Similarities for Fuzzy and crisp Probabilistic Expert Systems.

Cristina Coppola, Giangiacomo Gerla, Tiziana Pacelli

Dipartimento di Matematica e Informatica, Università degli Studi di Salerno
via Ponte don Melillo, 84084 Fisciano (SA), Italy
{ccoppola, gerla, tpacelli}@unisa.it

Abstract. In accordance with an idea in [8], in this paper we sketch a method to design expert systems, probabilistic in nature. Indeed, we assume that the probability that an individual satisfies a property is the percentage of similar individuals satisfying such a property. In turn, we call “similar” two individual satisfying the same observable properties. Such an approach is extended to the case of vague properties. We adopt a formalism very close to the one of formal concept analysis.

Keywords: Fuzzy Formal Context, Fuzzy Similarity, State, Probabilistic Expert Systems.

1 Introduction

Imagine that we claim that
(i) the probability of the statement *"a bird is able to fly "* is 0.9,
and compare such a claim with the following one
(ii) the probability of the statement *" Tweety is able to fly "* is 0.9.
Then, as emphasized by F.Bacchus in [1] and J.Y.Halpern in [2] the justifications of these probabilistic assignations looks to be very different. In fact, (i) expresses a statistical information about the proportion of fliers among the set of birds. Such information, related to the class of birds, is statistical in nature. Instead it seem very hard to justify (ii) from a statistical point of view and this since (ii) one refer to a particular bird (Tweety) and not to a class of elements. As a matter of fact either Tweety is able to fly or not and the probabilistic valuation in (ii) is a *degree of belief* depending on the level of my knowledge about the capabilities of Tweety. In [8] it is proposed the idea that in such a case we can refer to the class of birds “similar” to Tweety. More precisely, the belief expressed in (ii) is based on the past experience about the percentage of birds similar to Tweety and able to fly. Obviously, the valuation of the similarity depends on the information on Tweety we have. So, both the probabilistic assignments in (i) and in (ii) are statistical in nature.

On the basis of such an idea, in [3] a method to design probabilistic expert systems was proposed. The crucial notion of similarity is defined in accordance with Leibniz’ principle. Indeed, two individual are called “similar” provided that they share the same *observable* properties.

In this paper we reformulate the approach sketched in [3] and we extend it in order to admit vague properties. In doing this, we adopt a new formalism which is very close to *formal concept analysis* (see [7],[16], [18]) and which is adequate for a suitable extension to the fuzzy framework. This leads also to consider the crucial notion of *state* [11], [21].

2. Probabilistic valuations of the formulas in classical logic

In this section we recall some basic notions of probabilistic logic. In the following we denote by F the set of formulas of a classical zero-order language.

Definition 1. Let $B = (B, \vee, \wedge, -,0,1)$ be a Boolean algebra. A *Boolean valuation* of F (briefly *B-valuation*) is any map v: $F \rightarrow B$ satisfying the following properties, for any α and β:

- $v(\alpha \vee \beta) = v(\alpha) \vee v(\beta)$,
- $v(\alpha \wedge \beta) = v(\alpha) \wedge v(\beta)$,
- $v(\neg\alpha) = 1 - v(\alpha)$.

If B is $(\{0, 1\}, \vee, \wedge, -, 0, 1)$, then the *B-valuation* coincides with the usual truth assignment of the formulas in classical logic. Observe that a *B-valuation* is truth-functional by definition, i.e. the truth value of a compound formula depends on the truth values of its compounds, unambiguously. A formula α is called *tautology* if $v(\alpha) = 1$ for every B-valuation v and *contradiction* if $v(\alpha) = 0$ for every *B-valuation* v. Two formulas α and β are called *logically equivalent* if $v(\alpha) = v(\beta)$ for any v.

Definition 2. A *probability valuation* of F is any map μ: $F \rightarrow [0,1]$ such that:

1. $\mu(\alpha) = 1$ for every tautology α,
2. $\mu(\alpha \vee \beta) = \mu(\alpha) + \mu(\beta)$ if $\alpha \wedge \beta$ is a contradiction,
3. $\mu(\alpha) = \mu(\beta)$ if α is logically equivalent to β .

Let us observe that if μ is a probability valuation, then $\mu(\alpha) = 0$ for every contradiction α. Indeed, in such a case, since α is logically equivalent to $\alpha \vee \alpha$ and $\alpha \wedge \alpha$ is a contradiction, by **2.** and by **3.**, we have that $\mu(\alpha) = \mu(\alpha \vee \alpha) = \mu(\alpha) + \mu(\alpha)$. This entails that $\mu(\alpha) = 0$.

As it is well known, probability valuations are not truth-functional. Nevertheless, the truth-functionality can be obtained by means of the notion of B-valuation.

Definition 3. A *B-probability valuation* of F is a structure (B, v, p) where

- B is a Boolean algebra,
- v: $F \rightarrow B$ is a B-valuation (truth-functional),
- p: $B \rightarrow [0,1]$ is a finitely additive probability on B.

The notion of *B*-probability valuation and that one of probability valuation are strictly related as it is asserted in the following proposition (see [3]) .

Proposition 4. *Let* (*B, v, p*) *be a B-probability valuation and let us define* μ*:F*$\rightarrow$ [0,1] *by setting* $\mu(\alpha)= p(v(\alpha))$ *for every* $\alpha\in F$*. Then* μ *is a probability valuation. Conversely, let* μ*: F*$\rightarrow$ [0,1] *be any probability valuation in F. Then a Boolean algebra B and a B-probability valuation* (*B, v, p*) *exist such that* $\mu(\alpha) = p(v(\alpha))$.

Due to the Representation Theorem of Boolean algebras [2], [12], it is not restrictive to assume that *B* is an algebra of subsets of a set *S*. Moreover, we prefer identifying the subsets of a set with the related characteristic functions. So we refer to Boolean algebras as $\{0,1\}^S$ instead of $P(S)$ as we will see later on.

3. Formal contexts, statistical inferential bases and indiscernibility

The first important step to design a probabilistic expert system is to create a database storing information about past cases we consider related to the actual one (see [3], [8]). The notion of *formal context* seems suitable to represent this kind of collected information. This concept is a basic notion of *formal concept analysis* [7], [18], which is usually used to identify patterns in data and which recognizes similarities between sets of objects based on their attributes.

Definition 5. A *formal context* is a structure (*Ob, AT, tr)* where:
- *Ob* is a finite set whose elements we call *objects*,
- *AT* is a finite set whose elements we call *attributes*,
- *tr*: $Ob \times AT \rightarrow \{0,1\}$ is a binary relation from *Ob* to *AT*.

Given an object *o* and an attribute α, $tr(o,\alpha) = 1$ means that the object *o* possesses the attribute α while $tr(o, \alpha) = 0$ means that *o* doesn't satisfy α. Is easy to represent a formal context by a table, where the objects are the elements of the rows, the attributes are the elements of the columns and in all the cells of the table there are 0 or 1. We consider as set of objects a set of "past cases" and we distinguish two types of attributes: we call *observable* the properties for which it is possible to discover directly whether they are satisfied or not by the examined case. Otherwise, a property is called *non observable*. As an example, an event that will happen in the future is a non observable property. Furthermore the past cases are classified according to observable properties; the "actual case", i.e. the new examined case different from past cases, is considered "*analogous*" to a class of past cases if it satisfies their same observable properties.

Definition 6. A (*complete*) *statistical inferential basis* is a structure *SIB* = (*PC, AT, OBS, an, tr, w*) such that
- (*PC*, *AT*, *tr*) is a formal context,
- *OBS* is a subset of *AT*,
- *an*: $PC\rightarrow\{0,1\}$ is a map from *PC* to {0,1},

- w: $PC \rightarrow \mathbb{N}$ is a function called *weight function*.

We call the elements of *PC past cases* and the map tr: $PC \times AT \rightarrow \{0,1\}$ *information function*. The set *OBS* is the subset of the *observable attributes* and the map *an* is regarded as the (characteristic function of the) set of past cases analogous to the actual one. The meaning of the number $w(c) = n$ is that the past case c is the representative of n analogous cases. Then, we set the *total weight* of a statistical inferential basis *SIB* as

$$w(SIB)= \sum\{w(c)an(c) \,/\, c\in PC\}.$$

It corresponds to the number of the past cases analogous to the actual case represented globally by *SIB*. If $w(SIB)\neq 0$ then we say that the statistical inferential basis is *consistent*.

We denote by F (by F_{obs}) the set of formulas of the propositional calculus whose set of propositional variables is *AT* (is *OBS*, respectively). As usual, the function *tr* can be extended to the whole set F of formulas by setting, for every formula α and β,

- $tr(c, \alpha\wedge\beta) = min\{tr(c, \alpha),\ tr(c, \beta)\}$,
- $tr(c, \alpha\vee\beta) = max\{tr(c, \alpha),\ tr(c, \beta)\}$,
- $tr(c, \neg\alpha) = 1-tr(c, \alpha)$.

In this way, any past case is associated by *tr* with a classical valuation of the formulas in F.

In accordance with the basic notions of probabilistic logic, exposed in the previous section, now we provide some definitions of valuations associated to a statistical inferential basis *SIB*.

Proposition 7. *Every consistent statistical inferential basis SIB* = (*PC, AT, OBS, an, tr, w*) *defines a B-probability valuation* (*B, v, p*) *in F such that:*

- *B is the Boolean algebra* $(\{0,1\}^{PC}, \cup, \cap, \sim, c_{\emptyset}, c_{PC})$,
- $v(\alpha)$: $PC\rightarrow\{0,1\}$ *is* (*the characteristic function of*) *the set of past cases satisfying* α, *i.e.*

$$v(\alpha)\,(c)= tr(c, \alpha)\,,$$

- p: $B\rightarrow[0,1]$ *is the probability in B defined by setting, for any* $s\in\{0,1\}^{PC}$,

$$p(s) = \frac{\sum\{w(c)an(c)s(c) \,/\, c\in PC\}}{w(SIB)}\,.$$

In particular, we have that $p(c) = p(\{c\}) = \dfrac{w(c)an(c)}{w(SIB)}$.

As a consequence of Proposition 9 and Proposition 11 any statistical inferential basis *SIB* can be associated with a probability valuation μ of the formulas. So we have, for every formula α,

$$M(\alpha) = p(v(\alpha)) = \frac{\sum \{w(c)an(c)tr(c,\alpha) / c \in PC\}}{w(SIB)} .$$

In other words, $\mu(\alpha)$ represents the percentage of past cases (analogous to the actual case) in which α is true according to the stored dates.

According to the main idea we refer, it is important to specify which relation we take into account in order to consider "analogous" two cases. In the following, we will introduce a formalism very close to Pawlak' s one [14] and remembering Leibniz's indiscernibility principle for which two individuals are indiscernible if they share the same properties.

Definition 8. Let A be a subset of AT. Let $\leftrightarrow$ be the operation corresponding to the equivalence in the classical zero-order language and e: $PC \times PC \to \{0,1\}$ be a relation on PC defined by setting

$$e(c_1, c_2) = Inf_{\alpha \in A}\ tr(c_1, \alpha) \leftrightarrow tr(c_2, \alpha). \quad (1)$$

If $e(c_1, c_2) = 1$ we call the two cases c_1 and c_2 *A-indiscernible* and we write $c_1 \cong_A c_2$.

Let us observe that two cases are A-indiscernible if $tr(c_1, \alpha) = tr(c_2, \alpha)$ for every $\alpha \in A$, i.e. if they satisfy the same properties in A. It is immediate that $\cong_A$ is an equivalence relation in PC. Then, for every case c, we can consider the corresponding equivalence class $[c]_A$ and, obviously, the quotient of PC modulo $\cong_A$.

In particular, we are interested to identify the past cases satisfying the same observable properties of the actual case. Let us recall that by "actual case" we intend a case different from past cases in which the only available information is that expressed by the set F_{obs} in the language of "observable" properties. To our aim it is important to give an adequate definition of actual case.

Definition 9. We call *actual case* any map a_c: $OBS \to \{0,1\}$ from the set of the observable OBS to $\{0,1\}$. We call *piece of information about* a_c any subset of a_c, i.e. any partial map T: $OBS \to \{0,1\}$ such that a_c is an extension of T. We say that T is *complete* if $T = a_c$.

So, we identify the actual case with the "complete information" about its observable properties. As we will see in the next sections, we can collect pieces of information about a_c by a query process. In the following we denote the actual case by a_c or by the family $\{(\alpha, a_c(\alpha)\}_{\alpha \in OBS}$, indifferently. We extend the information function tr to the actual case by setting $tr(a_c, \alpha) = a_c(\alpha)$ for every $\alpha \in OBS$ and then to the whole set F_{obs} of observable formulas in the usual way. We also extend the relation e by considering pieces of information on the actual case a_c. Indeed, given a piece of information T, we set

$$e_T(c, , a_c) = Inf_{\alpha \in Dom(T)}\ tr(c, \alpha) \leftrightarrow tr(a_c, \alpha)\ .$$

If $e_T(c, a_c) = 1$, then c is a past case which is *OBS–indiscernible from the actual case a_c given the information T*. If T is complete then we write $e(c, a_c)$ instead of $e_T(c, a_c)$.

Definition 10. Let *SIB* be a statistical inferential basis. We say that a piece of information T is *consistent with SIB* if there exists a past case $c\in PC$ such that $an(c)\neq 0$ and $e_T(c, a_c)\neq 0$.

Let us observe that if T is consistent with *SIB* there is a past case c analogous to the actual case according to the available information T, i.e. a past case c exists such that it satisfies the same observable property of a_c with respect to T.

Given a statistical inferential basis $SIB = (PC, AT, OBS, an, tr, w)$, representing the basic information, and a piece of information on the actual case $T = \{(\alpha_1, T(\alpha_1)),..., (\alpha_n, T(\alpha_n))\}$, we are able to obtain a new statistical inferential basis $SIB(T)$ from *SIB*.

Definition 11. Let *SIB* be a statistical inferential basis and T a piece of information on a_c consistent with *SIB*. We call *statistical inferential basis induced by T in SIB* the structure:

$$SIB(T) = (PC, AT, OBS, an_T, tr, w),$$

where an_T is defined by setting $an_T(c) = an(c)\, e_T(c, a_c)$.

In accordance with Proposition 11 and also considering the B-probability valuation (B, v, p) associated to *SIB*, the statistical inferential basis $S(T)$ defines a B-probability valuation (B, v_T, p_T) where:

- B is the Boolean algebra $(\{0,1\}^{PC}, \cup, \cap, \sim, c_\emptyset, c_{PC})$,
- v_T: $F\rightarrow B$ is a B-valuation of the formulas in F defined by

$$v_T(\alpha)(c) = an_T(c)v(\alpha)(c) = an(c)e_T(c,a_c)tr(c,\alpha)\ ,$$

i.e. $v_T(\alpha)$ is (the characteristic function of) the set of past cases which are indiscernible from a_c (given the available information T) and verifying α,

- p_T: $B\rightarrow[0,1]$ is the probability on B defined by setting, for any $s\in \{0,1\}^{PC}$,

$$p_T(s) = \frac{\sum \{w(c)an_T(c)s(c) / c \in PC\}}{w(SIB(T))}\ . \qquad \textbf{(1)}$$

As usual, we have a probability valuation μ_T of the formulas defined, for every formula α, as $\mu_T(\alpha) = p_T(v_T(\alpha))$, i.e.

$$\mu_T(\alpha) = \frac{\sum \{w(c)an_T(c)tr(c,\alpha) / c \in PC\}}{w(SIB(T))} \qquad \textbf{(2)}$$

The number $\mu_T(\alpha)$ is the percentage of the past cases verifying α among the cases in S considered analogous to a_c taking into account the available information T.

Obviously, the probability p_T, defined in (1), can be regarded as the *conditioned probability* $p(\_/m_T)$, where m_T denotes the set of past cases indiscernible from a_c given T. Indeed, for any $s\in \{0,1\}^{PC}$ and by 0, we have

$$p_T(s)= \frac{\sum\{w(c)s(c)an_T(c)/\, c\in PC\}}{\sum\{w(c)an_T(c)/\, c\in PC\}}=$$

$$= \frac{\sum\{w(c)s(c)an_T(c)/\, c\in PC\}}{w(S)}\frac{w(S)}{\sum\{w(c)an_T(c)/\, c\in PC\}}=$$

$$= \frac{\sum\{w(c)s(c)an(c)e_T(c,a_c)/\, c\in PC\}}{w(S)}\frac{w(S)}{\sum\{w(c)an(c)e_T(c,a_c)/\, c\in PC\}}=$$

$$= \frac{\sum\{w(c)s(c)an(c)m_T(c)/\, c\in PC\}}{w(S)}\frac{w(S)}{\sum\{w(c)an(c)m_T(c)/\, c\in PC\}}=$$

$$= \frac{p(s\cap m_T)}{p(m_T)} = p(s\,/m_T).$$

Consequently, for every formula α, also the probability valuation μ_T can be regarded as the *conditioned probability* $\mu_T(\alpha) = \mu(\alpha\,/\,m_T)$.

4. A step-by-step Inferential Process

In this section we describe how the step-by-step inferential process works. We imagine an expert system whose inferential engine contains an *initial* statistical inferential basis *SIB*, i.e. a statistical inferential basis such that *an* is constantly equal to 1. This means that initially and in absence of information on a_c we assume that all the past cases are analogous to the actual case. Successively, we can obtain information on a_c by a sequence α_1 , ... α_n of queries about observable properties. So, we set $T_0 = \varnothing$ and, given a new query α_i, we set $T_i = T_{i-1}\cup\{(a_i,\lambda_i)\}$ where $\lambda_i = 1$ if the answer is positive (the actual case verifies α_i) and $\lambda_i = 0$ otherwise. As a consequence, we obtain a sequence of corresponding inferential statistical basis $\{SIB(T_i)\}_{i=1,\dots,n}$. At every step we can evaluate the probability that a_c *satisfies* β *given the available information*. Obviously, we are interested to a non observable property β.

Definition 12. Let *SIB* be an initial statistical inferential basis and β be a formula in F. Let T_n be the available information on a_c obtained by a sequence of n queries. Then we call *probability that* a_c *satisfies* β *given the information* T_n, the probability of β in the statistical inferential basis $SIB(T_n)$ induced by T_n in *SIB*.

More precisely, we have the following step-by-step process:

1. Set $T_0 = \emptyset$ and $SIB_0 = SIB(\emptyset) = SIB$.
2. Given T_k and SIB_k=$SIB(T_k)$, after the query α_{k+1} and the answer λ_{k+1},
put $T_{k+1}= T_k \cup\{(\alpha_{k+1}, \lambda_{k+1}))\}$ and $SIB_{k+1}= SIB(T_{k+1})$.
3. If the information is "*sufficient*" or *complete* goto 4, otherwise goto 2.
4. Set $\mu(\beta) = \mu_{Tk+1}(\beta)$ as defined in (2).
5. If T_{k+1} is inconsistent with SIB_{k+1} then the process is "*failed*".

Let us observe that we have different processes depending on the choice of the queries and on the "*stop*"-criterion expressed by the term "*sufficient*". As an example, the query α_i can be selected in order to minimize the expected value of the entropy. This is achieved by minimizing the value $|\mu(\alpha_i)–\mu(\neg\alpha_i)|$ where μ is the valuation related to SIB_i. Also, let us notice that once a complete information on a_c is obtained (in the language of the observable properties), $T_n = a_c$ and the inferential process necessarily terminates by giving a probabilistic valuation of the formulas by

$$\mu(\beta) = \frac{\sum\{w(c)e(c,a_c)tr(c,\alpha)/\ c \in PC\}}{\sum\{w(c)e(c,a_c)/\ c \in PC\}}$$

In othe words
"The probability that the actual case a_c satisfies the property β is given by the percentage of the cases OBS- indiscernible from a_c that in the past verified β".
Such a point of view gives an answer to the question about the probabilities related to singular cases [8].

5. Vague properties and similarities

In the previous sections we have considered only the presence of crisp attributes. An object satisfies or doesn't satisfy a property. But the real world has a fuzzy nature. In the most real situations an object verifies a property with a "degree". So, if we consider the presence of eventually "vague" properties, it is necessary to extend the notions we have considered so far. Firstly, we give some basic notions in multi-valued logic. In many-valued logics [4], [5], [9], [12] truth degrees are not two yet, but three or more and many different algebraic structures are used for the evaluation of formulas.

In this section we present a class of these structures, the class of *MV-algebras*, devised by C.C.Chang [4], and then we introduce some other notion concerning multi-valued logic, such as *fuzzy set* and *fuzzy-similarity*.

Definition 13. An *MV-algebra* [5] is a structure $A = (A, \oplus, \neg, 0)$ such that $(A, \oplus, 0)$ is a commutative monoid satisfying the following additional properties:
1. $\neg\neg a = a$;
2. $a\oplus\neg 0=\neg 0$;
3. $\neg(\neg a \oplus b) \oplus b = \neg (\neg b \oplus a) \oplus a$.

On each *MV*-algebra *A* we define the element 1 and the operation $\otimes$ as follows:

$$1=\neg 0 \quad \text{and} \quad a\otimes b = \neg(\neg a \oplus \neg b) .$$

A well known example of *MV*-algebra is given by the *Lukasiewicz algebra* ([0,1], $\oplus$, $\neg$, 0), where $\oplus$ is the *Lukasiewicz disjunction* defined by

$$a\oplus b = min(1, a+b)$$

and $\neg a=1-a$. As a consequence the operation $\otimes$ is the *Lukasiewicz conjunction* defined by

$$a\otimes b= max(0, a+b-1) .$$

Lukasiewicz conjuction and disjunction are, respectively, examples of *t-norm* and *t-conorm* [9], [12].

Definition 14. A *triangular norm* (briefly *t-norm*) is a binary operation $\otimes$ on [0,1] such that, $\otimes$ is commutative, associative, isotone in both arguments, i.e., $x_1 \leq x_2 \Rightarrow x_1 \otimes y \leq x_2 \otimes y$ and $y_1 \leq y_2 \Rightarrow x \otimes y_1 \leq x \otimes y_2$, and $\otimes$ verifies the boundary conditions, i.e. $1\otimes x = x = x \otimes 1$ and $0 \otimes x = 0 = x \otimes 0$, for all $x, y, z, x_1, x_2, y_1, y_2 \in [0,1]$.

Definition 15. A *t-conorm* is a binary operation $\oplus : [0,1]^2 \rightarrow [0,1]$ such that $\oplus$ is commutative, associative, isotone in both arguments, and such that $0 \oplus x = 0 = x \oplus 0$ and $1 \oplus x = x = x \oplus 1$.

Moreover, the t-conorm $\oplus$ is *dual* to a given t-norm $\otimes$ if, for every $x, y \in [0,1]$,

$$x \oplus y =1-((1-x) \otimes (1-y)) .$$

For each t-norm, we can consider the *associated biresiduation*, suitable to represent the truth function of equivalence. In the case of Lukasiewicz conjunction, it is defined by

$$a \leftrightarrow_{\otimes} b = 1-|a-b| .$$

and some its properties are listed in the following:

- $x \leftrightarrow_{\otimes} x = 1$,
- $x \leftrightarrow_{\otimes} y = 1 \Leftrightarrow x = y$,
- $(x \leftrightarrow_{\otimes} y) \otimes (y \leftrightarrow_{\otimes} z) \leq x \leftrightarrow_{\otimes} z$,
- $x \leftrightarrow_{\otimes} y = y \leftrightarrow_{\otimes} x$.

Fuzzy set theory [19] can be regarded as an extension of the classical one, where an element either belongs or does not belong to a set. Fuzzy set theory permits the gradual assessment of the membership of elements to a set, by a generalized characteristic function.

Definition 16. Let S be a set and let us consider the complete lattice [0,1]. We call *fuzzy-subset* of S any map $s{:}S \to [0,1]$ and we denote by $[0,1]^S$ or by $\mathfrak{F}(S)$ the class of all the fuzzy-subsets of S.

Given any x in S, the value $s(x)$ is the "degree of membership" of x to s. In particular, $s(x)= 0$ means that x is not included in s, whereas 1 is assigned to the elements fully belonging to s. Any fuzzy subset s such that $s(x)\in\{0,1\}$, for any $x\in S$, is called *crisp set*. Given $\lambda\in[0,1]$, we denote by s^λ the fuzzy set constantly equal to λ.

Definition 17. Let $\otimes$ be the Lukasiewicz conjunction and $\oplus$ be the Lukasiewicz disjunction. We define the *union*, the *intersection* and the *complement* by setting, respectively, for any $s, s'\in \mathfrak{F}(S)$ and for every $x\in S$,

- $(s \cup_\oplus s')(x) = s(x) \oplus s'(x)$
- $(s \cap_\otimes s')\,(x) = s(x) \otimes s'(x)$
- $(\sim s)(x) = -\, s(x)$.

Proposition 18. *The structure* $(\mathfrak{F}(S),\cup_\oplus, \cap_\otimes, \sim, s^0, s^1)$ *is an MV-algebra extending the Boolean algebra* $(P(S)\,, \cup, \cap, \sim, \emptyset, S)$ *of the subsets of* S .

In the following we denote this *MV*-algebra also by $(\mathfrak{I}(S), \oplus, \sim, s^0)$.

A special class of fuzzy sets is given by the concept of *similarity* [18], which is essentially a generalization of an equivalence relation.

Definition 19. Let $\otimes$ be the Lukasiewicz conjunction. A $\otimes$-*fuzzy-similarity* on a set S is a fuzzy-relation on S, i.e. a fuzzy subset of $S\times S$, E: $S\times S\to [0,1]$, satisfying the following properties

1. $E(x,x)=1$ (*reflexivity*)
2. $E(x,y) = E(y,x)$ (*symmetry*)
3. $E(x,y) \otimes E(y,z) \leq E(x,z)$. ($\otimes$- *transitivity*)

The logical meaning of the $\otimes$- transitivity is that "if x is similar to y with a degree $E(x,y)$ and y is similar to z with a degree $E(y,z)$ then x is similar to z with a degree $E(x,z)$ greater or equal to $E(x,y) \otimes E(y,z)$".
Let us recall that for any t-norm we can have a corresponding notion of fuzzy similarity but we give the definition directly by the Lukasiewicz conjunction because we will use it in the proposed inferential process.
In the following we refer to the following basic theorem enabling to extend Proposition 8 to vague properties (see Valverde [17]) and, in a sense, related with Leibniz's indiscernibility principle.

Proposition 20. Consider a finite family $(s_i)_{i\in I}$ of fuzzy subsets of a set S and define the fuzzy relation

$$e(x,y) = \otimes_{i\in I}\, s_i(x)\leftrightarrow_\otimes s_i(y).$$

Then e is a $\otimes$-similarity in S.

6. Probabilistic logic in fuzzy framework.

In this section we extend the basic notions of probabilistic logic, exposed in Section 3 and, since we will admit the presence of eventually “vague” properties in the inferential process, we have to consider probabilistic valuation of fuzzy subsets. In particular, we refer to the concept of *state* [11], which is a generalization on *MV*-algebras of the classical notion of (finitely additive) probability measure on Boolean algebras. In the following, we denote by F the set of formulas in the language of a many-valued logic. More precisely, we refer to a logic whose propositional calculus assumes truth values in an *MV*-algebras.

Definition 20 Let $(A, \oplus, \neg, 0)$ be an *MV*-algebra. An *MV-valuation* is any map $v_f: F \rightarrow A$ satisfying the following properties:

- $v_f(\alpha \vee \beta) = v_f(\alpha) \oplus v_f(\beta)$,
- $v_f(\alpha \wedge \beta) = v_f(\alpha) \otimes v_f(\beta)$,
- $v_f(\neg\alpha) = \neg v_f(\alpha)$.

Trivially, v_f is a truth-functional map by definition. Moreover, a formula α is called *tautology* if $v_f(\alpha) = 1$ and it is called *contradiction* if $v_f(\alpha) = 0$, for any *MV*-valuation v_f. Two formulas α and β are *logically equivalent* if $v_f(\alpha) = v_f(\beta)$ for any valuation v_f.

Definition 21. A *state* of an *MV*-algebra A is a map p_f: $A \rightarrow [0,1]$ satisfying the following conditions:

1. $p_f(0) = 0$,
2. $p_f(1) = 1$,
3. $p_f(a \oplus b) = p_f(a) + p_f(b)$ for every $a, b \in A$ such that $a \otimes b = 0$.

A natural example of state in the *MV*-algebra $(\Im(X), \oplus, \sim, s^0)$, where we have Lukasiewicz disjunction, is given by [21]:

Proposition 22. *Let X be a finite set and $p:\{0,1\}^X \rightarrow [0,1]$ an arbitrary probability measure on $\{0,1\}^X$. Let the map p_f: $\Im(X) \rightarrow [0,1]$ be defined, for every $s \in \Im(X)$, by*

$$p_f(s) = \Sigma\{s(x)p(x) \ / \ x \in X\}.$$

Then p_f is a state of the MV-algebra $(\Im(X), \oplus, \sim, s^0)$.

We introduce the notions of *MV-probability valuation* of formulas and, then, of *A-probability valuation* which enables us to obtain the truth-functionality of the first one.

Definition 23. An *MV- probability valuation* of F is any map μ: $F \rightarrow [0,1]$ such that:

- $\mu(\alpha) = 1$ for every tautology α,
- $\mu(\alpha \vee \beta) = \mu(\alpha) + \mu(\beta)$ if $\alpha \wedge \beta$ is a contradiction,
- $\mu(\alpha) = \mu(\beta)$ if α is logically equivalent to β.

Let us observe that the only difference with Definition 7 is that the notions of "tautology", "contradiction" and "logically equivalent" are intended in the sense of Definition 17.

Definition 24. An *A-probability valuation* is a structure (A, v_f, p_f) where

- A is an *MV*- algebra,
- v_f: $F \rightarrow A$ is a truth-functional *MV*-valuation of formulas,
- p_f: $A \rightarrow [0,1]$ is a state on A.

The notion of *A*-probability valuation is connected to that one of *MV*-probability valuation [13].

Proposition 25. *Let* (A, v_f, p_f) *be an A-probability valuation and let us define* μ:$F \rightarrow [0,1]$ *by setting* $\mu(\alpha) = p_f(v_f(\alpha))$ *for every* $\alpha \in F$. *Then* μ *is an MV-probability valuation. Conversely, let* μ:$F \rightarrow [0,1]$ *be any MV-probability valuation in F. Then an MV-algebra A and an A-probability valuation* (A, v_f, p_f) *exist such that* $\mu(\alpha)=p_f(v_f(\alpha))$.

7. Fuzzy Statistical Inferential Bases

In order to create a database of past cases verifying eventually vague properties in this section we extend the definitions presented in Section 4. We refer to a generalization of the basic notion of *formal concept analysis* [7], [16].

Definition 26. A *fuzzy formal context* [14] is a structure (Ob, AT, tr_f) where:

- Ob is a finite set whose elements we call *objects*,
- AT is a finite set whose elements we call *attributes*,
- $tr_f : Ob \times AT \rightarrow [0,1]$ is a fuzzy binary relation from Ob to AT.

The fuzzy relation tr_f connects any object with any attribute, i.e. the value $tr_f(o, \alpha)$ is the truth degree of the claim "the object o satisfies the property α". As in Section 4, we consider as set of objects the set of "past cases" and we distinguish two types of attributes: we call *observable* the attributes for which it is possible to discover directly whether they are satisfied by the examined case, *non observable* the others. The "actual case", different from past cases, is considered "similar" to a class of past cases if it satisfies their same observable properties. We want to evaluate the probability with which the actual case verifies a non observable property.

Definition 27. A (*complete*) *fuzzy statistical inferential basis* is a structure $SIB_f = (PC, AT, OBS, sim, tr_f, w)$ such that

- (PC, AT, tr_f) is a fuzzy formal context,
- OBS is a subset of AT,
- sim: $PC \rightarrow [0,1]$ is a fuzzy subset of PC,
- $w : PC \rightarrow \mathbb{N}$ is a function called *weight function*,

The set PC is that one of *past cases* and the map tr_f is called *fuzzy information function.* It provides the degree with which a past case satisfies an attribute. The set OBS is the (classical) subset of the observable attributes and the map sim is interpreted as the fuzzy set of past cases "similar" to the actual one. The value $w(c)$ gives the number of past cases whose representative is c. Then, we set the *total weight* of a fuzzy statistical inferential basis SIB_f as

$$w(S_f)= \sum\{w(c)sim(c) \ / \ c\in PC \} .$$

If $w(S_f)\neq 0$ then we say that S_f is *consistent.*
Differently from the total weight for a SIB, this formula doesn't yields the actual number of past cases similar to the actual one, but it furnishes a "fuzzy" weight to the whole inferential basis by giving a higher value if there are a lot of "similar" past cases.

As in the previous section, we denote by F (by F_{obs}) the set of formulas of a multivalued propositional calculus whose set of propositional variables is AT (or OBS, respectively). So we extend tr_f to the whole set F of formulas by setting

- $tr_f(c, \alpha\wedge\beta) = tr_f(c, \alpha) \otimes tr_f(c, \beta)$,
- $tr_f(c, \alpha\vee\beta) = tr_f(c, \alpha) \oplus tr_f(c, \beta)$,
- $tr_f(c, \neg\alpha) = 1- tr_f(c, \alpha)$.

By referring to the notions introduced in Section 8, we provide definitions of valuations associated to a fuzzy statistical inferential basis.

Proposition 28. *Every consistent fuzzy statistical inferential basis SIB_f defines an A-probability valuation (A, v_f , p_f) in F such that:*

- *A is the MV-algebra $(\mathfrak{F}(PC), \oplus, \sim, s^0)$,*
- *$v_f(\alpha)$ is the fuzzy subset of the past cases satisfying the formula α, i.e.*

$$v_f(\alpha)(c) = tr_f(c, \alpha)$$

- *p_f: $A\to[0,1]$ is the state on A defined, for any $s\in \mathfrak{F}(PC)$, as in Proposition 19, i.e.*

$$p_f(s) = \Sigma\{s(c)p(c) \ / \ c\in PC\} ,$$

where p is the probability on $\{0,1\}^{PC}$, defined by

$$p(c) = \frac{w(c)sim(c)}{w(SIB_f)} .$$

By Proposition 22, any fuzzy statistical inferential basis SIB_f can be associated with an MV-probability valuation μ of the formulas, defined, for every α, by

$$\mu(\alpha)= p_f(v_f(\alpha))= \Sigma\{tr_f(c,\alpha)\, p(c) \ / \ c\in PC\} .$$

In other words we have $\mu(\alpha)= \dfrac{\sum\{w(c)sim(c)tr_f(c,\alpha) \ / \ c\in PC\}}{w(SIB_f)}$, and this value represents the percentage of past cases similar to the actual case in which α is verified.

8. The Actual Case and its Similar Past Cases

The indiscernibility relation in Definition 12, used for “crisp” properties, isn’t sufficient anymore for “vague” properties. Indeed, in a classification process, given a set of (eventually vague) properties A, and a property $\beta \in A$, if $tr(c_1, \alpha) = tr(c_2, \alpha)$ for every $\alpha \in A \setminus \{\beta\}$ and $tr(c_1, \beta) = 0{,}8$ and $tr(c_2, \beta) = 0{,}9$ it is not reasonable to consider the two case c_1 and c_2 not “analogous”. Therefore, it is necessary to take into account an extension of the relation such that it results appropriate to a classification handling “vague” properties and in order to consider “similar” two cases whit respect to these properties. As an immediate consequence of Proposition ... we obtain the following one where $\oplus$ and $\otimes$ denotes the Lukasiewicz conjunction and disjunction, respectively.

Proposition 30. *Let SIB_f be a fuzzy statistical inferential basis and let (A, v_f, p_f) be the A-probability valuation associated to it. Then, for any subset B of AT, the fuzzy relation $E: PC \times PC \to [0,1]$, defined by setting*

$$E(c_1,c_2) = \otimes_{\alpha \in B}(v_f(\alpha)(c_1) \leftrightarrow_{\otimes} v_f(\alpha)(c_2)) , \qquad \textbf{(3)}$$

is a $\otimes$- fuzzy similarity.

Since the fuzzy set $v_f(\alpha): PC \to [0,1]$ of past cases satisfying the property α is defined by $v_f(\alpha)(c) = tr_f(c, \alpha)$, we can rewrite (4) as

$$E(c_1,c_2) = \otimes_{\alpha \in B}(tr_f(c_1, \alpha) \leftrightarrow_{\otimes} tr_f(c_2, \alpha)) ,$$

The value $E(c_1,c_2)$ furnishes the “degree of similarity” between the two past cases c_1 and c_2 in SIB_f. From the logical point of view, it is the valuation of the claim “every property satisfied by c_1 is satisfied by c_2 and vice-versa ”.

As usual, a similarity can be interpreted in terms of *fuzzy similarity classes*, one for each element of the universe. In our situation, for every case c_j, we can consider a fuzzy subset $sim_{cj}: PC \to [0,1]$ as the fuzzy class of the past cases “similar” to c_j, by setting $sim_{cj}(c) = E(c ,c_j)$. In particular, we have to identify the past cases similar to the actual one. Let us recall that by “actual case” we intend a case different from past cases in which the only available information is that one expressed by the set F_{obs} in the language of “observable” properties. The definition of “actual case” in the fuzzy situation is a generalization of that one in the crisp case.

Definition 30. We call (*fuzzy*) *actual case* any map $a_c:OBS \to [0,1]$ from the set of observable properties to the interval [0,1]. We call *piece of information about a_c* any subset of a_c, i.e. any partial map T: $OBS \to [0,1]$ such that a_c is an extension of T. We say that T is *complete* if $T = a_c$.

We denote the actual case by a_c or by the family $\{a_c, a_c(\alpha)\}_{\alpha \in OBS}$, indifferently. The last notation is more useful in describing the inferential process, where we identify the actual case with the “information” about its observable properties, collected by a query process. We extend the fuzzy information function tr_f and the similarity E, given in (3) to the actual case by setting $tr_f(a_c, \alpha)= a_c(\alpha)$ for every $\alpha\in OBS$ and, given a piece of information T about a_c, we set

$$E_T(c, a_c)= \otimes_{\alpha\in Dom(T)} \ (tr_f(c, \alpha) \leftrightarrow_{\otimes} tr_f(a_c, \alpha)) \ .$$

So, it results that the past case c is *similar to the actual case* a_c, *given the information* T, *with degree* $E_T(c, a_c)$. If T is complete then we write $E(c, a_c)$ instead of $E_T(c, a_c)$.

Definition 31. Let SIB_f be a fuzzy statistical inferential basis. We say that a piece of information T is *consistent with* SIB_f if there exists $c\in PC$ such that $sim(c)\neq 0$ and $E_T(c, a_c)\neq 0$.

If T is consistent with SIB_f, then in our database there is at least a past case c similar to a_c according to the available information T.

9. Fuzzy Statistical Inferential Bases Induced by a Piece of Information and the Step-by-Step Inferential Process

Given a fuzzy statistical inferential basis SIB_f and a piece of information T on the actual case a_c, we obtain a new fuzzy statistical inferential basis $SIB_f(T)$ from SIB_f.

Definition 32. Let SIB_f be a consistent fuzzy statistical inferential basis and T be a piece of information on a_c. We call *fuzzy statistical inferential basis induced by T in* SIB_f the structure $SIB_f(T)=(PC, AT, OBS, sim_T, tr_f, w)$, where sim_T is defined by setting $sim_T(c) = sim(c) \otimes E_T(c, a_c)$.

Let us observe that sim_T can be regarded as the fuzzy class of the past cases "similar" to a_c given the information T. Then, in accordance with Proposition 25 and given the A-probability valuation (A, v_f, p_f) associated to SIB_f, the induced fuzzy statistical inferential basis $SIB_f(T)$ defines an A-probability valuation (A, v_f^T, p_f^T) where:

- A is the MV-algebra $(\Im(PC), \oplus, \sim, s^0)$,
- $v_f^T : F\rightarrow A$ is an MV-valuation of the formulas defined by

$$v_f^T(\alpha)(c) = sim_T(c) \otimes v_f(\alpha)(c) = sim(c) \otimes E_T(c, a_c) \otimes tr_f(c, \alpha) \ ,$$

i.e. $v_f^T(\alpha)$ is the fuzzy set of the past cases similar to a_c (given the information T) and verifying the formula α ,

- $p_f^T: A\rightarrow[0,1]$ is the state on A defined by setting, for any $s\in \Im(PC)$,

$$p_f^T(s)= \Sigma\{s(c)\, p_T(c) \,/\, c\in PC\} \ ,$$

where p_T is the probability on $\{0,1\}^{PC}$ given by

$$p_T(c) = \frac{w(c)sim_T(c)}{w(SIB_f(T))} = \frac{w(c)sim_T(c)}{\sum \{w(x)sim_T(x) / x\in PC\}} \ .$$

So, give a fuzzy statistical inferential basis SIB_f and a piece of information T, we obtain an *MV*-probability valuation μ_T of the formulas, defined, for every formula α, by

$$\mu_T(\alpha)= p_f^{\,T}(v_f^{\,T}(\alpha)) = \Sigma\{[sim_T(c)\otimes v_f(\alpha)(c)]\, p_T(c) \,/\, c\in PC\}\ . \qquad \textbf{(4)}$$

Let us observe that we obtain

$$\mu_T(\alpha)=\frac{\sum \{w(c)sim_T(c)[sim_T(c)\otimes tr_f(c,\alpha)]/\ c\in PC\}}{\sum \{w(c)sim_T(c)/\ c\in PC\}}\ , \qquad \textbf{(5)}$$

and it represents the percentage of the past cases verifying α among the cases in SIB_f considered similar to a_c according to the available information T.

Now, let us imagine the expert system has to evaluate the probability that an actual case a_c verifies a non observable formula β. Let us suppose that in the *initial* fuzzy statistical inferential basis SIB_f, the map *sim* is constantly equal to 1, i.e. we are considering all the past cases "similar" to the actual one. The information on a_c can be obtained by a query-strategy. Let us denote by $\alpha_1,\ldots,\alpha_n$ a sequence of appropriate queries about observable properties of a_c. Then, we set $T_0=\emptyset$ and, given a new query α_i, we set $T_i=T_{i-1}\cup\{(\alpha_i, \lambda_i)\}$, where $\lambda_i\in[0,1]$ is the degree with which the actual case verifies the property α_i. Consequently, we obtain a sequence of corresponding fuzzy inferential statistical basis $\{SIB_f(T_i)\}_{i=1,\ldots,n}$ .At every step we have the probability that a_c *satisfies* β *given the available information*.

Definition 33 Let SIB_f be an initial fuzzy statistical inferential basis and β be a formula in F. Let T_n be the available information on a_c obtained by a sequence of n queries. Then we call *probability that* a_c *satisfies* β *given the information* T_n, the probability of β in the fuzzy statistical inferential basis $SIB_f(T_n)$ induced by T_n in SIB_f.

More precisely, the step-by-step inferential process is analogous to that in Section 7:

1. Set $T_0=\emptyset$ and $SIB_0 = SIB_f(\emptyset) = SIB_f$.
2. Given T_k and $S_k=S_f(T_k)$, after the query α_{k+1} and the answer λ_{k+1},
put $T_{k+1}=T_k\cup\{(\alpha_{k+1}, \lambda_{k+1}))\}$ and $SIB_{k+1}= SIB_f(T_{k+1})$.
3. If the information is "*sufficient*" or *complete* goto 4, otherwise goto 2.
4. Set $\mu(\beta) = \mu_{Tk+1}(\beta)$ as defined in (4).
5. If T_{k+1} is inconsistent with SIB_{k+1} then the process is "*failed*".

Let us notice that if the information on a_c is complete, then $T_n=a_c$ and the inferential process terminates. In such a case in (5), $sim_T(c)=E(c,a_c)$ and (5) represents the percentage of the past cases similar to a_c satisfying β in the past.

Let us observe that the proposed process is also interpretable in the framework of *case-based reasoning* (see for example [15]). Indeed, if we interpret the set of non observable attributes *AT-OBS* as the set of probable "solutions" and the value $tr_f(c, \alpha)$, with $\alpha\in$ *AT-OBS*, as the "validity degree" of the solution α for the case c collected in the database, the final value $\mu_T(\alpha)$ represents a "validity degree" of the solution α for the actual case. In such a case $\mu_T(\alpha)$ is the percentage of the past cases similar to a_c for which α was a "good solution". Our approach is close to *case-based reasoning*, systems since we make a prediction on a new case by observing precedent cases. On the other hand, the prediction, probabilistic in nature, is obviously different from that one used in other approaches, generally possibilistic in nature [6].

10. Future works and open questions

The researches related with the sketched expert systems are at an initial state and several questions are open. The main one is to test such an idea in some concrete cases. To this regard, observe that there is no difficulty in realizing a suitable program in any relational database management system (see [3]). ???

A second question is related with the difficulties of interpreting the probabilistic valuation of a formula in $SIB_f(T)$ as a conditioned probability in SIB_f (as we have made for the crisp case in Section 6). In fact, we can define the *conditioned state*, as in the classical probability theory, by setting $p_f(s/t) = p_f(s\otimes t)/\ p_f(t)$ and, due to the associativity of $\otimes$, p_f satisfies the *iteration rule* of the classical conditioning for a probability, i.e. $p_f(s/t\cap_\otimes v)= p_f(s\cap_\otimes t/v)\ /\ p_f(t/v)$. This is a basic property that it is useful in the inferential process and for a possible implementation of the expert system. Unfortunately, p_f doesn't result a state, since $\otimes$ is not distributive with respect to $\oplus$. So, we might look for an adequate definition of state, such that the corresponding conditioned state verifies the *iteration rule*.

Another question is related to the kind of the available information on the actual case. Indeed, it should be natural to admit that this information can be expressed by intervals, i.e. $T = \{a_c, I(\alpha)\}$ where $I(\alpha)$ is a closed interval in [0,1]. In fact, it is natural to admit that the truth value of α cannot be given in a precise way and that the truth value of α is approximate by an interval $I(\alpha)$. The intended meaning is that the precise truth value of α in a_c is in $I(\alpha)$. In other words we can refer to interval-valued fuzzy subsets to represent the extension of a vague predicate. If we admit such a possibility, then it is necessary to find an analogue of Proposition Valverde ... ? enabling us to define an "interval-valued" similarity by considering interval-valued fuzzy subsets.

Finally, a basic question is related with the search for an optimal strategy in the choice of the queries in the fuzzy framework. Indeed, it is not clear if also in the case of the vague properties, a good strategy is obtained by minimizing the expected value of the entropy.

References

1. Bacchus F., Lp, a Logic for Representing and Reasoning with Statistical Knowledge, Computational Intelligence, 6 (1990) 209-231.
2. Burris S., Sankappanavar H.P., A Course in Universal Algebra, The millennium edition.
3. Calabrò D., Gerla G., Scarpati L., Extension principle and probabilistic inferential process, Lectures on Soft Computing and Fuzzy Logic, Springer-Verlag, (2001) 113-127.

4. Chang C.C., Algebraic analysis of many valued logics, Trans. AMS, 93, (1958) 74-80.
5. Cignoli R., D'Ottaviano I.M.L., Mundici D., Algebraic Foundations of many-valued reasoning, volume 7 of Trends in Logic. Kluwer, Dordrecht, 2000.
6. Dubois D., Hüllermeier E., Prade H., Fuzzy set-based methods in instance-based reasoning, IEEE Transactions on Fuzzy Systems 10, 3, (2002) 322-332.
7. Ganter B., Wille R., Formal Concept Analysis: Mathematical Foundations, Springer, Berlin, 1999.
8. Gerla G., The probability that Tweety is able to fly, International Journal of Intelligent Systems, 9 (1994) 403-409.
9. Hájek P., Metamathematics of Fuzzy Logic, Kluwer Academic Publishers, Dordrecht, 1998.
10. Halpern J. Y., An analysis of first-order logic of probability, Artificial Intelligence, 46, (1990) 331-350.
11. Mundici D., Averaging the truth-value in Lukasiewicz logic, Studia Logica, 55, 1, (1995) 113–127.
12. Novak V., Perfilieva I., Mockor J., Mathematical Principles of Fuzzy Logic, Kluwer Academic Publishers, London, 1999.
13. Pacelli T., Similarities, distances and incomplete information, Ph.D Thesis, DMI, Università degli Studi di Salerno, 2006.
14. Pawlak Z., Rough sets, International Journal of Information and Computer Science, 11, (1982) 341-356.
15. Plaza E., Esteva F., Garcia P., Godo L., Lopez de Mantaras R., A logical approach to case-based reasoning using fuzzy similarity relations, Information Sciences 106,(1998) 105-122.
16. Quan T.T., Hui S.C., Cao T.H., FOGA: A Fuzzy Ontology Generation Framework for Scholarly Semantic Web, ECML/PKDD-2004 KDO Workshop.
17. Valverde L., On the Structure of F-Indistinguishability Operators, Fuzzy Sets and Systems, 17, (1985) 313-328.
18. Wille R., Restructuring lattice theory: an approach based on hierarchies of concepts, in: R. Ivan Rival (Ed.), Ordered Sets, Reidel, Dordecht, Boston, (1982) 445–470.
19. Zadeh L.A., Fuzzy Sets, Information and Control, 8, (1965) 338-353.
20. Zadeh L. A., Similarity relations and fuzzy orderings, Inf. Sci., 3, (1971) 177- 200.
21. Zadeh. L.A., Probability measures of fuzzy events, Journal of Mathematical Analysis and Applications, 23, (1968) 421- 427.