KYBERNETIKA — VOLUME 36 (2000), NUMBER 6, PAGES 689–705

FUZZY DECISION TREES TO HELP FLEXIBLE QUERYING

CHRISTOPHE MARSALA

Fuzzy data mining by means of the fuzzy decision tree method enables the construction of a set of fuzzy rules. Such a rule set can be associated with a database as a knowledge base that can be used to help answering frequent queries. In this paper, a study is done that enables us to show that classification by means of a fuzzy decision tree is equivalent to the generalized modus ponens. Moreover, it is shown that the decision taken by means of a fuzzy decision tree is more stable when observation evolves.

1. INTRODUCTION

Inductive learning raises a *particular* to the *general*. A set of classes $\mathcal{C}$ is considered, representing a physical or a conceptual phenomenon. This phenomenon is described by means of a set of attributes $\mathcal{A} = \{A_1, \dots, A_N\}$. Each attribute A_j can take a value v_{jl} in a given universe X_j. A *description* is an N-tuple of attribute-value pairs (A_j, v_{jl}). Each description is associated with a particular class c_k from the set $\mathcal{C} = \{c_1, \dots, c_K\}$ to make up an *instance* (or *example*, or *case*) e_i of the phenomenon.

Inductive learning is a process to generalize from a *training set* $\mathcal{E} = \{e_1, \dots, e_n\}$ of examples to a general law to bring out relations between descriptions and classes of $\mathcal{C}$.

From a training set, inductive learning enables us to extract knowledge. Such knowledge is used to associate any forthcoming description with a correct decision. For instance, an inductive learning process is the construction of a *decision tree* that is used to classify unknown descriptions. A decision tree is a natural structure of knowledge. Each node in such a tree is associated with a test on the values of an attribute, each edge from a node is labeled with a particular value of the attribute, and each leaf of the tree is associated with a value of the class [29]. However, when the values of attributes for the description change slightly, the decision associated with the previous description can vary greatly. It is a reason to introduce fuzziness in decision trees to obtain *fuzzy decision trees* [5, 18].

Fuzzy decision trees (FDTs) are a generalization of decision trees to handle attributes with numeric-symbolic values, i. e. attributes associated either with a numerical value (for instance, the size is *172 cm*) or a symbolic value corresponding to a numerical measure (for instance, the size is *big*).

FDT is a very promising learning tool because it represents induced knowledge in a very expressive way. The knowledge represented as a FDT is understandable and it differs from black box systems as neural networks.

Moreover, a FDT is equivalent to a set of fuzzy rules [6]. And such kind of induced rules can be introduced to optimize the query process of the database [7, 33] or to deduce decisions from data [1, 2, 15, 16]. FDTs enable us to obtain various kinds of such rules [19]. Thus, it is a powerful knowledge representation. A FDT, as a set of fuzzy rules, can be used as knowledge base to help flexibly querying a database.

Nowadays, literature related to FDT construction is very active. However, few works study the classification step with such a tree. In particular, classification of observation described by means of fuzzy values is not studied.

In this paper, we study the method to use a FDT to classify new cases and to infer new knowledge. We show that a FDT is really equivalent to a fuzzy rule base, and that the use of a FDT is equivalent to the application of the generalized modus ponens. Moreover, there exist several concepts to study the variability of the result provided by a fuzzy system with regard to modifications of the inputs of the system: the most commonly used are the *sensitivity*, *the robustness* and the *stability*. Here, we will deal with the stability of the decision taken by means of a FDT.

This paper is composed as follows: in Section 2 we recall the method of classification by means of a classical decision tree and we highlight its link with the classical modus ponens. In Section 3, we present the method to use FDT to classify new cases. This method is based on the use of a measure of satisfiability to compare fuzzy values. Thus, we show that this method of classification is equivalent to the use of the generalized modus ponens. In Section 4, we study the stability of the decision deduced from a FDT when the observation evolves. In Section 5, we describe our Salammbô software to construct and to test FDT, and we present the application of this kind of software to fuzzy data mining. Moreover, we show the utilization of induced fuzzy knowledge to help flexible querying. Finally we conclude and we present some future works.

2. CLASSIFYING NEW CASES

Inductive learning is composed of two steps. The first one is to bring out knowledge from a set of cases or observations, for instance to construct a decision tree from a set of cases. The second step is to apply this knowledge to classify new cases with this new knowledge.

In this part, we recall the method of classification by means of a classical decision tree, and we show that this method is equivalent to the application of the *modus ponens* with the whole set of rules induced by the tree.

A new example e to classify is described by means of values for each attribute $\{A_1 = w_1; \ldots; A_n = w_n\}$. This example has to be associated with a value for the class. The process of classification with the tree starts with the comparison of testing values of the attribute A_{l_1} present at the root of the tree with the corresponding value of e. Depending on the value w_{l_1}, a vertex going out of that node is used to reach either the next test node, or a leaf of the tree. If a test node is reached, a new comparison is done and the process is resumed until a leaf of the tree is reached. When a leaf is reached, the new example e is associated with the class present in this leaf. All the test nodes encountered from the root to the leaf during this process constitute a path of the decision tree.

It is easy to see that a path of a decision tree is equivalent to an IF... THEN rule. The premises for such a rule r are composed by tests on values of attributes, and the conclusion is the value of the class that labels the leaf of the path: `if` $A_{l_1} = v_{l_1}$ `and` $A_{l_2} = v_{l_2}$ `and ...and` $A_{l_p} = v_{l_p}$ `then` $C = c_k$. So, when the description of e matches each premise of a rule, the class of the conclusion of that rule is associated with e. Otherwise, the rule is not fired.

The modus ponens is the classical way of deduction of knowledge. It has been proposed in classical logic to formalize human reasoning. It is also used in knowledge-based systems.

$$\begin{array}{lccc} \text{Rule :} & P & \Longrightarrow & C \\ \text{Observation :} & P & & \\ \hline \text{Deduction :} & & & C \end{array}$$

The notation $P \Longrightarrow C$ is equivalent to the production rule denoted by `if` P `then` C

Classification by means of a decision tree reflects the modus ponens process. The whole tree is a set of production rules R_i which can be noted $P_i \Longrightarrow C_i$, each rule associated with a path in the tree from the root to a leaf. When classifying a new case, the description of this case is compared with all premises P_i of rules and is associated with the class C_i associated with this premise.

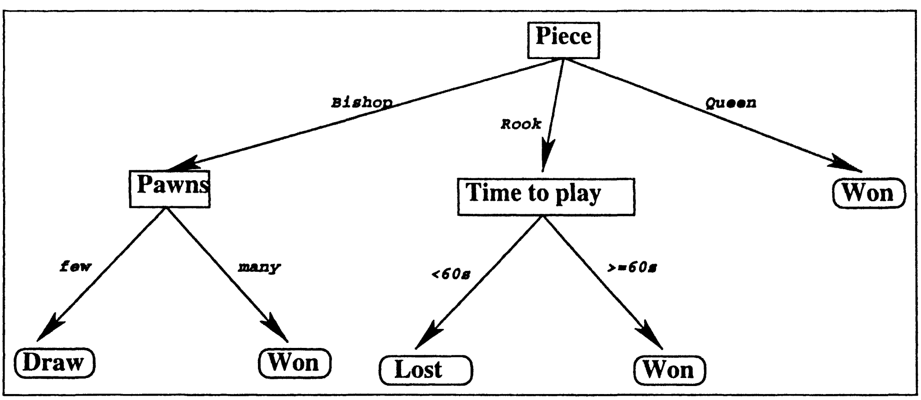

Fig. 1. Example of decision tree.

For instance, five rules are defined by the tree given in Figure 1:

```
R1 :  if Piece = Bishop and Pawns = Few then Draw
R2 :  if Piece = Bishop and Pawns = Many then Won
R3 :  if Piece = Rook and Time < 60s then lost
R4 :  if Piece = Rook and Time ≥ 60s then Won
R5 :  if Piece = Queen then Won
```

The description [Piece = Rook, Time = 56s, Pawns = Few] is compared with the premises of the five rules. It appears that this description fires the rule R3. We have:

Rule :	Piece = Rook and Time < 60s	$\Longrightarrow$	Lost
Observation :	Piece = Rook and Time = 56s		
Deduction :			Lost

3. CLASSIFYING BY MEANS OF A FUZZY DECISION TREE

In a FDT, the values occurring in vertices of the tree are fuzzy values. Thus, a FDT is now equivalent to a fuzzy rule base.

Now, the new example e to classify is described by means of fuzzy values for each attribute $\{A_1 = w_1; \ldots; A_n = w_n\}$.

The comparison between two values[1] highlights the degree that matches their resemblance with regard to a given criterion. In classical decision tree, two values are either similar or not. In fuzzy settings, two objects are similar with a degree of graduality. Several methods exist to measure the resemblance between two values [10, 11, 12, 13, 24].

A hierarchy is proposed by [4] to classify measures of resemblance with regard to their properties. Given an observed value w and a test value v, considered also as fuzzy sets of a given universe X, to measure the similarity $s(w, v)$ between w and v, the following properties are required for s:

- $s(w, v) = 1$ when $w \subseteq v$,
- $s(w, v) = 0$ when $w \cap v = \emptyset$,
- $s(w, v)$ is increasing with $w \cap v$.

The first property is required to normalize the measure. It reflects the fact that the similarity is maximum when the observed value w is equal or included in the test value v. The second property reflects the fact that there is no similarity between two disjoint values. The third property reflects the fact that the more the observed value w is included in the test value v, the higher the similarity of w and v.

These properties are based on the following. The inclusion between subsets of X is defined as: for all v, w, $w \subseteq v \Leftrightarrow w \cap v = w$. Thus, given two fuzzy sets v and

[1] We denote "value" both symbolic, or numerical value, and also fuzzy value.

w with membership functions μ_v and μ_w, if $\mu_{w\cap v}$ is the membership function of the intersection of these two sets, we have $w \subseteq v \Leftrightarrow \mu_{w\cap v} = \mu_w$. The intersection $w \cap v$ is increasing when for all v, w_1, and w_2, we have $(w_1 \cap v) \subseteq (w_2 \cap v) \Leftrightarrow s(w_1, v) \leq s(w_2, v)$.

From the hierarchy of measures by [4], it can be shown that the measure s is not necessarily a classical measure of similarity because the property of symmetry is not required (i e. it is not required that for all v, w, $s(w, v) = s(v, w)$) It can be shown that the measure s is a measure of satisfiability (also known as measure of inclusion) in the sense of [4]. An example of such a measure of satisfiability is given by [4, 32]:

$$s(w, v) = \frac{\mathcal{M}(w \cap v)}{\mathcal{M}(w)} \tag{1}$$

where $\mathcal{M}$ is a measure of fuzzy sets[2] [32] It can be noted that this measure is the normalized version of the measure of fuzzy sets introduced by [10].

For the construction of fuzzy decision trees, as previously introduced by [31], the measure $\mathcal{M}(A)$ generally used is defined on $\Omega_A \subseteq \Omega$, for all $A \in F(\Omega)$, with a membership function μ_A in $[0, 1]$, as follows:

- if Ω_A is a continuous universe of values: $\mathcal{M}(A) = \int_{\Omega_A} \mu_A(x)\,\mathrm{d}x$
- if Ω_A is a discrete set of values: $\mathcal{M}(A) = \sum_{x\in\Omega_A} \mu_A(x)$.

When Ω_A is a discrete set of values, $\mathcal{M}$ is the well-known *relative sigma-count* [35]. In this paper, for the sake of simplicity, we denote $\int_{\Omega_A} \mu_A(x)\,\mathrm{d}x$ the measure $\mathcal{M}$ when Ω_A is either a continuous universe or a discrete set.

Thus, the measure of satisfiability used in the software *Salammbô* (see Section 5) of construction of FDT [18, 22] is:

$$\mathtt{Deg}(w, v) = \frac{\int_X \mu_{w\cap v}\,\mathrm{d}x}{\int_X \mu_w\,\mathrm{d}x} \quad \text{if} \quad \int_X \mu_w\,\mathrm{d}x \neq 0 \tag{2}$$

where μ_w is the membership function associated with the value w, $\mu_{w\cap v}$ is the membership function associated with the intersection $w \cap v$ of w and v, X is the universe of values where μ_w and $\mu_{w\cap v}$ are defined. In the case where $\int_X \mu_w\,\mathrm{d}x = 0$, we set $\mathtt{Deg}(w, v) = 0$.

3.1. Comparison of an observation and premises

Given the rule **if** $A_{l_1} = v_{l_1}$ **and** ...**and** $A_{l_p} = v_{l_p}$ **then** $C = c_k$, a comparison is done between the description of example e and the premise of the rules.

To value the resemblance between $W = (w_{l_1}, \ldots, w_{l_p})$ and $V = (v_{l_1}, \ldots, v_{l_p})$, the resemblance of each component w_{l_i} and v_{l_i} is done. We denote $\mathtt{Deg}(w_{l_i}, v_{l_i})$ this

[2] We recall that a measure of fuzzy sets is defined as:

Definition 1 (Measure of fuzzy sets). Let Ω be a set of elements, $F(\Omega)$ be the set of subsets of Ω, and given $\subseteq$ an order on elements of $F(\Omega)$. A *measure of fuzzy sets* $\mathcal{M}$ is a function defined from $F(\Omega)$, to $\mathbf{R}^+$, such that $\mathcal{M}(\emptyset) = 0$ and $\forall A, B \in F(\Omega)$, $B \subseteq A \Rightarrow \mathcal{M}(B) \leq \mathcal{M}(A)$.

measure. Thus, to obtain the whole resemblance degree $\mathtt{Deg}(W, V)$ for W and V, all these degrees $\mathtt{Deg}(w_{l_i}, v_{l_i})$ are aggregated:

$$\mathtt{Deg}(W, V) = \mathtt{Deg}((w_{l_1}, \dots, w_{l_p}), (v_{l_1}, \dots, v_{l_p})) = \top_{i=1\dots p} \mathtt{Deg}(w_{l_i}, v_{l_i}). \quad (3)$$

This aggregation is done thanks to the triangular norm (t-norm) $\top$ in order to highlight the conjunctive link between all these values in the premise. Depending on the measure of satisfiability used, the t-norm must be chosen carefully in order to preserve the equality. That will enable us to ensure the cognitive process used in this kind of reasoning: the resemblance between two complex objects is related to the resemblance between each of their corresponding parts.

Thus, in our context, it can be shown that this property is obtained for the measure of satifiability we used (eq. 2) and the product t-norm.

Proof. For the sake of simplicity, in this part, we denote w_{l_i} by w_i.
From (eq. 2), when $\int_X \mu_{(w_1,\dots,w_p)}\,\mathrm{d}x \neq 0$, we have:

$$\mathtt{Deg}((w_1, \dots, w_p), (v_1, \dots, v_p)) = \frac{\int_X \mu_{((w_1,\dots,w_p)\cap(v_1,\dots,v_p))}\,\mathrm{d}x}{\int_X \mu_{(w_1,\dots,w_p)}\,\mathrm{d}x} \quad (4)$$

with $X = X_1 \times \dots \times X_p$, where X_i is the universe of values of attribute A_i.

The membership function $\mu_{(w_1,\dots,w_p)}$, defined on X, cartesian product of the X_i's, is deduced from the membership functions μ_{w_i} by means of a t-norm: for all $x = (x_1, \dots, x_p) \in X$, $\mu_{(w_1,\dots,w_p)}(x) = \top_{i=1\dots p}\ \mu_{w_i}(x_i)$. Here, $\int_X \mu_{(w_1,\dots,w_p)}\,\mathrm{d}x = 0$ is equivalent to $w_i(x_i) = 0$ for at least one i, and thus, for that i, we have $\mathtt{Deg}(w_i, v_i) = 0$. In this case, (eq. 3) is trivial:

$\prod_{i=1\dots p} \mathtt{Deg}(w_i, v_i) = \mathtt{Deg}((w_1, \dots, w_p), (v_1, \dots, v_p)) = 0.$

The membership function $\mu_{((w_1,\dots,w_p)\cap(v_1,\dots,v_p))}$ is defined from the membership functions $\mu_{(w_1,\dots,w_p)}$ and $\mu_{(v_1,\dots,v_p)}$ by means of the t-norm used as intersection operator. It is generally the same t-norm as the cartesian product: $\mu_{((w_1,\dots,w_p)\cap(v_1,\dots,v_p))} = \top(\mu_{(w_1,\dots,w_p)}, \mu_{(v_1,\dots,v_p)})$.

In the case where the product t-norm is used as both conjunctive operator of fuzzy sets, and cartesian product, we have, for all $x = (x_1, \dots, x_p) \in X$,

$\mu_{((w_1,\dots,w_p)\cap(v_1,\dots,v_p))}(x) = \prod_{i=1\dots p} \mu_{w_i}(x_i) \,.\, \mu_{v_i}(x_i)$. Thus, on $X = X_1 \times \dots \times X_p$, we have:

$\int_X \mu_{((w_1,\dots,w_p)\cap(v_1,\dots,v_p))}\,\mathrm{d}x = \int_X \left(\prod_{i=1\dots p} \mu_{w_i}(x_i) \,.\, \mu_{v_i}(x_i)\right) \mathrm{d}x$. And thus, if the universes X_i are supposed independent from each other, (in this case $\mathrm{d}x = \mathrm{d}x_1 \dots \mathrm{d}x_p$, we have: $\int_X \mu_{((w_1,\dots,w_p)\cap(v_1,\dots,v_p))}\,\mathrm{d}x = \prod_{i=1\dots p} \int_{X_i} \mu_{w_i}(x_i) \,.\, \mu_{v_i}(x_i)\,\mathrm{d}x_i$. With the same hypothesis of independence of the universes of values, we have $\int_X \mu_{(w_1,\dots,w_p)}\,\mathrm{d}x = \prod_{i=1\dots p} \int_{X_i} \mu_{w_i}(x_i)$. As a consequence, (eq. 3) can be found from (eq. 4):

$\mathtt{Deg}((w_1, \dots, w_p), (v_1, \dots, v_p)) = \prod_{i=1\dots p} \frac{\int_{X_i} \mu_{w_i \cap v_i}(x_i)\,\mathrm{d}x_i}{\int_{X_i} \mu_{w_i}(x_i)\,\mathrm{d}x_i}$ and thus
$\mathtt{Deg}((w_1, \dots, w_p), (v_1, \dots, v_p)) = \prod_{i=1\dots p} \mathtt{Deg}(w_i, v_i)$. □

3.2. Classifying with a fuzzy decision tree

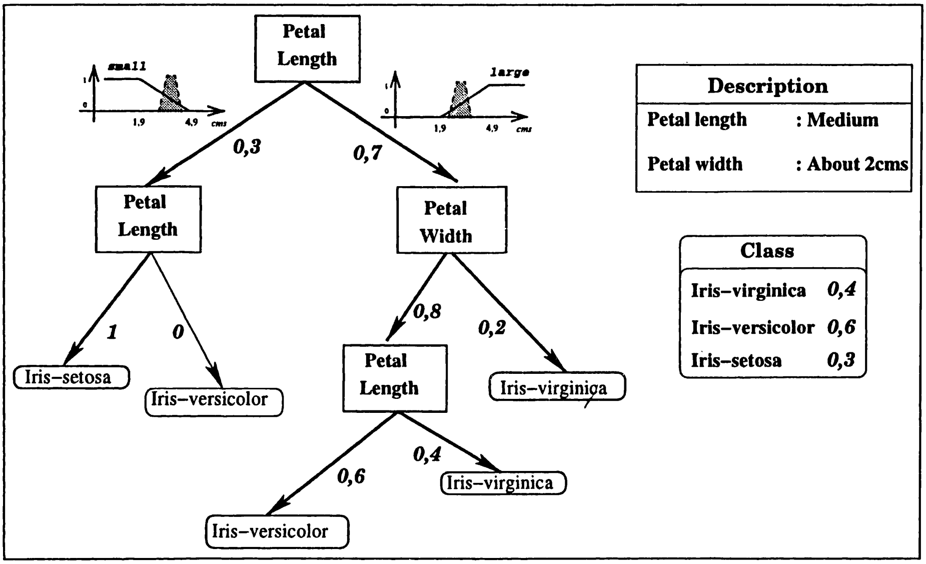

Fig. 2. Classification with a fuzzy decision tree.

In a FDT, a leaf can be labeled by a set of classes $\{c_1, \ldots, c_K\}$, each c_j associated with a weight $P^*(c_j|(v_{l_1}, \ldots v_{l_p}))$ computed during the learning phase. This weight is the probability of c_j for training examples with values $(v_{l_1}, \ldots v_{l_p})$ weighted by their membership degree to the leaf. This weight does not exist in classical decision trees because training values are crisp, and a training example belongs to only one leaf. So, in this case, $P^*(c_j|(v_{l_1}, \ldots v_{l_p}))$ is equal to 1 for each c_j. Thus, a path of a FDT is equivalent to the following rule: if $A_{l_1} = v_{l_1}$ and ...and $A_{l_p} = v_{l_p}$ then $C = c_1$ with the degree $P^*(c_1|(v_{l_1}, \cdots v_{l_p}))$ and ...and $C = c_K$ with the degree $P^*(c_K|(v_{l_1}, \cdots v_{l_p}))$.

For each premise, a degree of satisfiability $\texttt{Deg}(w_{l_i}, v_{l_i})$ is computed for the corresponding value w_{l_i}. Finally, given a rule r, a description is associated with the class c_j with a *final degree* of satisfiability $\texttt{Fdeg}_r(c_j)$ that corresponds to the satisfiability of the description to the premises of the rule r weighted by the conditional probability for c_j according to the rule r in order to take into account the confidence of the rule: $\texttt{Fdeg}_r(c_j) = \top_{i=1\ldots p}\texttt{Deg}(w_{l_i}, v_{l_i}).P^*(c_j|(v_{l_1}, v_{l_2}, \cdots v_{l_p}))$.

Final degrees computed from all the rules are aggregated by means of a triangular conorm $\perp$ (for instance, the *maximum* triangular conorm) to obtain a single degree of satisfiability $\texttt{Fdeg}(c_j)$. If n_p is the number of rules given by the FDT, we have $\texttt{Fdeg}(c_j) = \perp_{r=1\ldots n_p}\texttt{Fdeg}_r(c_j)$.

For each value of the class, the description e is associated with such a degree of satisfiability $\texttt{Fdeg}(c_j)$ for each class c_j computed from the whole set of rules. The class c_e associated with e can be chosen as the class with the highest degree of satisfiability: $\texttt{FDeg}(c_e) = \max_{j=1\ldots K} \texttt{FDeg}(c_j)$. Such a process of aggregation of degrees is used in order to have meaningful values of degrees for each class.

For instance, the whole process of classification of an example is summarized in Figure 2. A given example, the *description* of which is described by means of fuzzy values, must be classified. Each of its values for corresponding attributes are compared with each value labeling a vertex in the tree. The value of this degree is reported on the corresponding edge of the tree. The final degree for each class is computed as mentioned and is reported in the figure.

3.2.1. The fuzzy classification is a generalized modus ponens

The *generalized modus ponens* (GMP) is an extension of the classical modus ponens to handle fuzzy data [34]:

$$\begin{array}{lccc} \text{Rule :} & P & \Longrightarrow & C \\ \text{Observation : } . & P' & & \\ \hline \text{Deduction :} & & & C' \end{array}$$

Thus, observing a value P' close to the premise P of the rule enables the construction of a conclusion C' close to C. This deduction process is smoother than the classical one and reflects better the human process of reasoning: in spite of the fact that it is extremely rare that the exactly same cause appears several times, the human brain is always able to deduce knowledge.

The membership function of the conclusion C' is deduced from the membership functions of P, C, and P' as [34]:

$$\forall y \in Y, \;\; \mu_{C'}(y) = \sup_{x \in X} \top_m(\mu_{P'}(x), f_R(x,y)) \tag{5}$$

with $\mu_{C'}$ and $\mu_{P'}$ as the respective membership functions of C' and P', Y the universe where the conclusion is defined, X the universe where the premise is defined, f_R is the fuzzy implication that describes the link between the premise and the conclusion, and $\top_m$ is the GMP operator [3].

It can be shown that, under particular hypothesis, the classification by means of a fuzzy decision tree described in Section 3.2 is equivalent to a GMP.

Let the product t-norm be the implication operator (which does not generalize the classical implication), and the GMP operator. Thus, if μ_C and μ_P are the respective membership functions of C and P, we have $f_R(x,y) = \mu_P(x) \,.\, \mu_C(y)$, and $\top_m(\mu_{P'}(x), f_R(x,y)) = \mu_{P'}(x) \,.\, f_R(x,y)$. So, in this case, the equation 5 becomes for all $y \in Y$, $\mu_{C'}(y) = \sup_{x\in X}(\mu_{P'}(x) \,.\, \mu_P(x) \,.\, \mu_C(y))$ and thus for all $y \in Y$, $\mu_{C'}(y) = \mu_C(y) \,.\, \sup_{x\in X}(\mu_{P'}(x) \,.\, \mu_P(x))$. Now, let the product t-norm be the intersection operator. If $\mu_{P'}$ and μ_P are normalized on X, the measure $\sigma(P', P)$ defined as $\sigma(P', P) = \sup_{x\in X}(\mu_{P'}(x) \,.\, \mu_P(x))$ is a measure of satisfiability.

Proof. It is easy to see that:

— if $P' \subseteq P$ then $\exists x \in X,\ \ \mu_{P'}(x) = \mu_P(x) = 1$ thus $\sigma(P', P) = 1$

— if $P' \cap P = \emptyset$ then $\forall x \in X,\ \ \mu_{P'}(x) \,.\, \mu_P(x) = 0$ thus $\sigma(P', P) = 0$

— if $P' \cap P$ is increasing, then $\sup_{x \in X}(\mu_{P'}(x) \,.\, \mu_P(x))$ is increasing too, as a consequence $\sigma(P', P)$ is increasing with respect to $P' \cap P$. □

Consequently, $\sigma(P', P)$ can be considered as the global degree $\mathtt{Fdeg}_r(C)$ obtained according to the rule r for the class C. Thus, the analogy between the generalized modus ponens and the classification by means of a fuzzy decision tree has been pointed out.

4. STABILITY OF FUZZY DECISION TREES

We do not consider fuzzy control where the stability of a system is an important part of research. This stability of a system concerns the research of an equilibrium point after a modification of the input data of the system [9].

In other domains, several works have been done on the study of the stability of fuzzy models. For instance, the sensitivity and the robustness of a system are studied in [27]. The authors introduced two kind of measures, the measure of sensitivity $se(x, \varepsilon)$ which is defined as the distance between a linguistic label $A(x)$ (a fuzzy set defined on values x from a given universe X) and a disturbed version $A(x + \varepsilon)$, and the measure of robustness $ro(x, \varepsilon)$ which is defined as $ro(x, \varepsilon) = 1 - se(x, \varepsilon)$.

A study of the robustness or the sensitivity of a binary operator is done in other works [25, 26].

A more general study is introduced in [28]. In this work, the sensitivity of fuzzy models based on linguistic rule bases is studied. The modification of the input data leads to a modification of the output value of the fuzzy system. The link between these two modifications is studied considering the operators used to infer the output value from the input data.

In this part, a study of the stability of fuzzy decision trees is done. The stability of these trees when classifying evoluting observations is proved.

4.1. Evolution of an observation and classical decision tree

With a classical decision tree, a small variation of values near the boundaries of the values present in the node of the tree can imply an important change of the decision. As mentioned, when the decision is inferred by means of the classical method, values of the description are compared to premises to fire rules. Thus, a small change in the value describing an observation will modify the matching of this observation with premises of rules, it will change the fired rule and thus, it can give a new unpredictable value for the decision.

4.2. Evolution of an observation and fuzzy decision tree

On the contrary, the value of the decision taken by a FDT is continuous relatively to small changes in the values of the description of the observation. This continuity results from the continuity of the measure of satisfiability and the continuity of the inference method used to aggregate the satisfiability degrees. Moreover, the measure of satisfiability in the fuzzy case is computed as a value from $[0,1]$ instead of a crisp value from $\{0,1\}$ in classical trees. Therefore, changes from a rule to another is gradual and fired rules are less dependent on small variations of the values of the description. This gradation is related to the chosen measure of satisfiability and the chosen operators of aggregation.

In a first step, given a reference value v, we study the continuity of the measure of satisfiability Deg when a new fuzzy value $\tilde{w}$, near a previous fuzzy value w, is observed. This new value has a satisfiability degree $\mathtt{Deg}(\tilde{w}, v)$. When $\tilde{w}$ is slightly different from w, then it can be said that an error function ϵ is defined on X, lying on $[-1, 1]$: for all $x \in X$, $\mu_{\tilde{w}}(x) = \mu_w(x) + \epsilon(x)$.

For instance, such an error function is given in Figure 3.

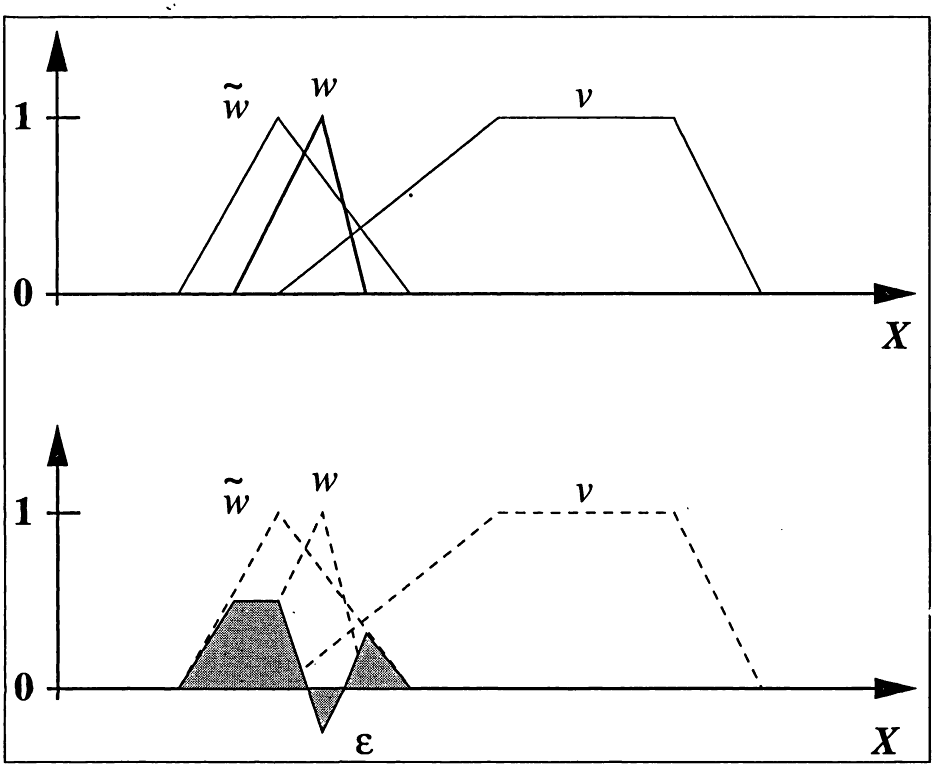

Fig. 3. Example of error function

Continuity of the measure of satisfiability

To establish the continuity of the measure of satisfiability Deg with w, given the measure (eq. 2), it can be proven that for all σ, there exists α, such that for all $x \in X$, $|\mu_{\tilde{w}}(x) - \mu_w(x)| \leq \alpha \Rightarrow |\mathtt{Deg}(\tilde{w}, v) - \mathtt{Deg}(w, v)| \leq \sigma$.

P r o o f. To study the continuity of Deg with regard to w, we study the evolutions of $|\text{Deg}(\tilde{w},v) - \text{Deg}(w,v)|$ with regard to $|\tilde{w} - w|$. For instance, we use the degree of satisfiability previously given (eq. 2). With the product t-norm as intersection operator, we have[3]

$$\text{Deg}(w,v) = \frac{\int_X \mu_w \mu_v \, dx}{\int_X \mu_w \, dx} = \frac{\int_X (\mu_{\tilde{w}} - \epsilon)\mu_v \, dx}{\int_X \mu_w \, dx} = \left(\text{Deg}(\tilde{w},v)\frac{\int_X \mu_{\tilde{w}} \, dx}{\int_X \mu_w \, dx}\right) - \frac{\int_X \epsilon\mu_v \, dx}{\int_X \mu_w \, dx}.$$

Thus, we have $\text{Deg}(\tilde{w},v) - \text{Deg}(w,v) = \text{Deg}(\tilde{w},v)\left(1 - \frac{\int_X \mu_{\tilde{w}} \, dx}{\int_X \mu_w \, dx}\right) + \frac{\int_X \epsilon\mu_v \, dx}{\int_X \mu_w \, dx}$,

and finally $\text{Deg}(\tilde{w},v) - \text{Deg}(w,v) = -\text{Deg}(\tilde{w},v)\frac{\int_X \epsilon \, dx}{\int_X \mu_w \, dx} + \frac{\int_X \epsilon\mu_v \, dx}{\int_X \mu_w \, dx}$.

Thus, we have $|\text{Deg}(\tilde{w},v) - \text{Deg}(w,v)| = \left|-\text{Deg}(\tilde{w},v)\frac{\int_X \epsilon \, dx}{\int_X \mu_w \, dx} + \frac{\int_X \epsilon\mu_v \, dx}{\int_X \mu_w \, dx}\right|$.

As a consequence, $|\text{Deg}(\tilde{w},v) - \text{Deg}(w,v)|$ can be upper bounded by means of the triangular inequality:

$$\begin{aligned} |\text{Deg}(\tilde{w},v) - \text{Deg}(w,v)| &\leq \text{Deg}(\tilde{w},v)\frac{\left|\int_X \epsilon \, dx\right|}{\int_X \mu_w \, dx} + \frac{\left|\int_X \epsilon\mu_v \, dx\right|}{\int_X \mu_w \, dx} \\ &\leq \frac{\int_X |\epsilon| \, dx}{\int_X \mu_w \, dx} + \frac{\int_X |\epsilon|\mu_v \, dx}{\int_X \mu_w \, dx}. \end{aligned} \quad (6)$$

Thus, if there exists a positive real value α such that for all $x \in X$, $|\epsilon(x)| \leq \alpha$ then $\int_X |\epsilon(x)| \, dx \leq \alpha \int_X \, dx$. And by denoting the supposed finite integrals as $M(w) = \int_X \mu_w \, dx$, $M(v) = \int_X \mu_v \, dx$ and $M = \int_X \, dx$, from (eq. 6) we have $|\text{Deg}(\tilde{w},v) - \text{Deg}(w,v)| \leq \frac{\alpha M + \alpha M(v)}{M(w)}$.

By definition, for all $x \in X$, $\mu_v(x) \leq 1$ and thus $M(v) \leq M$, we have $|\text{Deg}(\tilde{w},v) - \text{Deg}(w,v)| \leq \frac{2\alpha M}{M(w)}$. That concludes on the continuity of Deg with regard to w: for all $\sigma > 0$, $\exists\, \alpha = \frac{\sigma M(w)}{2M}$ such that for all $x \in X, |\epsilon(x)| \leq \alpha \Rightarrow |\text{Deg}(\tilde{w},v) - \text{Deg}(w,v)| \leq \sigma$. Or for all $\sigma > 0$, there exists $\alpha = \frac{\sigma M(w)}{2M}$ such that for all $x \in X$, $|\mu_{\tilde{w}}(x) - \mu_w(x)| \leq \alpha \Rightarrow |\text{Deg}(\tilde{w},v) - \text{Deg}(w,v)| \leq \sigma$. □

Thus, the variation of the degree of satisfiability is continuous with respect to the variation of the observation.

Evolution from a precise value to a fuzzy value

We focus now on the problem of the evolution of a precise value w for the observation into a fuzzy value $\tilde{w}$ such that $\tilde{w}$ is *near* w. Let v be a fuzzy value, the degree of satisfiability of the precise value w related to v is computed as $\text{Deg}(w,v) = \mu_v(w)$ (see Section 2) and we assume that μ_v is continuous.

When $\tilde{w}$ is also a precise value, the variation of the degree is continuous because in this case $\text{Deg}(\tilde{w},v) = \mu_v(\tilde{w})$.

When $\tilde{w}$ is a fuzzy value, $\text{Deg}(\tilde{w},v)$ converges to $\text{Deg}(w,v) = \mu_v(\tilde{w})$. Let $\tilde{w}$ be a fuzzy value with w as modal value and $[w - \beta, w + \beta]$ as support set. We suppose

[3]For the sake of simplicity, we denote in this part $\mu_{\tilde{w}}(x)$, $\mu_{\tilde{w}}(x)$ and $\epsilon(x)$ by $\mu_{\tilde{w}}$, $\mu_{\tilde{w}}$ and ϵ.

that w belongs to the support of the fuzzy value v and does not belong to its kernel. When $x \in [w-\beta, w]$, $\mu_{\tilde{w}}(x) = \frac{x+\beta-w}{\beta}$ and when $x \in [w, w+\beta]$, $\mu_{\tilde{w}}(x) = \frac{-x+\beta+w}{\beta}$. Moreover, let $\forall x \in [w-\beta, w+\beta]$, μ_v be a straight line valued as $\mu_v(x) = ax + b$.

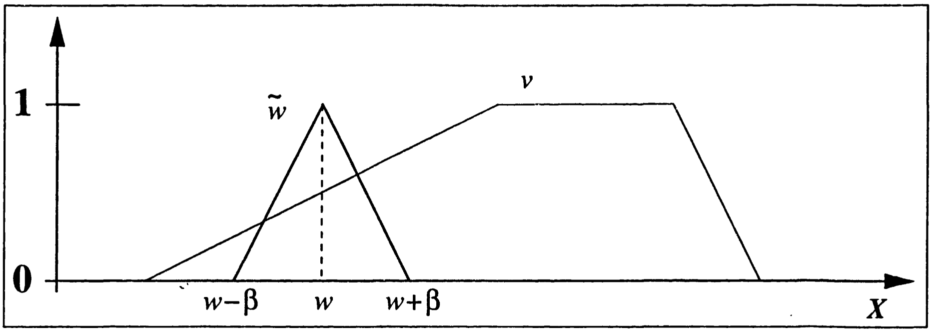

Fig. 4. Example of evolution from crisp to fuzzy value (case $(w-\beta) > -\frac{b}{a}$).

The degree of satisfiability is continuous if $\lim_{\beta\to 0} \texttt{Deg}(\tilde{w}, v) = \texttt{Deg}(w, v)$.

— When $(w-\beta) > -\frac{b}{a}$ (see Figure 4), we have $\int_X \mu_{\tilde{w}\cap v}(x)\, dx = \mu_v(w)\beta$, and thus: $\texttt{Deg}(\tilde{w}, v) = \frac{\mu_v(w)\beta}{\beta} = \mu_v(w) = \texttt{Deg}(w, v)$.

— When $w = -\frac{b}{a}$, we have $\texttt{Deg}(w, v) = \mu_v(w) = \mu_v(-\frac{b}{a}) = 0$, thus $\texttt{Deg}(\tilde{w}, v) = \frac{\frac{a\beta^2}{6}}{\beta} = \frac{a\beta}{6}$, and thus $\lim_{\beta\to 0} \texttt{Deg}(\tilde{w}, v) = \lim_{\beta\to 0} \frac{a\beta}{6} = 0$.

Finally, given a fuzzy value of reference v, the degree of satisfiability of $\tilde{w}$ tends towards the degree of satisfiability of w when β tends towards 0.

— When $w < -\frac{b}{a}$, the degree of satisfiability of w related to v is equal to 0. A modification of w such that $\tilde{w} = w + \beta \leq -\frac{b}{a}$ will have a degree of satisfiability to v also equal to 0. On the contrary, when β is such that $w + \beta > -\frac{b}{a}$, the previously mentioned proof is repeated to prove that the degree tends towards 0.

— When $w = -\frac{1-b}{a}$, the same proof is repeated to prove that $\texttt{Deg}(\tilde{w}, v)$ tends towards $\texttt{Deg}(w, v)$

All the proofs given here deal with observed values in one of the slopes of v. All other proofs are equivalent for each case and they are not given here.

To conclude, in the next part, we prove that with the given measure of satisfiability, a slight change in the observed value leads to a slight modification of the value of the degree of satisfiability.

4.2.1. Evolution of the degree for the conclusion of the rule

After having checked the continuity of the measure of satisfiability when the observation evolves, we prove now the continuity of the degree associated with the conclusion of a rule.

As mentioned, the degrees from all the premises of a rule are aggregated to compute a global degree (see Section 3.2). This aggregation is done by means of the product t-norm.

Let us call $D(w)$ the value of this product. When one value of the observation changes slightly, we proved previously that the variation of the degree is also slight: Let w_{l_k} be the prime value and let $\tilde{w_{l_k}} = w_{l_k} + \epsilon$ be the evolved value. We have proven that $\mathtt{Deg}(w_{l_k}, v_{l_k})$ becomes $\mathtt{Deg}(w_{l_k}, v_{l_k}) + \varepsilon$, with small values for $\varepsilon \in \mathbb{R}$. The global degree $D(\tilde{w})$ is:

$$D(\tilde{w}) = (\mathtt{Deg}(w_{l_k}, v_{l_k}) + \varepsilon) \prod_{i=1\ldots k-1,k+1,\ldots,p} \mathtt{Deg}(w_{l_i}, v_{l_i}).$$

Or

$$D(\tilde{w}) = D(w) + \varepsilon \left(\prod_{i=1\ldots k-1,k+1,\ldots,p} \mathtt{Deg}(w_{l_i}, v_{l_i}) \right),$$

and

$$|D(\tilde{w}) - D(w)| = \left| \varepsilon \left(\prod_{i=1\ldots k-1,k+1,\ldots,p} \mathtt{Deg}(w_{l_i}, v_{l_i}) \right) \right|,$$

and thus: $|D(\tilde{w}) - D(w)| \leq |\varepsilon|$ since $\forall\, w_{l_i}, v_{l_i}$, $\mathtt{Deg}(w_{l_i}, v_{l_i}) \leq 1$.

As a consequence, we deduce the continuity of $D(w)$ with respect to w: $\forall\, \sigma > 0$, $\exists\, \alpha = \sigma$ such that $\forall\, x \in X$, $|\varepsilon| \leq \alpha \Rightarrow |D(\tilde{w}) - D(w)| \leq \sigma$.

As a consequence, we prove that the classification by means of the fuzzy method is more stable than the classification by means of a classical method.

5. THE SALAMMBÔ SOFTWARE

We have implemented the Salammbô software to build fuzzy decision trees [5, 18]. This software enables us to test several kinds of parameters during the construction of the FDT, and during the use of this tree to classify a new case.

The construction of a FDT is done by means of measures chosen in a general family we have introduced on the basis of interesting properties required for the discrimination process (the choice of attributes to construct nodes of the tree) [23]. For instance, the FDT can be constructed by means of the Shannon entropy, the Gini test of impurity, or a fuzzy measure of entropy. Moreover, an automatic method to build a fuzzy partition on the set of values of numerical attributes is introduced that enables us to avoid the prior definition of fuzzy values of attributes by an expert [17].

During the process of classification of new cases with the FDT, various parameters (t-norms, t-conorms, implications operators) have been implemented to be used in the Salammbô software. It enables us to test them and to compare them in the process of classification on different kinds of databases [18]. These experiments highlight the higher accuracy of fuzzy decision trees when classifying a test set of examples.

5.1. Fuzzy decision tree to help flexible queries

Nowadays, a large amount of data is contained in a lot of databases. These data represent an important source of potential knowledge to use. Data mining is the process of mining databases in order to induce such knowledge [14]. Fuzzy data mining is concerned by the mining of fuzzy knowledge [8, 22].

For instance, the FDT construction method can be applied to mine a database and extract fuzzy knowledge [19]. The fuzzy rule base induced as a FDT is a new form of knowledge that can be associated with the database. This knowledge can be used in different ways.

First of all, it enables us to improve the querying process of the database. A set of rules can be introduced to optimize the query process of the database [7, 33] or to deduce decisions from data [1, 2, 15, 16]. Such a set of induced fuzzy rules can be associated with a database as a knowledge base that can be used to help answering frequent queries. A fast response can be found for a query on the value of an attribute. It can also lower the conditions on the values of attributes for a query, before the process of examination of the database [7]. Moreover, fuzzy rules can take advantage of the fuzziness of their values to take into account new numerical or fuzzy values. The method of classification with such a set of fuzzy rules is a good way to handle new values for attributes.

Secondly, a set of fuzzy rules is completely understandable and a decision taken by means of these rules is explainable. It can be used as a new knowledge on the domain of the database, and it can be understood by any expert of this domain.

5.2. Fuzzy decision trees applied to data mining

Tests have been conducted that highlighted the interest of fuzzy decision trees for data mining processes, laying in their expressiveness, and their qualities when handling numerical attributes.

For instance, in chemistry, associations between the structures of chemical compounds and the quality of their odors have been brought out by means of the construction of FDT with the *Salammbô* software [20]. Applications were conducted to extract knowledge from other kinds of domains: from a geographical oriented-object spatial database [21], and from an electrical domain database [18].

Here, we present some results of two construction methods, obtained on two commonly used training databases with a cross validation test. These databases are available on the ftp site of the University of Irvine, California[4].

The first comparison concerns the iris database. In this database, examples are described by means of 4 numeric attributes and there are 3 classes to recognize. The second comparison concerns the database of Breiman's waveforms. In this database,

[4]ftp://ftp.ics.uci.edu/pub/machine-learning-databases

Table 1. Comparison of classification by classical and fuzzy decision trees.

Base	*Method*	*Size*	*Classification rate*
Iris	Classic	8.5	95.2%
	Salammbô	4.0	96.0%
Waveform	Classic	44.6	72.7%
	Salammbô	66.9	78.2%

examples are described by means of 21 numeric attributes and there are 4 classes to recognize.

In Table 1, results with the *Classic* method concern the classical ID3 method and are those given by [30] for the C4.5 algorithm, adapted to numeric attributes. In this case, the decision trees are also pruned. Results with the Salammbô software concern the construction of fuzzy decision trees with a fuzzy measure of entropy as measure to select attributes (see [18]). The *Size* is the average number of paths of the built trees, and the *Classification rate* is the number of test examples that are well classified by means of the built tree.

It can be observed that the fuzzy decision trees provide better classification rate than the classical decision trees. In the case of the iris database, the size of fuzzy decision trees, even with no pruning, is highly smaller than the size of the classical decision tree. In the case of the waveform database, the size of the tree should be minimized by means of a pruning phase.

6. CONCLUSION AND FUTURE WORKS

In this paper, a study is done that enables us to show that classification by means of a fuzzy decision tree is equivalent to the generalized modus ponens.

The proposed method of classification by means of a fuzzy decision tree lies on the use of measure of satisfiability to compare observed values to testing values. It has been proved that such a method ensures the cognitive process used in this kind of reasoning.

Moreover, we proved that using a fuzzy decision tree instead of a classical decision tree, a slight change in the values of a description leads to a slight change in the value of the decision. Given a measure of satisfiability and a process of inference by means of a fuzzy decision tree, we showed that a continuity in the value of the decision is obtained relatively to the values of the description. Thus, the stability of fuzzy decision trees when classifying evoluting observations results from this continuity.

In future works, we will study the links existing between the measure of satisfiability and the inference method. Moreover, we will also analyze the stability of the method of construction of fuzzy decision tree when the slight changes occur in the data pertaining to the training set.

(Received June 13, 2000.)

REFERENCES

[1] R. Agrawal, T. Imielinski, and A. Swami: Mining association rules between sets of items in large databases. In: Proc. ACM-SIGMOD Internat. Conference on Management of Data, Washington DC 1993, pp. 207–216.

[2] R. Agrawal, T. Imielinski, and A. Swami: Database mining: A performance perspective. IEEE Trans. on Knowledge and Data Engineering *5* (1993), 6, 914–925.

[3] B. Bouchon–Meunier: La logique floue et ses applications. Collection Vie Artificielle. Addison Wesley France, 1995.

[4] B. Bouchon–Meunier, M. Rifqi, and S. Bothorel: Towards general measures of comparison of objects. Fuzzy Sets and Systems *84* (1996), 2, 143–153.

[5] B. Bouchon–Meunier, C. Marsala, and M. Ramdani: Learning from imperfect data. In: Fuzzy Information Engineering: a Guided Tour of Applications (D. Dubois, H. Prade, and R. R. Yager, eds.), Wiley, New York 1997, pp. 139–148.

[6] B. Bouchon–Meunier and C. Marsala: Learning fuzzy decision rules. In: Fuzzy Sets in Approximate Reasoning and Information Systems, volume 3 of Handbook of Fuzzy Sets, chapter 4 (D. Dubois J. Bezdek and H. Prade, eds.), Kluwer, Dordrecht 1999.

[7] I-M. A. Chen: Query answering using discovered rules. In: Proc. 12th Internat. Conference on Data Engineering (S. Y. W. Su, ed.), IEEE Computer Society Press, New Orleans 1996, pp. 402–411.

[8] K. J. Cios, W. Pedrycz, and R. W. Swiniarski: Data Mining – Methods for Knowledge Discovery. (Engineering and Computer Science.) Kluwer, Dordrecht 1998.

[9] D. Driankov, H. Hellendoorn, and M. Reinfrank: An Introduction to Fuzzy Control. Springer Verlag, Berlin 1993.

[10] D. Dubois and H. Prade: A unifying view of comparison indices in a fuzzy set-theoretic framework. In: Fuzzy Set and Possibility Theory (R. R. Yager, ed.), Pergamon Press 1982, pp. 3–13.

[11] D. Dubois, H. Prade, and C. Testemale: Weighted fuzzy pattern matching. Fuzzy Sets and Systems *28* (1988), 3, 313–331.

[12] D. Dubois, X. Mo, and H. Prade: Fuzzy-valued variables and fuzzy discrimination trees in pattern-directed inference. In: Eighth International Congress of Cybernetics and Systems, New York 1990.

[13] D. Dubois, X. Mo, and H. Prade: Fuzzy discrimination trees. In: Fuzzy Engineering toward Human Friendly Systems. Proc. Internat. Fuzzy Engineering Symposium IFES'91, volume 1, Yokohama 1991, pp. 250–260.

[14] U. M. Fayyad, G. Piatetsky–Shapiro, and P. Smyth: From data mining to knowledge discovery in databases. AI Magazine *17* (1996), 3, 37–54.

[15] X. Hu and N. Cercone: Mining knowledge rules from databases: A rough set approach. In: Proc. 12th Internat. Conference on Data Engineering (S. Y. W. Su, ed.), IEEE Computer Society Press, New Orleans 1996, pp. 96–105.

[16] B. Lent, A. Swami, and J. Widom: Clustering association rules. In: Proc. 13th Internat. Conference on Data Engineering, IEEE Computer Society Press, Birmingham 1997, pp. 220–231.

[17] C. Marsala and B. Bouchon-Meunier: Fuzzy partioning using mathematical morphology in a learning scheme. In: Proc. 5th IEEE Internat. Conference on Fuzzy Systems, volume 2, New Orleans 1996, pp. 1512–1517.

[18] C. Marsala: Apprentissage inductif en présence de données imprécises: construction et utilisation d'arbres de décision flous. PhD Thesis, Université Pierre et Marie Curie, Paris 1998.

[19] C. Marsala: Application of fuzzy rule induction to data mining. In: Proc. 3rd Internat. Conference FQAS'98 – LNAI 1495 (T. Andreasen, H. Christiansen, and H. L. Larsen, eds.), Springers, Roskilde 1998, pp. 260–271.

[20] C. Marsala, M. Ramdani, M. Toullabi, and D. Zakaria: Fuzzy decision trees applied to the recognition of odors. In: Proc. IPMU'98 Conference, volume 1, Editions EDK, Paris 1998, pp. 532–539.
[21] C. Marsala and N. Martini Bigolin: Spatial data mining with fuzzy decision trees. In: Proc. Internat. Conference on Data Mining (N. F. F. Ebecken, ed.), WIT Press, Rio de Janeiro 1998, pp. 235–248.
[22] C. Marsala and B. Bouchon–Meunier: An adaptable system to construct fuzzy decision trees. In: Proc. NAFIPS'99 (North American Fuzzy Information Processing Society), New York 1999, pp. 223–227.
[23] C. Marsala, B. Bouchon–Meunier, and A. Ramer: Hierarchical model for discrimination measures. In: Proc. IFSA'99 World Congress, Taiwan 1999, pp. 339–343.
[24] X. Mo: Compilation de bases de connaissances avec prise en compte de l'imprécision et de l'incertitude. PhD Thesis, Université Paul Sabatier, Toulouse 1990.
[25] H. T. Nguyen, V. Kreinovitch, and D. Tolbert: On robustness of fuzzy logics. In: Proc. FUZZ-IEEE Internat. Conference, volume 1, San Francisco 1993, pp. 543–547.
[26] H. T. Nguyen, V. Kreinovitch, and D. Tolbert: A measure of average sensitivity for fuzzy logics. Internat. J. of Uncertainty, Fuzziness and Knowledge-Based Systems *2* (1994), 4, 361–375.
[27] Pedrycz, Yubazaki, Ohtani, and Hirota: Robustness and sensitivity in fuzzy computational structures. In: Proc. IFSA'91 Conference, Brussels 1991, pp. 197–200.
[28] J.-J. Py: Éléments d'étude de la sensibilité des modèles flous. Université Paris IX, 1997.
[29] J. R. Quinlan: Induction of decision trees. Machine Learning *1* (1986), 1, 86–106.
[30] J. R. Quinlan: Improved use of continuous attributes in C4.5. J. Artificial Intelligence Research *4* (1996), 3, 77–90.
[31] M. Ramdani: Système d'Induction Formelle à Base de Connaissances Imprécises. PhD Thesis, Université P. et M. Curie, Paris 1994.
[32] M. Rifqi: Mesures de comparaison, typicalité et classification d'objets flous : théorie et pratique. PhD Thesis, Université P. et M. Curie, Paris 1996.
[33] S. Shekhar, B. Hamidzadeh, A. Kohli, and M. Coyle: Learning transformation rules for semantic query optimization: A data-driven approach. IEEE Trans. on Knowledge and Data Engineering *5* (1993), 6, 950–964.
[34] L. A. Zadeh: The concept of a linguistic variable and its application to approximate reasoning, part 3. Inform. Sci. *9* (1976), 43–80.
[35] L. A. Zadeh: The role of fuzzy logic in the management of uncertainty in expert systems. Fuzzy Sets and Systems *11* (1983), 199–227.
Reprinted in: Fuzzy Sets and Applications: selected papers by L. A. Zadeh (R. R. Yager, S. Ovchinnikov, R. M. Tong and H. T. Nguyen, eds.), pp. 413–441.

Dr. Christophe Marsala, LiP6, Pôle I.A., Université Pierre et Marie Curie, 8 rue du Capitaine Scott, 75015 Paris. France.
e-mail: Christophe.Marsala@lip6.fr