_{ij} (\underset{2}{\cap}_{pr}\, a) + ... + (-1)^{k+1} p_{1...k} f_{1...k} (\underset{k}{\cap}_{pr}\, a) .$$

■

In order to find the f_i, p_i and A_i, many algorithms inspired from the Dynamical Clustering Method, may be built and have to be compared, for instance, the following :

Starting step : a partition $A_1,...,A_k$ of A chosen at random or given by an expert.

Representation step : compute p_i and f_i such that

$$(p_i,f_i) = \text{Arg Max}\left\{ \prod_{a \in A_i} p_i\, f_i(a) \;/\; p_i \in [0,1],\ c_i \in C \right\}$$

where c_i are the parameters which caracterizes f_i among a set C. In the same way the $p_{i...\ell}$, $f_{i...\ell}$, $c_{i...\ell}$ are computed :

$$(p_{i...\ell}, f_{i...\ell}) = \text{Arg Max} \left\{ \prod_{a \in A_i \cap ... A_\ell} p_{i...\ell}\, f_{i...\ell}(a) \;/\; p_{i...\ell} \in [0,1],\ c_{i...\ell} \in C \right\}$$

Assignement step : $A_i = \{a/f_i(a) > f_j(a)\}$ $i=1,...,k$ or a is assigned at random to one of the class $A_1,...,A_k$ with probability $p_i f_i(a)/f(a)$ for $i=1,...,k$.

Notice that in the representation, we may compute successively for each $i...\ell$: $f_{i...\ell}$ which has to fit as well as possible $\bigcap_{j=i,\ell} {}_{pr} A_{j...}$, then b^* by solving the equations $f_{i...\ell}(a) = b^*(a/A_i \cap_{pr} ... A_\ell)$ and finally $p_{i...\ell}$ knowing b^* .

12.4. Decomposition of a generalisation by local fitting generalisation

In this section the problem is to decompose a generalisation b^* of $A = \{a_1, ..., a_n\}$ into k local generalizations of A, denoted $B = \{b_1, ..., b_k\}$ such that each b_i "fits" as well as possible a subset of A that it generalizes. Such generalization of A may be obtained for instance in the two following ways where $P = (P_1, ..., P_k)$ is a partition of A :

i) $b^* = \underset{i=1,k}{\text{Max}}\ b_i^*$ with $b_i = \text{Max}\ \{a/a \in P_i\}$

ii) $b^* = \underset{i=1,k}{\cup_x}\ b_i^*$ with $b_i = \cup_x \{a/a \in P_i\}$

It results that in both cases, b^* is a generalization of A as $\forall i = 1,k$ and $a \in P_i$, $b \geq b_i \geq a$. How can we get the P_i in order that each b_i generalizes and fit as well as possible P_i ? We may use the following criterion :

$$W(B,P) = \sum_i D(b_i, P_i) \text{ with } B = (b_1, ..., b_k) \in \mathfrak{a}^k$$

$$\text{and } D(b_i,P_i) = \Sigma \left\{ p_i(a)\, b_i^*(a)/a \in P_i,\ a \leq b_i,\ p_i(a) = \frac{\text{Ext } (b_i^*/P_i,\ b_i^*(a))}{\text{Ext } (b_i^*/A,\ b_i^*(a))} \right\}$$

In order to get a local maxima of W we may use the Dynamic Clustering Method in the following way :

Initialisation : start from $P^{(0)} = \{P_1^{(0)}, ..., P_k^{(0)}\}$ a partition of A, estimated by an expert or chosen at random.

Representation step : compute $B^{(1)} = \{b_1^{(1)}, \ldots, b_k^{(1)}\}$ such that

$$D(b_i^{(1)}, P_i^{(0)}) = \text{Max } \{D(b, P_i^{(0)})/b \in \mathcal{a}\} \text{ and } p_i^{(1)}(a) = \frac{\text{Ext } (b_i^{(1)}/P_i^{(0)}, b_i^{(1)}(a))}{\text{Ext } (b_i^{(1)}/A, b_i^{(1)}(a))} .$$

Assignement step : compute $P^{(1)} = \{P_1^{(1)}, \ldots, P_k^{(1)}\}$ such that $P_i^{(1)} = \{a/p_i^{(1)}(a)\ b_i^{(1)}(a) \geq p_j^{(1)}(a)\ b_j^{(1)}(a)\ \ i \neq j\}$.

By using alternatively the assignement and representation step we define a sequence $\cup_n = W(B^{(n)}, P^{(n)})$ which increases at each iteration until convergence. The partition $P^{(N)}$ obtained at the convergence is an answer to the preceding question ; notice, that this algorithm also gives at convergence $B^{(N)} = \{b_i^{(N)}, \ldots, b_k^{(N)}\}$ where the $b_i^{(N)}$ may be considered as "local prototypes" of A.

12.5. A geometrical example

In this case, $A = \{a_1, \ldots, a_k\}$ is a set of k probabilist assertions defined on $\Omega = \mathbf{R}^2$; each a_i is associated to a point of the plane $w_i \in \Omega_1 = \{w_1, \ldots, w_k\} \subseteq \mathbf{R}^2$ such that $a_i = [y = e^{-d(w_i,.)}]$ where d is a dissimilarity measure and y is a mapping $\mathbf{R}^2 \to Q^{pr}$ where Q^{pr} is a set of mappings $q : O \to [0,1]$ such that $0 \leq q(v) \leq 1$ where O is a finite set of $\mathbf{R}^2$ chosen such that card O be generally larger than card Ω_1 = card A = k.

We associate to any point $w \in \mathbf{R}^2$ a symbolic representation denoted $w^s = [y = \delta_w]$ where δ_w is the Dirac mass (i.e. $\delta_w(w) = 1$ and $\delta_w(v) = 0$ if $v \neq w$). If $w \in O$ it results by definition that :

$a_i(w) = \langle e^{-d(w_i,.)}, \delta_w \rangle = \sum \{e^{-d(w_i,v)}, \delta_w(v)/v \in O\} = e^{-d(w_i,w)}$; hence, $a_i(w_i) = 1$.

More precisely, it is possible to define probabilist assertions associated to geometrical curves (circles, straight lines etc...) by defining a more general dissimilarity also denoted d, such that d(c,w) is the dissimilarity between a curve denoted c and a point $w \in O$; if we settle b $= [y = e^{-d(c,.)}]$ we get $b(w) = \langle e^{-d(c,.)}, \delta_w \rangle = e^{-d(c,w)}$; hence, b measures the closeness between a point and the curve c associated to b ; we also have :

$b^*(a_i) = \frac{1}{\text{card } O} \langle e^{-d(c,.)}, e^{-d(w_i,.)} \rangle = \frac{1}{\text{card } O} \sum_1 e^{-(d(c,v)+d(w_i,v))}$; hence $b^*(a_i)$ will be high if there are many points of O simultaneously close to the curve c and to w_i (see figure 14)

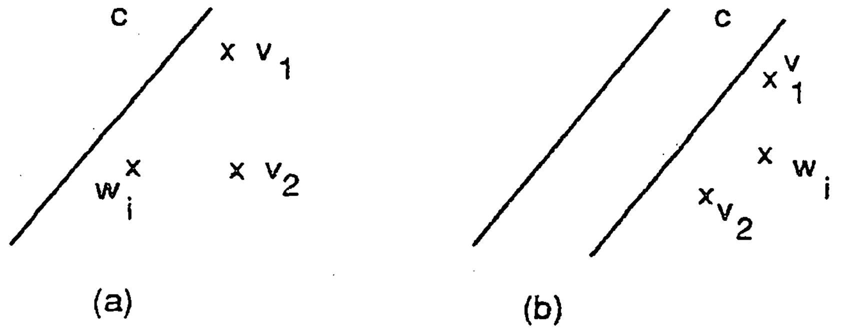

Figure 14 : In (a) $b^*(a_i)$ is lower than in (b) ; in other words w_i is considered to be closer to c in (b) than in (a).

If we denote $b_i = [y = q_i]$ with $q_i = e^{-d(c_i,.)}$ we have :
$b_1 \cup_{pr} b_2 = [y = q_1 \cup_{pr} q_2] = [y = q_1 + q_2 - q_1\ q_2] = [y = e^{-d(c_1,.)} + e^{-d(c_2,.)} - e^{-(d(c_1,.)+d(c_2,.))}]$; therefore $b_1 \cup_{pr} b_2(w) = e^{-d(c_1,w)} + e^{-d(c_2,w)} - e^{-(d(c_1,w)+d(c_2,w))}$; this means that $b_1 \cup_{pr} b_2(w)$ is high if w is close to c_1 or (not exclusive) from c_2.
$b_1 \cap_{pr} b_2\ (w) = e^{-(d(c_1,w)+d(c_2,w))}$ which means that $b_1 \cap_{pr} b_2\ (w)$ is high if w is similaneously close to c_1 and c_2.
We also have by definition $\overset{*}{b_1}(b_2) = <q_1, q_2> = \frac{1}{\text{card } O} \sum_{v} e^{-(d(c_1,v)+d(c_2,v))}$ which means that $\overset{*}{b_1}(b_2)$ will be high if there are many $v \in O$ simultaneously close to c_1 and c_2.

In this case a problem of fitting a set of symbolic objects A by maximizing a criterion without decomposition may be set in the following way :
Find $b = [y = e^{-d(c,.)}]$ which maximizes :

$$W(b) = \sum_{a \in A} b^*(a) = \frac{1}{\text{card } O} \sum_{w_i \in \Omega_1} \sum_{v \in O} e^{-(d(c,v)+d(w_i,v))}$$

$$= \frac{1}{\text{card } O} \sum_{v \in O} e^{-d(c,v)} \sum_{w_i \in \Omega_1} e^{-d(w_i,v)}$$

This means that b and is associated curve C, is a good solution when the more a $v \in O$ is "close" to the w_i the more it is close to the curve c.
A problem of fitting A by a generalization of a decomposition may be set in the following way :
Find $b = [y = e^{-d(c,.)}]$ which maximizes :

$W(B,P) = \sum_{j=1,k} D(b_j, P_j)$ where $B = (b_1, \ldots, b_k)$, $P = (P_1, \ldots, P_k)$ is a partition and
$D(b_j, P_j) = \sum_{a \in P_j} \overset{*}{b_i}(a)$; therefore :

$$W(B,P) = \sum_{j=1,k} \sum_{a \in P_j} \overset{*}{b_j}(a)$$

$$W(B,P) = \sum_{j=1,k} \left(\frac{1}{\text{card } O} \sum_{w_i \in \Omega_1^j} \sum_{v \in O} e^{-(d(c_j,v)+d(w_i,v))} \right.$$

where Ω_1^j $\{w \in \Omega_1 / a_w \in P_\ell\}$; we have

$$W(B,P) = \frac{1}{\text{card } O} \sum_{j=1,k} \sum_{v \in O} e^{-d(c_j,v)} \sum_{w_i \in \Omega_1^j} e^{-d(w_i,v)}$$

This means that B is a good solution if in each class P_ℓ the more a $v \in O$ is close to the w_i, the more it is close to the curve c_ℓ associated to b_ℓ.

13. Symbolic objects representation by categories and fractals

13.1. Categorical representations

The category theory was introduced by Filenberg and Maclane (1945) in algebraic topology in order to study geometric and algebraic interrelationships. The definition of a category (see for instance, Hydeheard and Burstall (1988)) is the following :

Definition

A category is a graph ($\mathcal{a}$, E, s, t) whose nodes $\mathcal{a}$ we call objects and whose edges E we call arrows. Associated with each object a in $\mathcal{a}$, there is an arrow $i_a : a \rightarrow a$, the identity arrow on a, and two pairs of arrows $f : a \rightarrow b$ and $g : b \rightarrow c$, there is an associated arrow $g\, f : a \rightarrow b$, the composition of f with g. The following equations must hold for all objects a, b, c and arrows $f : a \rightarrow b$, $g : b \rightarrow c$ and $h : c \rightarrow d$:

$$(h\, g)\, f = h(g\, f)$$
$$f\, i_a = f = i_b\, f.$$

Examples of categories

a) Partial orders

It is then easy to show that a partial order is a category. For instance, if we consider the four symbolic objects {a, b, c}, and the partial order $a \leq b \leq c$, $a \leq d$, we have a category defined by $\mathcal{a} = \{a, b, c\}$, $E = \{f/a \leq b \Rightarrow f : a \rightarrow b\}$; it is possible to represent this category by the graph given in figure 15.

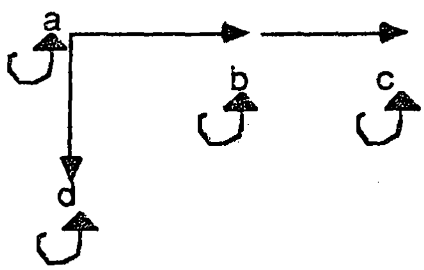

Figure 15

b) Pyramids

A pyramid see Diday (1986), Bertrand Diday (1991) is a category where a is the set of individuals or classes of individuals associated to each mode of the pyramid and E = $\{f/a \subseteq b \Leftrightarrow f : a \rightarrow b\}$.

c) Lattice of complete symbolic objects

In the same way such a lattice (see § -) may be considered to be a category where a is the set of complete objects associated to each node of the lattice and $E = \{f/a \leq b \Leftrightarrow f : a \rightarrow b\}$. It is also possible to show that a pyramid of symbolic objects see Brito Diday (1989) is in the same way, also a category.

Definition of a functor

A functor $F : A \rightarrow B$ from category B is a pair of functions :

a) F : Obj (A) (i.e. the set of objects of the category A) $\rightarrow$ Obj (B) $F : a \rightarrow F(a)$
b) F : Arrow (A) $\rightarrow$ Arrow (B), $F : (f : a \rightarrow b) \rightarrow F(f) : F(a) \rightarrow F(b)$. Satisfying $F(i_a) = i_{F(a)}$ and $F(g\,f) = F(g)\,F(f)$ whenever gf is defined.

Examples of functions :

Both a pyramid of complete symbolic objects (Brito, Diday (1989)) and its geometrical representation in the plane, are categories. The mapping which associates to the algebraic definition a pyramid of complete symbolic objects its geometrical representation is a function. There is also a functor for relating the algebraic properties of a lattice of complete symbolic objects and its geometrical representation (see Wille (1980), Duquesne (1986)).

13.2. Fractal representation

By using duality (see section 9.2) the objects defined on individuals at one level become the individuals at the next level ; if this transformation has some regularity the sequence of dualities defines a fractal. For instance, we may imagine that several distributions (birth weight, size etc. of babies, for example) vary in the same way from local district to county and from county to regions etc. ; in this case the first set of symbolic assertion denoted a describes local district, for instance, by : $ld_j = \hat{}_{i\,pr}\ [y_i = q_j^i]$; if we consider only the birth weight, in order to simplify, we get $ld_j = [y = q_j]$ where q_j is the distribution of the birth weight of the babies born in the j th local district; hence, ld_j is defined on Ω, the set of babies born for instance, in 1992. If weight

(w) is the weight of the baby $w \in \Omega$, we have $ld_j(w) = q_j$ (weight (w)) ; (this comes from the fact that if we set $w^s = [y = r_w]$ where r_w (weight (w)) = 1 and $r_w(t) = 0$ if $t \neq$ weight (w) we get by definition of probabilistic objects :
$ld_j(w) = \langle q, r_w \rangle = \Sigma\{q(v)\, r_w(v) / v \in O\} = q(\text{weight}(w))$ were O is the set of possible birth weigths). Let Q be the set of probability distributions $q : O \rightarrow [0,1]$; and = $\{ld_j / ld_j = [y = q_j], q_j \in Q\}$ the set of probabilist objects associated to a local district ; at a higher level, let be $\mathcal{A}^*$ the set of probabilist objects associated to departments, such that

$$\mathcal{A}^* = \{a^*/a^* : \mathcal{A} \rightarrow [0,1],\ y^* : \mathcal{A} \rightarrow Q,\ a^* = [y^* = q_j^*]\} ;$$

for instance $a_j^* = [y^* = q_j^*]$ describes the j th department by the birth weight distribution $q_j^* : O \rightarrow [0,1]$ of the local district contained in it. Hence, we get :

$$a_j^*(par_i) = \langle q_j^*, q_i \rangle = \Sigma\{\, q_j^*(v)\, q_i(v) / v \in O\}.$$

We may continue the process by defining regions of departments by a set of probabilistic objects

$$\mathcal{A}^{**} = \{a^{**}/a^{**} : \mathcal{A} \rightarrow [0,1],\ y^{**} : \mathcal{A}^* \rightarrow Q,\ a^{**} = [y^{**} = q^{**}]\}$$

and so on. We get a prefractal of order 2 if there exists a mapping $k : \mathcal{A}_{pr} \rightarrow \mathcal{A}_{pr}$ such that $h(a) = a^*$ and $h(a^*) = a^{**}$ (which means that there is a mapping $h_1 : Q \rightarrow Q$ such that $h_1(q) = q^*$ and $h_1(q^*) = q^{**}$). If instead of representing only the babies birth weights, O represents also this size (i.e. $O = O_1 \times O_2$ where O_1 and O_2 are respectively the sets of possible weights and size) figure 16 represents a prefractal of order 2.

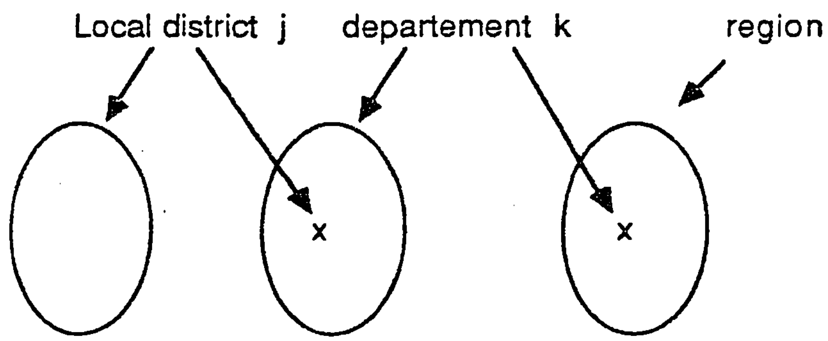

Figure 16 : a prefactal of order 2 of probabilist objects $a_j = [y = q_j]$, $a_k^* = [y^* = q_k^*]$ and $a_\ell^{**} = [y^{**} = q_\ell^{**}]$, where q_j, q_k^*, q_ℓ^{**} represents the probabilist distribution of birth weight and size of babies, in a local district, department and region such that there exists $h_1 : h_1(q_j) = q_k^*$ and $h_1(q_k^*) = q_\ell^{**}$.

More generally, by setting $n* = ** \ldots *$, n times, we obtain a "symbolic fractal" when we have a sequence $u_n = a^{n*} = \hat{\wedge}_x \{y_i^{n*} = q_i^{n*}\}$ where $a^{n*} : a_x^{n*} \rightarrow [0,1]$ and $Q_i^x = \{q\}$, where q is a mapping $O \rightarrow [0,1]$ associated to x ; $y_i^{n*} : a_x^{(n-1)*} \rightarrow Q_i^x$, such that there exists a mapping

$h : a_x^{(n-1)*} \rightarrow a_x^{n*}$ which satisfies : $h(a^{(n-1)*}) = a^{n*}$; sometimes h may be decomposed in $h = \{h_i\}$ such that $\forall n, \forall i, h_i(q_i^{(n-1)*}) = q_i^{n*}$.

Let F_p be the set of prefractals of order p, in other words $F_p = \{B = (w, b, \ldots, b^{(p-1)*}) / B \in (\Omega, a_x, \ldots, a_x^{(p-1)*}) / \exists\, h : a_x^{(n-1)*} \rightarrow a_x^{n*}$, such that $h(b^{(n-1)*}) = b^{n*}\}$.

It is then possible to define the extension of a prefractal of order p, denoted $A_p = (a, a^*, \ldots, a^{p*})$ by

$$\text{Ext}\,(A_p / F_{p-1}) = \{(B, A_p(B)) / B \in F_{p-1}, A_p(B) = (a(w), a^*(b), \ldots, a^{p*}(b^{(p-1)*}))\}$$

An extension of A_p at a level α is defined by

$$\text{Ext}(A_p / F_p, \alpha) = C \text{ with } C = \{B \in F_{p-1} / \text{Min}\,\{a(w), a^*(b), \ldots, a^{p*}(b^{(p-1)*})\} \geq \alpha\}.$$

It is also possible to define probabilist, possibilist and belief union, intersection etc. between prefractals by setting : $A_1 *_x A_2 = (a_1 *_x a_2, a_1^* *_x a_2^*, \ldots, a_1^{p*} *_x a_2^{p*})$.

There is a wide class of applications of sýmbolic fractals in all domains where there is a known organisation of the individuals, classes of individuals, classes of classes and so on, when all units are defined by symbolic objects related together by a mapping f ; for instance, when we have an official organisation of geographical regions in official statistics, in order to study pollution in the air, in forests, in the water or to detect variations of species of insects and frequence of illnesses or when we have an official division of departments in a company, by using the frequency of the words used at each level, in order to study the quality of the division. In all these cases symbolic fractals constitutes the representation of an ideal situation and so they give the possibility of detecting anomalies.

Example :

A region is divided into n departments which are divided themselves into m districts which may be also divided and so on ; our aim is to give a geometrical fractal representation of statistical informations by associating to each unit (for instance a district or a department) or a set of squares which contains the distributions associated to this unit. In order to do so (see figure 16) we have to combine two fractals : a geometrical one (the squares) and a statistical one : (the probability distributions of birth weights, for instance). More formally, this combination may be represented in the following way : at the first step : the ith district is represented in $a_i = [y_1 = q_i] \wedge_x [y_2 = t_i]$ which is a mapping $\Omega = \Omega_1 \times \Omega_2 \rightarrow [0,1]$ where Ω_1 is the set of babies born in 1992 and $\Omega_2 = R^2$, O_1 is the set of possible weights, $O_2 = R^2$ $y_1 : \Omega_1 \rightarrow Q_1 = \{q/q : O_1 \rightarrow [0,1]$ is a probability distribution of the birth weights$\}$; $y_2 : \Omega_2 \rightarrow Q_2 = \{q/q : O_2 \rightarrow \{0,1\}$, $q(v) = 1$ if v is on the geometrical figure (a square, for instance) represented in q$\}$; here, $f_x (L_1, L_2) = L_1.L_2$. Hence $a_i (w_1, w_2) = q_i (\text{weight} (w_1)).t_i(w_2)$; we may associated to a_i, a method denoted [method i] which says that the distribution q_i must be represented in the geometrical figure defined by t_i.

At step n, we have : $a_j^{n*} = [y^{n*} = q_j^{n*}] \wedge [y_2^{n*} = t_j^{n*}] \wedge [\text{method}j]$; the sequence $\cup_n = a^{n*}$ is defined by the mappings h_1 and h_2 such that $q_j^{(n+1)*} = h_1 (q_{j1}^{n*}, q_{j2}^{n*}, \ldots, q_{j5}^{n*})$ and $t^{(n+1)*} = h_2 (t_{i1}^{n*}, \ldots, t_{i5}^{n*})$ where h_1 associates to five distributions representing 5 districts, a mean distribution which represents a department ; if a distribution is too far from this mean, we say that there is an anomaly ; h_2 associates at the beginning to five squares, representing 5 districts, a department represented by 5 squares of equal sizes and centered at points : (0,0), (1,0), (0,1), (-1,0), (0,-1) ; at step 2, h_2 associates to 5 departments a region represented by 5^2 squares in such a way that each department is centered in (0,0), (2,0), (0,2), (-2,0), (O,-2), at step p, h_2 associates to $t_{i1}^{(p-1)*}, \ldots, t_{i5}^{(p-1)*}$ a figure of 5^p squares composed by 5 times, 5^{p-1} squares each one centered in (0,0), $(n^p,0)$,$(0,n^p)$, $(-n^p,0)$,$(0,-n^p)$. At step p the method j consists of representing in the central square (i.e. of center (0,0)) of the 5^p squares the ditribution q_ℓ^{p*} which is the closest to the mean of the q_{ji}^{p*} , $i \in \{1, \ldots, 5\}, i \neq j$ and to order the 4 other distributions according to their similarity, in the other central squares (i.e. of center $(n^p,0)$, $(0,n^p)$, $(-n^p,0)$, $(0,-n^p)$).

In figure 17 we represent a region of 5 departments on the left in the ideal situation of a regular fractal ; on the right, when some anomalies appear ; in this case the corresponding squares or 5^p squares are distributed from their center on their axis.

Notice that the method may be extended easily to the case where the number of districts, departments, regions etc. are not equal ; it suffices to replace the number 5 by the number ℓ = Max (card (districts), card (departments), card (regions), ...) and to omit at each level the useless squares in a regular way.

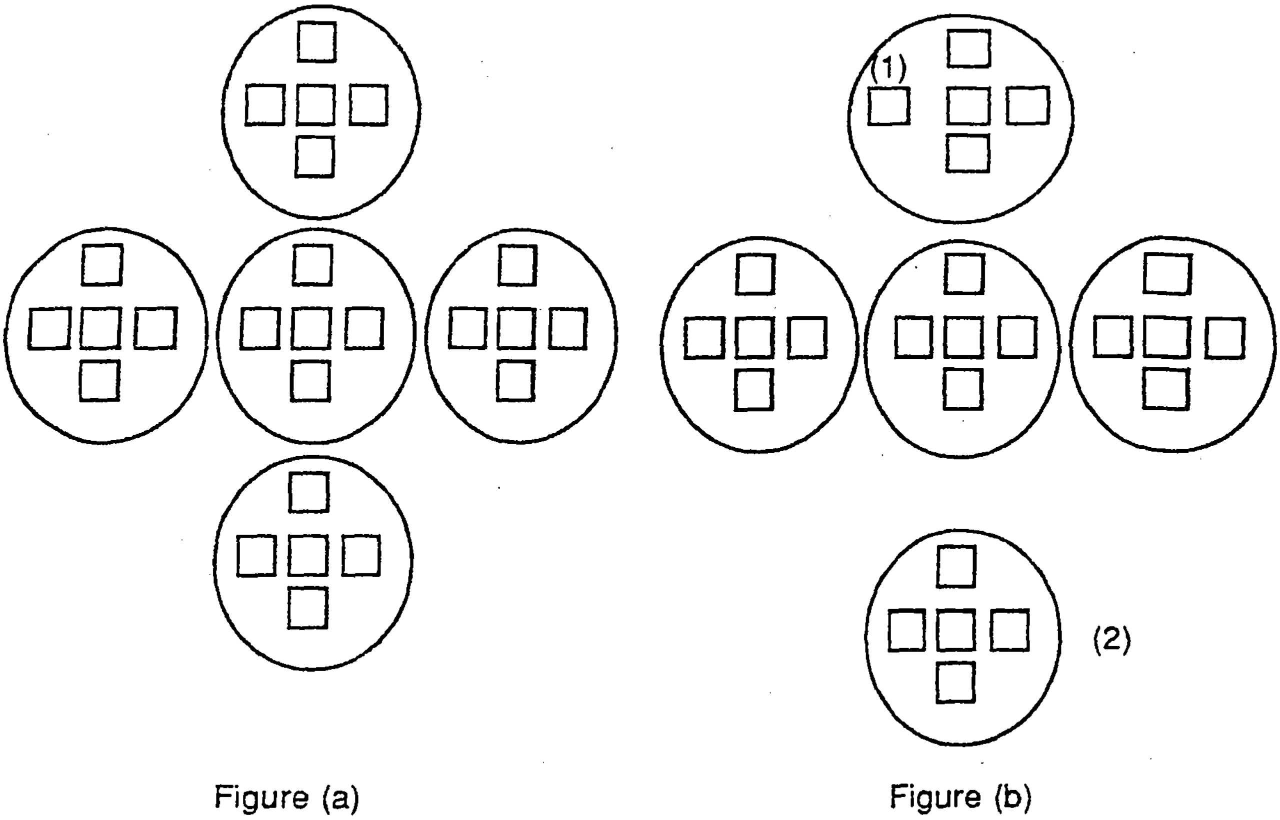

Figure 17 : 25 districts, 5 departments and 1 region represented in a) by an ideal fractal ; in b) with an anomaly of a district (top-left) and of department at the bottom.

13.3. Fractals and categories

There are several ways of making a link between fractals and categories ; a first way may happen when at a given step, categories of units remain at the next step of duality the same category . For instance, at the first level, a pyramid of villages of a omit is transformed in the useless pyramid but of districts at the level of a department.

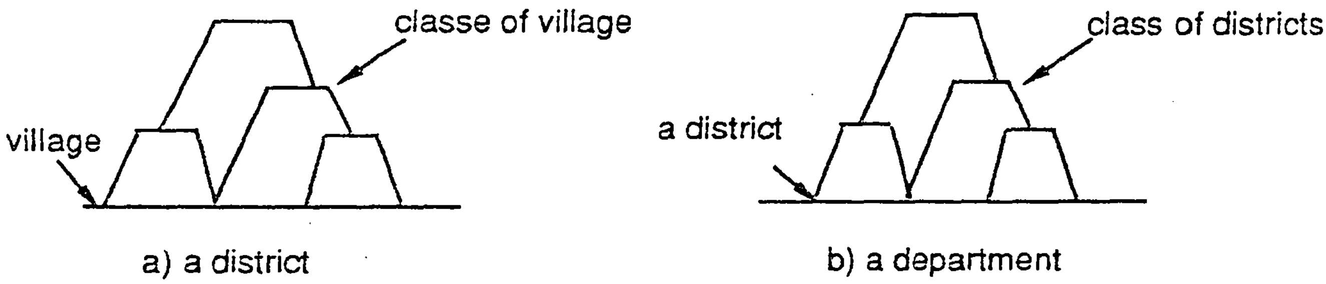

Figure 18 : A first possible link between categories and fractals.

A second way may be described by a square of symbolic objects $u_p = a^{p*} = \bigwedge_i [y_i^{p^*} = q_i^{p^*}]$ where $q_i^{p^*}$ is a probability distribution of categories $q_i^{p^*} = h_i\,(q_i^{(p-1)^*})$.

Example : fractal decomposition of a sequence of letters.

We consider a large sequence of units where each unit is a letter taken from a finite set of letters L, each $q_i^{p^*}$ is associated to intervals of the sequence (called windows) of size L and represents the probability distribution of a type of category computed on a sequence of p_i of such windows. For instance, if

i) q_2 is associated to windows of size L two,

ii) L = {A, C, G, T} (as in a sequence of DNA in the genoma)

iii) the sequence is AT CG CC AG GT CC CA CG TT and we consider only hierarchical categories with $p_i = 3$, we obtain the hierarchies H_a, H_b, H_c of figure 19, respectively associated to the sequences of windows a) AT CG CC, b) AG GT CC, c) CA CG TT ; for

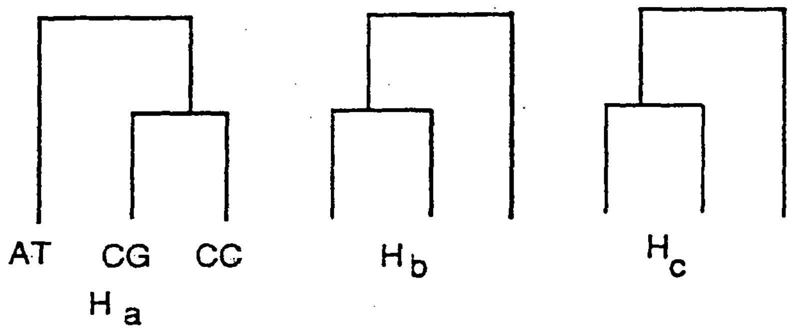

Figure 19 : Hierarchies associated to three sequences of three windows of size 2.

instance, the first hierarchy H_a merges first CG and CC which have 1 difference (the letters G and C) at level 1, then at level 2, AT is merged with CG $\cup$ CC because there are two differences between those two classes. Finally, q_i represents the ditribution of categories defined by Prob $(H_a) = \frac{1}{3}$ and Prob (H_b) = Prob $(H_c) = \frac{2}{3}$. Therefore in this way we get the following probabilist event : $a = [y = q_i] = [y = \frac{1}{3} H_a, \frac{2}{3} H_b, \frac{2}{3} H_c]$.

In order to obtain a fractal on a given sequence, it is possible for instance to start from a large region represented by a_ℓ^{**} ; then dividing it into two or more parts in order to get departments represented by $a_k^* = a_\ell^{**}$, such that $h : h(a_k^*) = a_\ell^{**}$ exists and repeat the process with departments, districts and so on until anomalies appear (i.e. h doesn't exist) in some part ; in these parts new fractals may be built with other h_i, and (or) by omiting some of the $[y_i^{p^*} = q_i^{p^*}]$, etc.

14. Stages of a Symbolic Data Analysis

Roughly speaking, we may characterize a Symbolic Data Analysis by the following steps :

1) Start from a set of more or less complex individual objects .

2) Build classes from a clustering, a factorial analysis, a category (lattice, hierarchy, pyramid, etc.).

3) Describe these classes in order to obtain metadata associated to each class ; these metadata may be given directly by experts, thereby dropping steps 1 and 2.

4) Build for each class a symbolic object associated to these metadata (i.e. define f_x, g_x, $\cup_x$, $\cap_x$, C_x) in order to be able for instance, to compute their extension.

5) Analyse, synthetise, classify, discriminate, organize by different methods of symbolic data analysis the set of symbolic objects obtained at step 4).

6) Extract from step 5 metaknowledge (knowledge on knowledge) for instance rules relating symbolic objects extrected from a category built in step 5.

15. An example of application in road transportation

In the French National Institute of Transportation (INRETS), D. Fleury, C. Fline and J.F. Peytavin have designed "Scenarios of accidents" in order to improve the French road network.
The scenarios are expressed by sentences which describe some characteristics of the user, the kind of displacement of the vehicle, the scene, the moment, the place etc.

Example :

"Man between 30-50 years old, loosing the control of his vehicle (local user, experimented, often drunk), accident happening during the day".
The scenarios are based on the experience of the experts, by working on the field and also by using large data bases of accidents. Under the direction of D. Fleury (INRETS) and M. Gettler-Summa (Paris 9 - Dauphine University), A. Regnier has applied several steps of a Symbolic Data Analysis approach in order to improve, complete and organize the knowledge base of scenarios given by the experts. In this work, each scenario is designed as a symbolic object represented by a probabilist assertion.

The following steps have been used :

a) Expressing the expert's scenarios by probabilist assertions.

Example :

Scenario = [day={70% monday,30% sunday}] $\wedge_{pr}$ [road condition={snow-covered,icy}] $\wedge_{pr}$ [road signs = {60% step, 40% give way}]

b) Compute the extensions of the probabilist assertions associated to each scenario.

Example :

A scenario is described by :

scen = [day = {70% monday, 30% sunday}] $\wedge_{pr}$ [time = 7-9am]

An accident in the data base is defined by :

acc = [day = {monday}] $\wedge$ [hour = 8]

By applying the definition of probabilist objects we get

scen (acc) = $(70 \times 1 + 30 \times 0 + 100 \times 1)/2 = 85\%$

If the experts decide that we have a prototype for a probabilist assertion a = $\wedge_{i=1,n} [y_i = q_i]$ when $a(w) \geq \frac{1}{n} \sum_{i=1,n} \text{Max} \{q_i (v)/v \in O_i\}$, we may say that acc is a prototype of scen ; as $\frac{1}{n} \sum_{i=1,n} \text{Max} \{q_i (v)/v \in O_i\} = \frac{70 + 100}{2} = 0.85 = a(w)$.

c) Improving the scenarios given by the experts.

The extension in the data base of some scenarios appears to be too large, other are too small ; in the first case the experts have to add some events ; in the second case they have to drope some of them.

d) Building new scenarios.

The study has concerned 579 accidents in a French department (Eure et Loir) ; from this information 12 scenarios were created by the experts ; the union of the extension of the associated probabilist assertions covered 286 accidents ; hence we had to induce new probabilist assertions from the 293 remaining accidents, in order to get a knowledge base of scenarios whose extension covers as well as possible the 579 accidents of the data base. A K-means clustering algorithm applied to the 293 accidents made it possible to builts 14 clusters whose intension provided 14 probabilist assertions. Notice that also a pyramidal clustering building at each step a probabilist assertion could also be done (see Diday et al (1992)).

e) Organisation of the scenarios.

Steps a) b) c) d) has finally provided a knowledge base of 26 scenarios, each one represented by a probabilist assertion whose extension covers 93.4% of the data base of accidents. In order to get a synthetic organisation of this base a pyramid of symbolic objects have been summerized by the experts. A piece of this pyramid is provided in figure x which corresponds to accidents due to collision.

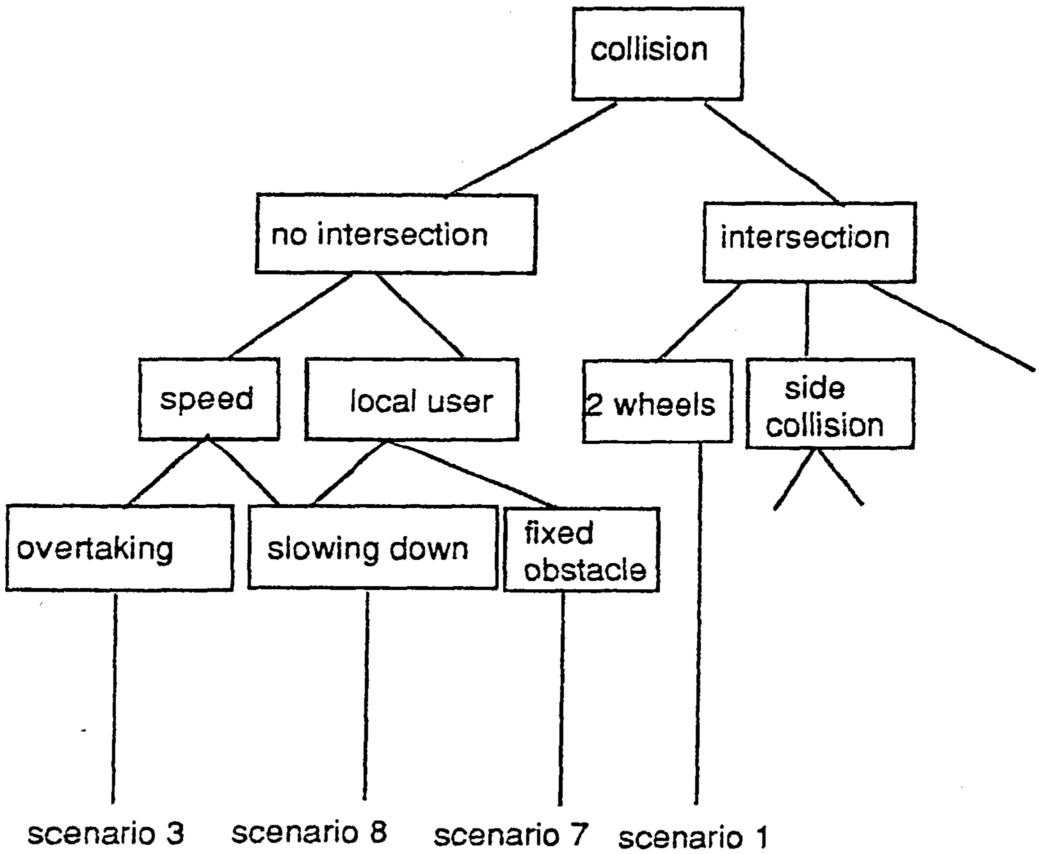

Figure 20 : Organisation of scenarios by a pyramid of symbolic objects concerning accidents due to collision.

f) Metaknowledge.

The organisation of the scenarios obtained at step e) has already provided knowledge on the knowledge base of scenarios. Other kinds of metaknowledges may also be provided ; for instance, the scenarios have been partitioned in two classes : the "strong" (or "prototypes") and

the weak ; a scenario is considered to be "strong" when it has a large extension even if it is defined by a large number of events; if not, it is called "weak". It turns out that the initial 12 scenarios provided by the experts were strong, which has validated their study.
Finally this study has confirmed, completed, organised the scenarios and provided a better knowledge of the structure of the accidents in Eure et Loire. We are now studying the robustness of the results for the other French departments. The next step would be to study the existence of fractals associated to each scenario on geographical zones of growing size (villages, districts, departments, region etc.), in order to detect anomalies.

16. Symbolic objects with a mixture of semantics

16.1. Definition of mixed symbolic objects

This kind of object may be written $a = \hat{\uparrow}_i a_i$ where a_i may be a boolean, possibilist, probabilist or belief object. There are several ways to define it ; in the first way, $a = \hat{\uparrow}_i a_i$ is a mapping $\Omega \rightarrow [0,1]$ such that $a(w) = \prod_i a_i(w)$ where each $a_i(w)$ is computed according to the semantic associated to x_i ; to do so it must be supposed that w^s is a boolean symbolic object taking a single value with weight 1 for each variable y_i.

The second way, closer to natural language considers $\text{sent}(a) = \hat{\uparrow}_{x_i} a_i$ as a mapping from Ω in $L = \prod_i L_i$ name (a) where each L_i is a set of words associated to the semantic x_i and name (a) is the name given to a ; such that $\text{sent}(a(w)) = \prod_i \ell_i$ name (a) where $a_i(w) = \ell_i$. This kind of object is illustrated by an example given in 16.3.

16.2. Monothetic, polythetic and prototypic aspects of a mixed symbolic object

By representing a class by a mixed symbolic object defined by $a = \hat{\uparrow}_i a_i : \Omega \rightarrow [0,1]$, where a_i may be a boolean, possibilist, probabilist or belief object, we obtain at the same time a description which is

i) monothetic at a given level α, in terms of conjunction of properties such that $a(w) = \text{true}$ iff $\forall i \; a_i(w) \geq \alpha$;

ii) polythetic, as the properties defined by a modal symbolic object expresses tendances or ideal situations to be attained more or less by the individuals of the class ; hence, it has no sense to say that they are sufficient or necessary conditions to be satisfied ;

iii) prototypic, by giving the possibility to compute individuals or symbolic objects which best satisfy the mixed symbolic object (see 15 and the following example).

Example :

We define a set of scenarios of accidents by the following kind of mixed symbolic objects :

$$\text{scen}_j = \underset{i}{\wedge}\text{pr}\,[y_i = q_i^j] \wedge \underset{\ell}{\wedge}\text{pos}\,[y_\ell = q_\ell^j] \wedge \underset{k}{\wedge}\text{cr}\,[y_k = q_k^j]\,;$$

if we denote $a_1 = \underset{i}{\wedge}\text{pr}\,[y_i = q_i^j]$, $a_2 = \underset{\ell}{\wedge}\text{pos}\,[y_\ell = q_\ell^j]$ and $a_3 = \underset{i}{\wedge}\text{cr}\,[y_i = q_i^j]$ we get $\text{scen}_j\,(w) = a_1\,(w) \wedge a_2\,(w) \wedge a_3\,(w)$ if $a_i\,(w) \in [0,1]$; we may also define a sentence by the following way :
$\text{sent}(\text{scen}_j(w)) = \ell_1\,\ell_2\,\ell_3$ name (a) if ℓ_1 is a word related to probability, ℓ_2 to possibility and ℓ_3 to belief. More precisely, if

$$a = [\text{place} = 0.9 \text{ town}, 0.1 \text{ suburbes}] \wedge [\text{time} = 1 \text{ pm}, 0.3 \text{ am}]_{pos}$$

where the place is described by probabilist mappings and the time by possibilist mappings ; then for w = [place = suburbes] $\wedge$ [time = pm] we get scen (w) $= 0.1 \times 1 = 0.1$; if we associate the word ℓ_1 = "rare" to values of a probabilistic object a_{pr} such that $a_{pr}\,(w) < 0.2$, the word ℓ_2 = "possible" to the values taken by a possibilist object a_{pos} such that $a_{pos}\,(w) > 0.9$ and to the object a, the phrase "citizen scenario", we get sent(a(w)) = ℓ_1 name (y_1) ℓ_2 name (y_2) name (a) = rare place possible time citizen scenario.

Finally, in this case a prototype may be defined by

$$w_{pro} = \underset{i}{\wedge}\text{pr}\,[y_i = \arg\max_v q_i\,(v)] \wedge \underset{\ell}{\wedge}\text{pos}\,[y_i = \arg\max_v q_\ell\,(v)] \wedge \underset{k}{\wedge}\text{cr}\,[y_k = \arg\max q_k\,(v)]$$

Therefore in our example we get :

$$w_{pro} = [\text{place} = \text{town}] \wedge [\text{time} = 1 \text{ pm}].$$

CONCLUSION

Considering a data base (Ω,Δ') where any individual object $w\in\Omega$ 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 technic) ; 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 informations. 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, 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 individuals objects ; for instance, theorem 2 shows that in case of probabilities a^* defined on $\mathfrak{a}_{pr}$ (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 by 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 $A_x = \cup_x \{q/q\in A\subseteq\mathfrak{a}_x\}$ is called x- set, then in the case of possibilities (x=pos) A_{pos} is a fuzzy set in the original sense given by Zadeh (1985) ; 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 kind of operators we will have to try to satisfy stability. Several computer programs of symbolic data analysis have been already implemented independently, see for instance in this issue : histograms of symbolic objects (De

Carvalho 1991), symbolic pyramidal clustering (P. Brito). Decision tree on symbolic objects (C. Jacq), extracting rules from a special kind of symbolic objects (M. Sebag). More generally in the framework of the Esprit II program MLT ("Machine learning toolbox") an interface between Makey (Lebbe, Vignes (1990)) and SICLA (Celeux et al 1989) 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, belief and possibilist objects, for instance) and new kinds of queries, expressed by a modal assertion a_x, when the extension is composed by individual objects or by dual modal assertion a_x^* when the extension is computed on a set of assertions $A_x \subseteq a_x$.

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 (house, 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, SDA may also introduce powerful methods (as factorial analysis), among many others, widely used in data analysis but neglected in 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

$a^*(\mathfrak{a}_x) = f_x(\{g_x(q_i, \{\bigcup_{j} {}_x\, q_i^j \,/\, q_i^j \in Q_i^{\mathfrak{a}_x}\})\}_i)$; where, by definition $\mathfrak{a}_x$ is the set of im assertions associated to x and $Q_i^{\mathfrak{a}_x} = \{q_i^\ell / a_\ell = \hat{\wedge}_i\, [y_i = q_i^\ell] \in \mathfrak{a}_x\} = Q_i^x$ the set of any $\cup_x$, $\cap_x$, c_x combination of elements $q_i^j \in Q_i$ associated to x. Hence, we have $a^*(\mathfrak{a}_x) = f_x(\{g_x(q_i, \bigcup_{j} {}_x \{q_i^j \,/\, q_i^j \in Q_i^x\})\}_i)$. We set $1_a = \hat{\wedge}_i [y_i = 1_{O_i}]$ where $1_{O_i}(v) = 1\ \forall v \in O_i$. We denote $q_i^A = \{\cup_x\, q/q \in Q_i^{Ax}\}$, where $A_x \subseteq \mathfrak{a}_x$ and Q_i^{Ax} is defined as in §9 by $Q_i^{Ax} = \{q_i/a = \hat{\wedge}_i [y_i=q_i] \in A_x\}$ which means that Q_i^{Ax} 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 $\forall u_j \in \mathbb{R}$, $\bigcup_{j=1,n} {}_x\, u_j = u_1 \cup_x u_2 \ldots \cup_x u_n$, where in the case of possibilities we have $u_1 \cup_p u_2 = \mathrm{Max}(u_1,u_2)$ 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.

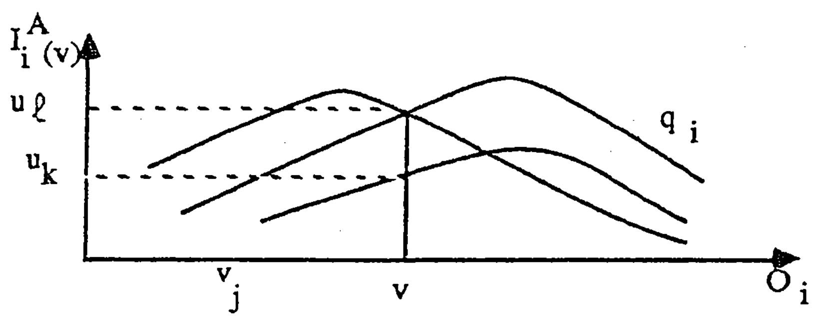

Figure 21 : When q_i varies in Q_i^{Ax}, $q_i(v) = u_\ell$ is repeated each time that it exists a different $q_j \in Q_i^{Ax}$ such that $q_j(v) = q_i(v) = u_\ell$.

Lemma 1

If for any sequence$\{u_n\}$ of rational numbers dense in $I_i^A(v)$, the sequence

$U_n = \bigcup_{\ell=1,n} u_\ell$ converges towards the same limit U, then $q_i^A(v) = U$.

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 $\frac{1}{2^k}$; 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 $\frac{1}{2^k}$ 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 $\mathbb{R}$, it is possible to decompose the intervals defined by its boundaries, also by a partition of lenght $\frac{1}{2^k}$; we associate to each of these intervals which contains at least one element of O_i, a rationnal number from this interval. The set of these numbers is denoted O_i^k and the elements of O_i^k are denoted $v_1, v_2, \ldots, v_{n_k}$, with n_k = card Q_i^k ($\leq 2^k$ 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 $I_A^k = I_A^k(v_1) \times \ldots \times I_A^k(v_{n_k})$ and we suppose that $v_\ell \in \{v_1, \ldots, v_{n_k}\}$; we associate to any $r = (r_1, \ldots, r_{n_k}) \in I_A^k$ 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 : $C_{r_\ell}^k(v_\ell) = \{q_r \in Q_i^A / r = (r_1, \ldots, r_{n_k}) \in I_A^k, r_\ell$ fixed, $q_r(v_j) \in I_{r_j}^k(v_j)$, q_r 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 time 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 $q_i^k(v) = \bigcup_{u_\ell \in S_k} \times\, u_\ell^k = \bigcup_{\ell = 1, n_k} \times\, u_\ell^k$ 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 $u \in I_i^A(v)$, $\exists\, u_\ell^k \in I_A^k(v) \subseteq S^k$ such that
$|u_\ell^k - u| \leq \frac{1}{2^k}$. 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 it 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 :

$U = \bigcup_{u \in I_i^A(v)} \times \bigcup_{q \in C_i^u(v)} \times\, u$ 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 $|u_k - u| \leq \frac{1}{2^k}$ 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, $\forall\, v \in O_i$, $\exists\, v_k \in O_i^k$, such that $|v_k - v| \leq \frac{1}{2^k}$.

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 $v_i \in O_i^k$ $|q_k(v_i) - q(v_i)| \leq \frac{1}{2^k}$ as by construction the length of the interval $I_{r_i}^k$ is $\frac{1}{2^k}$; 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 $q_i^k(v) = \bigcup_{x\, u\in I^{A}_k(v)} \bigcup_{x\, q\in C^k_u(v)} u$, we get :

$$U = \bigcup_{x\, u\in I^{A}(v)} \bigcup_{x\, q\in C_i^u(v)} u = \bigcup_x \{q/(v)/q \in Q_i^A\} = q_i^A(v). \blacksquare$$

We recall that 1_{O_i} and I_i^A are the mapping $O_i \rightarrow [0,1]$ 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 $q_i^A = \{\bigcup_x q/q \in Q_i^A\}$.

Lemma 2

If $\forall\, u_1, u_2 \in [0,1]$, $u_1 \bigcup_x u_2 \geq \mathrm{Max}\,(u_1, u_2)$, then $\forall\, A_x \subseteq \mathfrak{a}_x$ and $\forall v \in O_i$, we have $q_i^A(v) = \mathrm{Max}\,\{u/u \in I_i^A(v)\}$ and $q_i^{\mathfrak{a}_x} = 1_{O_i}$.

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 $\forall n$ $U_n = U_{n-1} \bigcup_x u_n \geq U_{n-1}$, since $u \bigcup_x v \geq \mathrm{Max}\,(u,v)$, and so the sequence $\{U_n\}$ is increasing, as it is majored by $U = \mathrm{Max}\{u/u \in I_i^A(v)\}$. 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 $I_i^A(v)$, it would exist a k such that $u < u_k < 1$ and $U_k = U_{k-1} \bigcup_x u_k \geq u_k > u$, 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 $n \rightarrow +\infty$, $a_n \rightarrow -\infty$ and $b_n \rightarrow +\infty$. We may say that the theorem remains true with $O_i =]-\infty, +\infty[$ since when $n \rightarrow +\infty$ $\forall$ a, b $\in \mathbb{R}$ there exists N_1, N_2 enough large such that the theorem remains true on $]a_{n_1}, b_{n_1}, [$with $a_{n_1} < a$ and $b_{n_2} > b$ for (n_1,n_2) such that $n_1 > N_1$ and $n_2 > N_2$.

Proof of theorem 1 (possibilist objects)

i) $a^*(\mathfrak{a}_p) = 1$, $a^*(\phi) = 0$.

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

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

b) $\forall$ u $\in [0,1]$, $1 \cup_p u = 1 = \mathrm{Max}(1,u) = 1$; $0 \cup_p 0 = \mathrm{Max}(0,0) = 0$; therefore $\forall i$, $q_i^{a_p} = 1_{O_i}$. Therefore $a^*(a_p) = f_p(\{g_p(q_i, 1_{O_i})\}_i) = \mathrm{Max}(\{\sup_{v\in O_i} (\mathrm{Min}(q_i(v), 1_{O_i}(v)))\}_i) = \mathrm{Max}(\{\sup_{v\in O_i} (q_i(v))\}_i) = \mathrm{Max}(\{1\}_i)$ 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 $a^*(a_p) = 1$. By definition we have $a^*(\phi) = f_p(\{g_p(q_i,\phi_i)\}_i)$ with $g_p(q_i,\phi_i) = \sup_{v\in O_i} (\mathrm{Min}(q_i(v),\phi_i(v))) = \sup_{v\in O_i} (\phi_i(v))=0$ as $\phi_i(v) = 0\ \forall\ v\in O_i$. Therefore $a^*(\phi) = \mathrm{Max}(\{0\}_i) = 0$.

ii) $a^*(A_1 \cup_p A_2) = \mathrm{Max}(a^*(A_1),a^*(A_2))$. By definition we have : $A_1 \cup_p A_2 = \wedge_i [y_i = q_i^{A_1} \cup_p q_i^{A_2}]$ with $q_i^{A_k} = \{ \cup_{j} {}_p q_i^j / q_i^j \in Q_i^{A_k} \}$;

Since the assumptions of the lemma 2 are satisfies, q_i^A exists. Hence, we may write :

$a^*(A_1 \cup_p A_2) = f_p(\{g_p(q_i,q_i^{A_1} \cup_p q_i^{A_2})\}_i)$ where $g_p(q_i,q_i^{A_1} \cup_p q_i^{A_2}) = \mathrm{Sup}_{v\in O_i} (\mathrm{Min}\{q_i(v), \mathrm{Max}(q_i^{A_1}(v), q_i^{A_2}(v))\})$. Since $\mathrm{Min}\{a,\mathrm{Max}(b,c)\} = \mathrm{Max}\{\mathrm{Min}(a,b), \mathrm{Min}(a,c)\}$, we have : $g_p(q_i,q_i^{A_1} \cup_p q_i^{A_2}) = \mathrm{Sup}_{v\in O_i} (\mathrm{Max}\{\mathrm{Min}\{q_i(v),q_i^{A_1}(v)\}, \mathrm{Min}\{q_i(v),q_i^{A_2}(v)\}\})$. But since $\mathrm{Sup}_{v\in O_i} (\mathrm{Max}\{a(v),b(v)\}) = \mathrm{Max}\{\mathrm{Sup}_{v\in O_i} (a(v)), \mathrm{Sup}_{v\in O_i} (b(v))\}$, we get : $g_p(q_i,q_i^{A_1} \cup_p q_i^{A_2}) = \mathrm{Max}\ \{\mathrm{Sup}_{v\in O_i} (\mathrm{Min}\{q_i(v),q_i^{A_1}(v)\}), \mathrm{Sup}_{v\in O_i} (\mathrm{Min}\{q_i(v),q_i^{A_2}(v)\})\}$. Hence, as $\mathrm{Max}_i (\{\mathrm{Max}(a_i,b_i)\} = \mathrm{Max}(\mathrm{Max}_i (\{a_i\}), \mathrm{Max}_i (\{b_i\}))$ and $a^*(A_1 \cup_p A_2) = \mathrm{Max}_i (\{g_p(q_i,q_i^{A_1} \cup q_i^{A_2}\})$, by definition, we get :

$a^*(A_1 \cup_p A_2) = \mathrm{Max}\ \{\mathrm{Max}_i \mathrm{Sup}_{v\in O_i} (\mathrm{Min}\{q_i(v),q_i^{A_1}(v)\}), \mathrm{Max}_i \mathrm{Sup}_{v\in O_i} (\mathrm{Min}\{q_i(v),q_i^{A_2}(v)\}\}$ and finally : $a^*(A_1 \cup_p A_2) = \mathrm{Max}(a^*(A_1),a^*(A_2))$.

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 : $u_1 \cup_{pr} u_2 = u_1 + u_2 - u_1 u_2 \geq u_1 + u_2 (1 - u_1) \geq u_1$, and $u_2 + u_1 (1-u_2) \geq u_2$, so that $u_1 \cup_{pr} u_2 \geq \mathrm{Max}(u_1, u_2)$.

b) $\forall$ u $\in [0,1]$, $1\ U_x\ u = 1 + u - u = 1$;

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

Hence it follows from this lemma that $q_i^{\mathfrak{a}_{pr}} = 1_{O_i}$ and $\forall A_x \subsetneq \mathfrak{a}_x$, $q_i^{A_x}$ exists.

Now we may prove theorem 2 :

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

From $q_i^{\mathfrak{a}_{pr}} = 1_{O_i}$ we get : $a^*(\mathfrak{a}_{pr}) = f_{pr}(\{g_{pr}(q_i, 1_{O_i})\}_i)$;

were $g_{pr}(q_i, 1_{O_i}) = \Sigma\ \{q_i(v)\ 1_{O_i}(v)\ /\ v \in O_i\} = \Sigma\ \{q_i(v)\ /\ v \in O_i\} = 1$.

Therefore : $a^*(\mathfrak{a}_{pr}) = f_{pr}(\{1\}_i) = 1$.

By definition we have $a^*(\phi) = f_{pr}(\{g_{pr}(q_i, \phi_i)\}_i)$, 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 $a^*(\phi) = \text{Mean}\ (\{0\}_i) = 0$.

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

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

$a^*(A_1 \cup_{pr} A_2) = f_{pr}(\{g_{pr}(q_i, q_i^{A_1} \cup_{pr} q_i^{A_2})\}_i)$ with

$$g_{pr}(q_i, q_i^{A_1} \cup_{pr} q_i^{A_2}) = < q_i,\ q_i^{A_1} + q_i^{A_2} - q_i^{A_1} \cap_{pr} q_i^{A_2} >$$
$$= < q_i, q_i^{A_1} > + < q_i, q_i^{A_2} > - < q_i, q_i^{A_1} \cap_{pr} q_i^{A_2} >$$

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

$$a^*(A_1 \cup_{pr} A_2) = \text{Mean}(\{< q_i, q_i^{A_1} > + < q_i, q_i^{A_2} > - < q_i, q_i^{A_1} \cap_{pr} q_i^{A_2}\}_i)$$
$$= \text{Mean}\ (\{< q_i, q_i^{A_1} >\}_i) + \text{Mean}\ (\{q_i, q_i^{A_2}\}_i) - \text{Mean}(\{< q_i, q_i^{A_1} \cap_{pr} q_i^{A_2}\}_i)$$
$$= a^*(A_1) + a^*(A_2) - a^*(A_1 \cap_{pr} A_2) .$$

■

REFERENCES

[1] Adanson M. (1757), "*Histoire Naturelle du Sénégal - Coquillages*", Bauche Paris.

[2] Adanson, M. (1763), "*Famille des plants*", Vol. 1, Vincent, Paris.

[3] Arnault A. and Nicole P. (1662), "*La logique ou l'art de penser*", reprinted by Froman, Stuttgart (1965).

[4] Backhoff G. (1967), "*Lattice Theory*", Amer. Math. Soc. Providence (ed. 3).

[5] Barbut M., Monjardet B. (1971), "*Ordre et classification*", T.2 Hachette, Paris.

[6] Beckner (1959), "*The Biological Way of thought*", Columbia University Press, New York, 200p.

[7] Belson (1959), "*Matching and prediction on the principle of biological classification*", Applied Statistics, vol. VIII.

[8] Benzecri J.P. et al (1973), "*L'analyse des données*", Dunod, Paris.

[9] Bochenski I.M. (1970), "*A history of formal logic*", I. Thomas, trans., New York : Chelsea Publishing Co.

[10] Breiman L., Friedman J.H., Olsken R.A., Stone C.S. (1984), "*Classification and regression trees*", Belmont, Wadsworth.

[11] Brito P. and Diday E., "*Pyramidal representation of symbolic objects*", in Knowledge, Data and Computer Assisted Decisions, Schader M. and Gaul W. (ed) NATO ASI serie F : Computer and System Sciences Vol. 61.

[12] Brito P., (1993), "Symbolic objects : order structure and pyramidal clustering", in this issue.

[13] Brito P., Diday E., (1990), "*Pyramidal representation of symbolic objects*", in NATO ASI Series, Vol. F 61, Knowledge Data and computer-assisted Decisions edited by Schader and W. Gaul. Springer Verlag.

[14] Carnap R. (1947), "*Meaning and necessity : a study in Semantic and Modal Logic*", The University of Chicago Press, Chicago.

[15] Celeux G., Diday E., Govaert G., Lechevallier Y., Ralambondrainy H., (1989), "*Classification automatique : environement Statistique et Informatique*", Dunod.

[16] Celeux G., Diebolt J., (1985), "*The SEM algorithm : A probabilist teacher algorithm derived from the EM algorithm for the mixture problem*", Computational Statistics Quarterly 2 p. 73-82.

[17] Chomsky N. (1966), "*Cartesian linguistics : a chapter in the history of rationalist thought*", Harper & Row, New York, French transl. Seuil, Paris 1969.

[18] Choquet G., (1953), "*Théorie des capacités*", Ann. Inst. Fourrier 5. 131-295.

[19] Dale M.B. and Anderson D.J. (1973), "*Inosculate analysis of vegetation data*", Aust. J. Bot. vol. 21, pp.253-276.

[20] Dallwitz (1974), "*A flexible computer program for generating diagnostic keys*", Syst. Zoology, 23 (1), pp. 50-57.

[21] De Carvalho F.A.T. (1991), "Histogramme en Analyse des Données Symboliques", Dissertation Univ. Paris 9 Dauphine.

[22] Dempster A.P., (1967), "*Upper and Lower Probabilites Induced by a Multivalued Mapping* ", Annals of Mathematical Statistics 38, 325-339.

[23] Descles J.P. (1986), "*Travaux de linguistique et de littérature*", XXIV, 1, Strasbourg, Klimcksieck.

[24] Descles J.P. (1991), "*La notion de typicalité : une approche formelle*", in Sémantique et Cognition, pp. 225-244, Editions du CNRS Paris.

[25] Descles J.P. and Kanellos I. (1991), "*La notion de typicalité : une approche formelle*", in Sémantique et Cognition, CNRS Paris, D. Dubois editor.

[26] Diday E. (1971), "*La méthode des nuées dynamiques*", Rev. Stat. Appliquée, vol. XIX, n°2, pp. 19-34.

[27] Diday E. (1976), "*Sélection typologique de variables*", Rapport INRIA.

[28] Diday E. and Simon J.C. (1976), "*Cluster Analysis*", in K.S. Fu (ed.), Digital Pattern Recognition, Springer Verlag, pp. 47-94.

[29] Diday E. et al. (1979), "*Optimisation en classification automatique*", 800p., INRIA (ed.), Rocquencourt 78150 Le Chesnay, France.

[30] Diday E. Lemaire J., Pouget J., Testu F. (1984), "*Eléments d'analyse des données*", Dunod.

[31] Diday E., (1990), "*Knowledge representation and symbolic data analysis*", in NATO ASI Series, Vol. F 61, Knowledge Data and computer-assisted Decisions edited by Schader and W. Gaul. Springer Verlag.

[32] Diday E., (1991), "*Des objets de l'analyse des données à ceux de l'analyse des connaissances*", in "Induction symbolique-numérique à partir de données", Y. Kodratoff, E. Diday, editors, CEPADUES (Toulouse).

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

[34] Diday E., Govaert G., Lechevallier Y., Sidi J. (1980),

[35] Dubois D. (1992), "*Representations catégorielles, prototypes et typicalité*", Le Courrier du CNRS "Sciences Cognitives", N°79, Octobre p.68.

[36] Dubois D., Prade H., (1986), "*A set-theoretic view of belief functions*", International Journal General Systems, Vol. 12, pp 193-226.

[37] Dubois D., Prade H., (1988), "*Possibility theory*", Plenum New York.

[38] Duquenne V. (1986), "*Contextual Implications between attributes and some representation properties for finite lattices*", Beitrage zur Begriffsanalysis Ganter, Wille, Wolf (ed.) Wissensthafts Verlag Mannheim.

[39] Fisher D. and Langley P. (1986), "*Conceptual Clustering and its relation to Numerical Taxonomy*", Workshop on Artificial Intelligence & Statistics. W. Gale (ed) Addison-Wesley.

[40] Ganascia J.G. (1991), "*Charade : apprentissage de bases de connaissances*", Y. Kodratoff and E. Diday (eds) Cepadues.

[41] Geach P. and Black M., "*Translation from the philosophical writtings of Gottlob Frege*", Oxford : Blackwell.

[42] Gower J.C. (1974), "*Maximal predictive classification*", Biomet. vol. 30, pp. 643-654.

[43] Gower J.C. (1975), "*Relating Classification to identification*", in R.J. Pankhurst (editor). Biological Identification with computer. pp. 65-72, London, Academic Press.

[44] Heidegger M. (1662), "*Die Frage nach dem Ding*", Max Niemeyer Verlag, Tübinger. In french "Qu'est-ce qu'une chose ?" Gallimard (1971).

[45] Jambu (1978), "*Classification automatique pour l'analyse des données*", Dunod, Paris.

[46] Jevons W.S. (1877), "*The principles of Science : A treatise on Logic and Scientific Method*", 2nd ed. rev. Macmillan London and New-York, 786p.

[47] Kant E. (1785), "*Fondement de la métaphysique des moeurs*", p.71, Delagrave, Paris.

[48] Kaplan A. and Schott H.F. (1951), "*A calculus for emprical classes*", Méthodes 3, 165-190.

[49] Kruskal J.B., Wish M. (1978), "*Multidimensional scaling*", Sagr., Beverly Hills, Calif.

[50] Latta R. and McBeath A. (1956), "*The element of logic*", London Macmillan (Original work published in 1929).

[51] Lauritzen S.L. and Spiegelhalter D.J. (1988), "*Local computation with probabilities on graphical structures and their application to expert system*".In Readings in uncertain reasoning (1990) edited by G. Shafer and J. Pearl.Morgan Kaufman Publishers.

[52] Le Guyader H. (1988), "*Théorie et Histoire en biologie*", J. Vrin, Paris.

[53] Lebbe J and Vignes R., "*Généeration de graphes d'identification à partir de descriptions de concepts*", in Induction symbolique numérique Y. Kodratoff and E. Diday (editors) Cepadues - éditions.

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

[55] Lebowitz M. (1983), "*Concept learning in a rich input domain*" Proc. of the Machine Learning Workshop, pp. 177-182.

[56] Lebowitz M. (1983), "*Generalisation from natural language text*", Cognitive Science 7, 1, pp. 1-40.

[57] Lerman I.C. (1981), "*Classification et analyse ordinale des données*", Dunod, Paris.

[58] Louis P. (1956), "*Aristote, les parties des animaux*", reprinted by "Les belles lettres".

[59] Michalski R. (1980), "*Knowledge Acquisition Though conceptual clustering : a theoretical framework and an algorithm for partitioning data into conjunctive concepts*", Int. Jour. of Policy Analysis and Information Systems, Vol. 4, N°3.

[60] Michalski R. and Stepp (1983), "*Automated Construction of Calssifications Conceptual*

Clustering versus Numerical Taxonomy", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.5, N°4, July.

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

[62] Michalski R.S., Carbonell J.G., Mitchell T.M., "*Machine learning, an Artificial Intelligence Approach*", Springer Verlag.

[63] Minsky M. (1975), "*A framework for representing knowledge "the psychology of computer vision" *", New York, Mac Graw-Hill.

[64] Panhurst R.J. (1978), "*Biological ientification. The principle and practice of identification methods in biology*", London : Edward Arnold.

[65] Pearl J., (1988), "*Probabilist reasoning in intelligent systems*" Morgan Kaufman, San Mateo.

[66] Quinlen J.R. (1986), "*Induction of decision trees*", Machine Learning 1: pp. 81-106, Kluwer Academic Publishers, Boston.

[67] Ralambondrainy H. (1991), "*Apprentissage dans le contexte d'un schéma de bases de données*", Y. Kodratoff and E. Diday (editors), Cepadues.

[68] Rosch E. (1978), "*Principle of categorization*" in E. Rosch and B. Lloyd (eds), Cognition and Categorisation pp. 27-48, Hillsdale, N.J : Erlbaum.

[69] Rosch E. and Mervis C.B. (1975), "*Family resemblances : studies in the internal structure of categories*" Cognitive Psychology, 7, pp. 573-605.

[70] Roux M. (1985), "*Algorithmes de classification*", Masson.

[71] Schafer G. (1976), "*A Mathematical theory of evidence*" Princeton University Press.

[72] Schafer G., (1990), "*Perspectives on the Theory and Practice of Belief functions*" International Journal of Approximate Reasoning. Vol 4, Numbers 5/6.

[73] Sebagh M., Diday E., Schoenauer M. (1980), "*Incremental learning from Symbolic Objects*", in Knowledge, Data and Computer Assisted Decisions, Schader M. and Gaul W. (ed) NATO ASI serie F : Computer and System Sciences Vol. 61.

[74] Shweizer B., Sklar A. (1960), "*Statistical matric spaces*" Pacific J. Math 10 : 313-334.

[75] Sokal R.R. and Sneath P.H.A. (1963), "*Principle of Numerical Taxonomy*", San Francisco W.H. Freeman (sec. edit 1973).

[76] Sutcliffe J.P. (1992), "*Concept, class, and category in the tradition of Aristotle*", in Categories and concepts theoretical views and inductive data analysis. Academic Press.

[77] Tukey J. (1977), "*Exploratory data analysis*", Addison-Wesley, Reading, Mass.

[78] Wagner H. (1973), "*Begriff*", in Handbuch philosophischer Grundbegriffe, eds H. Krungs, H.M. Baumgartner and C. Wild, Kösel, München 191-209.

[79] Ward J.H. (1963), "*Hierarchical grouping to optimize an objective function*", J. Amer. Stat. Assoc. 58, pp. 236-244.

[80] Wille R. (1989), "*Knowledge acquisition by methods of formal concept analysis*", in : Data Analysis, Learning symbolic and numeric knowledge, E. Diday (Ed.), Nova

Sciences Publishers.

[81] Zadeh L.A. (1971), "*Quantitative fuzzy semantics*". Informations Sciences, 159-176).

Unité de Recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)

Unité de Recherche INRIA Lorraine Technopôle de Nancy-Brabois - Campus Scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 VILLERS LES NANCY Cedex (France)
Unité de Recherche INRIA Rennes IRISA, Campus Universitaire de Beaulieu 35042 RENNES Cedex (France)
Unité de Recherche INRIA Rhône-Alpes 46, avenue Félix Viallet - 38031 GRENOBLE Cedex (France)
Unité de Recherche INRIA Sophia Antipolis 2004, route des Lucioles - B.P. 93 - 06902 SOPHIA ANTIPOLIS Cedex (France)

EDITEUR
INRIA - Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)

ISSN 0249 - 6399

RR.1936