Inductive Learning and Fuzziness

B. Bouchon-Meunier[1], C. Marsala[1] and M. Ramdani[1,2]

This paper introduces a framework for inductive learning systems in numerical-symbolic data realm. Two complementary methods are proposed to deal with such kind of data pertaining to the used training set. The first method is applicable when some expert knowledge is available, for instance when experts provide fuzzy partitions of the universes of attributes. A second method consists in inferring a fuzzy partition for an attribute when no expert knowledge is available. With these two methods, a fuzzy decision tree is constructed by means of a fuzzy form of entropy, the entropy-star. Finally, it is shown how such fuzzy trees are used to classify unknown data.

INTRODUCTION

Acquisition of knowledge pertaining to a specific domain of expertise is an essential step to achieve an effective reasoning.

A usual way to acquire knowledge is to obtain it from experts. Such knowledge is either simple (e.g. facts, evidence) or complex (e.g. rules, laws, relationships). Inductive learning systems infer complex knowledge from simple knowledge. It is a way of reasoning from many known facts towards a general law.

However, it is often difficult to deal with fuzzy or imprecise simple knowledge. The theory of fuzzy sets introduced by L.A. Zadeh enables coding and treatment of such knowledge. This leads to the integration of fuzzy set theory into inductive learning systems, in order to take into account the fuzziness of simple knowledge.

An expertise domain with examples of cases solved by experts is considered here. Such examples constitute a training set. Each element of this set is represented by a pair [description, class] where a description is a set of pairs [attribute, value].

The purpose is to find general rules which permit classification of any description. It is the induction scheme, which is generally achieved by means of the construction of a decision tree. Each vertex is associated with an attribute. The edges coming out of a vertex are associated with characterizations of the attributes. The simplest case corresponds to symbolic attributes with a finite number of modalities and, in this case, each characterization corresponds to a modality. The choice of the attributes is based on their efficiency with regard to the identification of a class.

Various methods fulfill this task, for instance, the Top Down Induction of Decision Trees (TDIDT) based on methods that use an evaluation function to arrange the attributes in the decision tree. For example, the ID3

1. LAFORIA-IBP, UPMC, Case 169, 4 place Jussieu, 75252 Paris Cedex 05, France.
2. Faculté des Sciences et Techniques de Mohammadia, Département d'Informatique, BP 146, Mohammadia, Maroc.

Scientia Iranica, Vol. 2, No. 4, © Sharif University of Technology, January 1996.

algorithm [1,2] uses a measure of entropy and the CART algorithm [3] is based on Gini's test of impurity. These kinds of methods are connected to the theory of questionnaires [4] and use various tools from information theory [5]. Other methods are based on the theory of data analysis, for example C-means methods or ascending methods.

However, all these methods do not very well fit the numerical nature of the data being dealt with. Furthermore, they do not take into account the imprecision and the fuzziness of the data pertaining to the training set. Sometimes, numerical data are considered as symbolic [6]. These kinds of methods build large trees and it is difficult to rely on induction from a single numerical datum.

Most of the above-mentioned systems discretize the universe of continuous attributes and construct the characterizations as crisp intervals. However, in such a case, the continuous aspect of the attribute is completely ignored (for example, problems occurring near the boundaries). To take into account this continuous aspect, other systems incorporate fuzzy sets in their methods.

Fuzzy techniques are also interesting in case of numerical-symbolic attributes. The training values for such attributes are numerical but experts on the domain, in their current language, use fuzzy symbolic values. The challenge is then to incorporate such data within the inductive systems.

There exist two kinds of systems that integrate fuzzy techniques. The first integrates fuzzy techniques during the learning phase, e.g. the system SAFI [7,8] or Janikow's system [9]. The other system uses such techniques during the classification phase, e.g. Catlett's and Jang's systems [10,11].

In this paper, an algorithm is presented that improves traditional inductive algorithms. In the first section, an approach is proposed to solve the problems of numerical-symbolic data during the construction of decision trees. In the second section, the algorithm which builds the tree is presented. In the third section the use of decision trees to classify new data is briefly discussed. And finally, interesting new developments that could be added are suggested.

A WAY OF HANDLING NUMERICAL-SYMBOLIC DATA

A list of attributes $A_1, \ldots, A_N$ and classes $c_1, \ldots, c_K$ is considered which can be regarded as modalities of a decision attribute C. A training set contains examples which are associated with both values of the attributes and a class. The problem is to find a way of determining the class, given the values of the attributes. From the training set, an order of the attributes is determined leading to the determination of a class, enabling us to associate a class with any new example described only by means of the values of $A_1, \ldots, A_N$. In the case where the data are not homogeneous, numeric or symbolic values of the attributes can be used depending on the case.

There are many difficulties when dealing with numerical-symbolic training data. New ways of handling such data are proposed here. The first way is applicable when expert knowledge regarding the attributes is available. An expert (or many experts) gives some linguistic characterizations for the attributes. For instance, he provides a symbolic partition for a numerical attribute. This kind of expert knowledge is often difficult to obtain or sometimes, it does not exist. The second proposed way enables us to infer a fuzzy partition if no expert knowledge is available.

CASE WHERE EXPERT KNOWLEDGE IS AVAILABLE

Let E be a training set with numerical-symbolic attributes. For example, to study the high jump specialty in the domain of athletics, determination of the relation of pertinence between the height that somebody is able to jump and his age, size and weight is desirable. The attributes of a person are his age, size and weight; the class is the height that he is able to jump (Table 1).

Table 1. Typical training set.

Example	Age	Size (m)	Weight (kg)	Height (m)
E1	20	1.85	80	1.70
E2	25	1.60	80	1.35
E3	35	1.70	60	1.35
E4	40	1.75	75	1.20
E5	29	1.65	90	1.25

Table 2. Symbolic values.

He is young	He is small	His weight is heavy	He jumps high
He is mature	He is tall	His weight is light	He jumps low

Table 3. Training set with degree of satisfiability for each attribute.

Example	Age		Size (m)		Weight (kg)		Height (m)	
	Young	Mature	Small	Tall	Light	Heavy	Low	High
E1	1	0	0	1	0.25	0.75	0	1
E2	1	0	1	0	0.25	0.75	0.75	0.25
E3	0.5	0.5	1	0	1	0	0.75	0.25
E4	0	1	0.5	0.5	0.75	0.25	1	0
E5	1	0	1	0	0	1	1	0

An expert is supposed to provide a list L_i of symbolic modalities (Table 2) for each attribute A_i, for instance {young, mature} for the attribute Age. In order to take into account these symbolic values, an interface is used that translates numerical values into symbolic ones. This is done with the help of fuzzy set theory. Symbolic values are represented by fuzzy sets of the universe corresponding to the numerical attribute (Figure 1).

Thus, symbolic values of L_i will replace the specific values of attribute A_i given in Table 1, with the adjunction of a degree of satisfiability (Table 3).

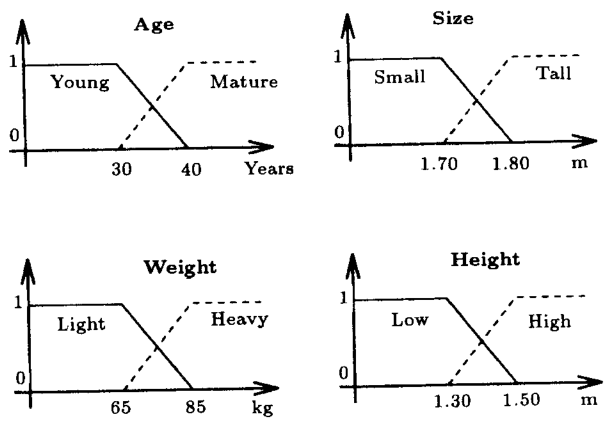

Figure 1. Typical fuzzy partitions.

Degree of Satisfiability

To determine degrees of satisfiability, let us consider the value ω of attribute A_i for a given example of the training set. Coefficients determining to which extent each symbolic value v of L_i can replace ω, or to which extent v is satisfiable for ω are being sought. The so-called degree of satisfiability is used, defined as follows.

Let V and W be two fuzzy sets in a universe X. fw and fv denote their membership functions. The degree of satisfiability [12] is derivated from a fuzzy relation of inclusion

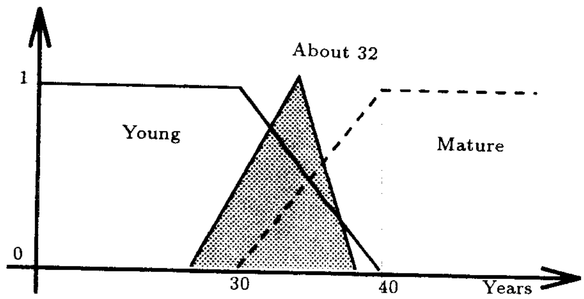

Figure 2. Example for the satisfiability degree.

introduced by Sanchez [13] and defined as:

$$\mathrm{Deg}(W \subset V) = \frac{\int_X f_{V\cap W} dx}{\int_X f_W dx} \text{ if } W \neq \varnothing$$

$$= 0 \text{ if } W = \varnothing$$

More particularly,

$$\mathrm{Deg}(W \subset V) = f_V(x_0) \text{ if } W = \{x_0\} ,$$

and it can be noted that $\int_X f_{V\cap W} dx$ (resp. $\int_X f_W dx$) is the area between $f_{V\cap W}$ (resp. f_W) and the horizontal axis.

For example, in the case of the attribute Age, the fuzzy value about 32 can be replaced by the modality Young with a degree of satisfiability computed in this way (see Figure 2).

Fuzzy Entropy as a Selection Criterion

For the sake of simplicity, it is supposed that all the values of a numerical attribute A are precise in the training set E. Let $X = \{x_1, x_2, \ldots, x_n\}$ be this set of values. Let $v_1, \ldots, v_m$ be the modalities that define a fuzzy partition on the universe of the values of A.

The fuzzy probability [14] of modality v_i is defined as:

$$P^*(v_i) = \sum_{1\leq l\leq n} f_{v_i}(x_l) P(x_l) ,$$

where the probability $P(x_l)$ is weighted by the frequency of x_l in E and f_{v_i} is the membership function of the fuzzy set representing v_i. In practice, if n_{x_l} is the number of examples in E with the value x_l for $A, P(x_l)$ is approximated with $P(x_l) = \frac{n_{x_l}}{|E|}$ ($|E|$ is denoted to be the total number of examples in E).

Let $Y = \{y_1, y_2, \ldots, y_q\}$ be the set of numerical values of the decision attribute C in E. The fuzzy conditional probability of modality c_j of C, given v_i, is defined as:

$$P^*(c_j|v_i) = \frac{P^*(c_j, v_i)}{P^*(v_i)} ,$$

with:

$$P^*(c_j, v_i) = \sum_{1\leq l\leq q} \sum_{1\leq k\leq n} \min(f_{c_j}(y_l), f_{v_i}(x_k)) P(y_l, x_k) .$$

In practice, $P(y_l, x_k)$ is approximated with $P(y_l, x_k) = \frac{n_{y_e \wedge x_k}}{|E|}$, where $n_{y_e \wedge x_k}$ is the number of examples in E with both the values x_k for A and y_l for C.

Entropy-star [8] is called the fuzzy entropy of the decision C related to the attribute A. It is defined as:

$$E_A^* = -\sum_i P^*(v_i) \sum_j P^*(c_j|v_i) \log P^*(c_j|v_i) .$$

This entropy measures the uncertainty of the fuzzy decision when the modalities of the numerical-symbolic attribute A are known. This entropy is the criterion chosen to order the attributes during the construction of the decision tree. Moreover, it generalizes the classical Shannon entropy [15].

Fuzzy Entropy-Based Adjustment of the Membership Functions

The selection of the parameters for the membership functions is crucial. A good definition by an expert is very important for the whole system. When the expert is not sure of his definition, these parameters have to be adjusted. Obviously, it seems convenient to adjust them with the values pertaining to the training set. The entropy-star is minimized in order to realize this adjustment.

However, usually, the best partition minimizing the entropy-star is a crisp partition, which is not of interest to us in this case. To avoid this kind of limitation, the notion of minimal spread degree [8] is introduced. This measure is the lower distance allowed between

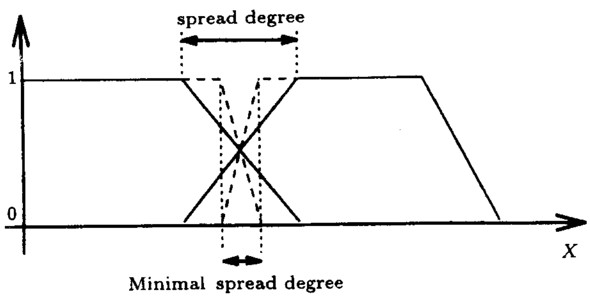

Figure 3. Minimal spread degree.

the kernels of two neighboring membership functions in the fuzzy partition of the universe (Figure 3).

Another way is the adjustment of the membership functions by means of genetic algorithms [16].

CASE WHERE NO EXPERT KNOWLEDGE IS AVAILABLE

In some particular cases, there is no available knowledge from any expert on the domain under study. In other cases, the fuzzy partition of the universe X_j of an attribute A_j has not been known.

Some existing methods adapt the ID3 algorithm to take into account the problem of such training sets of data. The method developed by Catlett [10] is one of those. The CART algorithm [3] is a particular example of how to handle such data. One possible way of improvement lies in the integration of fuzzy notions in order to smooth the boundaries found during the discretization stage when classifying a new example [11,17,18]. However, after some experiments [17,19], it is realized that this kind of fuzzification should be enhanced by another type of discretization. In fact, with such a fuzzification, the spread degrees between the modalities of the obtained partition for an attribute is very small. This fuzzy partition is too close to a crisp partition and thus, the gain in good classification is relatively poor.

The use of the notion of fuzzy entropy in this kind of problem is proposed. To achieve this purpose, a fuzzy partition on the universe A_j must exist. Techniques derived from mathematical morphology theory [20,21] are introduced in order to find such fuzzy partitions. These techniques are formalized using the formal language theory [22,23] in order to respect the particular structures of our data. Such kind of work, relating the formal language theory and pattern recognition techniques can be found in [24]. Our algorithm allows us to find a fuzzy partition of X_j with respect to the distribution of the classes $c_1, \ldots, c_K$ on X_j in E. This will make it possible to use the fuzzy entropy to sort the attributes. In this paper, concentration is focused on the case of crisp classes (non-fuzzy classes). In the presence of fuzzy classes, this method must be adapted, e.g. by using fuzzy mathematical morphology [25].

In the following section, our method to transform the universe of values of an attribute into fuzzy partitions is presented. For more details refer to [19] where all the algorithms are described.

The Training Set as a Word

In order to use our fuzzy measure of entropy, the entropy-star, a fuzzy partition on the universe of the training data must be available. To induce such fuzzy partition, the use of some operators of the mathematical morphology theory is proposed: erosion, dilatation, closure, opening and filter [20]. In our system, these operators are formalized with the use of the formal language theory. Basically, the training set is considered as a word on the alphabet of the classes, each class is viewed as a letter of this alphabet.

For instance, let E be the set $\{(5,+), (7,+), (8,-), (13,+), (14,+), (17,-), (20,+), (21,-), (22,+), (23,+), (25,-), (29,-), (30,-), (35,-), (36,+), (38,-), (40,-)\}$. The first element of each pair is the value of an attribute A_j defined on the universe $X_j \subseteq \mathbb{N}$. The second element is the class (here, there are two classes $+$ and $-$) corresponding to this value. This set of classes defines the alphabet $\{+,-\}$ on which the word $\{++-++-+-++----+--\}$ which is associated with it is deduced.

The two basic operators (erosion and dilatation) are represented as rewriting systems upon this alphabet, with a particular letter as a structuring element (for instance the letter + in the previous example). The erosion eliminates the very short sequences of letters in a word, the dilatation enables the system to merge two sequences of the same letter separated by a short sequence of different letters. The operators closure and opening are combinations of erosion and dilatation. A filter is a set of n closures followed by n dilatations ($n \in \mathbb{N}$).

With these techniques, it is possible to extract a significative sequence from the data (Figure 4). Such a sequence is a set of successive letters in the word induced by the training set. Each of these sequences is related to an interval of X_j. The lower boundary of this interval is the value of the attribute corresponding to the first letter of the sequence and the higher boundary is the value corresponding to the last one. For instance, with the sequences given in Figure 4, the following intervals are obtained: (5,14), (17,23) and (25,40).

In order to generalize a fuzzy partition from the data in the training set E to the whole universe X_j of the attribute A_j, the extreme interval is expanded to the limits of the universe.

In the example, (5,14) becomes (0,14) and (25,40) becomes $(25, +\infty)$.

Within such interval, most examples belong to the same class c_i (for instance, in the example $c_i = +$ or $c_i = -$). It is said that the class of this sequence is c_i. Some sequences, related to an interval where the classes of the data are highly mixed, are called uncertain.

In the example, the first interval defines a sequence +, the second one defines an uncertain sequence and the last one defines a sequence $-$ (Figure 4).

When a filter is applied upon a training set

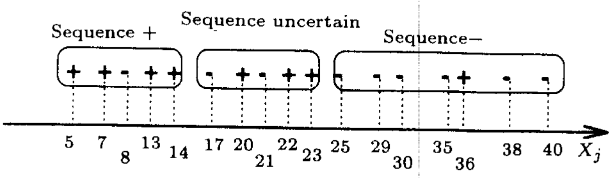

Figure 4. Sequences upon a training set.

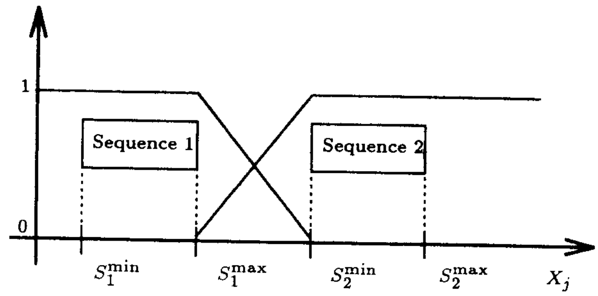

Figure 5. Induced fuzzy partition.

considered as a word, it is possible to eradicate uncertain sequences of classes in this word. The size of the sequences desired determines the number of applied filters. Therefore, a word with large sequences is obtained. Let r be the number of fuzzy modalities required for the attribute. The r largest sequences containing one class are selected, for instance $[\mathcal{S}_1^{\min}, \mathcal{S}_1^{\max}]$ and $[\mathcal{S}_2^{\min}, \mathcal{S}_2^{\max}]$ when $r = 2$. In the case where such r sequences cannot be found, either the number of applied filters can be reduced, or less sequences be selected. A fuzzy partition is inferred upon the universe of data with these r sequences: the considered intervals define the kernel of each fuzzy modality (Figure 5). For instance, from the previous example, the partition given by $\mathcal{S}_1^{\min} = 0$, $\mathcal{S}_1^{\max} = 14$, $\mathcal{S}_2^{\min} = 25$ and $\mathcal{S}_2^{\max} = +\infty$ is inferred.

Once the fuzzy partition is obtained, it is used as if it was given by an expert and the algorithms given in the previous section can be applied.

CONSTRUCTION OF DECISION TREES

The elements for the use of an extended version of the ID3 algorithm devoted to the case of numerical-symbolic data are now available. A decision tree is built with the entropy-star measure as a discrimination criterion. This criterion allows us to find the attribute that will split the current training set. The chosen attribute is the one that minimizes the entropy-star. The current training set is E itself at the first step of the construction of a decision tree. After this first step, E is splitted into several subsets corresponding to the characteristics of the attribute used at the root. Each of these subsets becomes the current training set for the following step.

The modalities of any chosen attribute split the current training set into various subsets. Each modality generates a new current training subset and labels an edge coming out of the current node of the tree. Each example with a non-null degree of satisfiability regarding a modality is conveyed through the edge associated with this modality. Its degree of satisfiability is also conveyed with it.

The decision tree is developed in an iterative way, from the root to the leaves. This development is continued until a particular criterion is fulfilled. For instance, the construction of the decision tree is stopped when:

- The size of the current training set is lower than a fixed threshold,

  $$|\text{training set}| < \text{threshold}_{\text{size}} \,.$$

- The value of the entropy-star is lower than a fixed threshold,

  $$E_C^* < \text{threshold}_{\text{entropy}} \,.$$

- No attributes can further split the training set.

This algorithm differs, depending on existence or non-existence of some additional domain knowledge. If some expert knowledge is available, the numerical-symbolic attribute is fuzzified and then the decision tree is built. If no expert knowledge is available, the attribute at each step of the algorithm is fuzzified. Our algorithm based on the rewriting systems given previously is used at each step of the building of the decision tree and before the computation of the entropy-star of an attribute.

For each class associated with a leaf, the decision is given with the fuzzy conditional probability computed during the development of the tree. For example, in the athletics domain, the decision tree represented in Figure 6 is built.

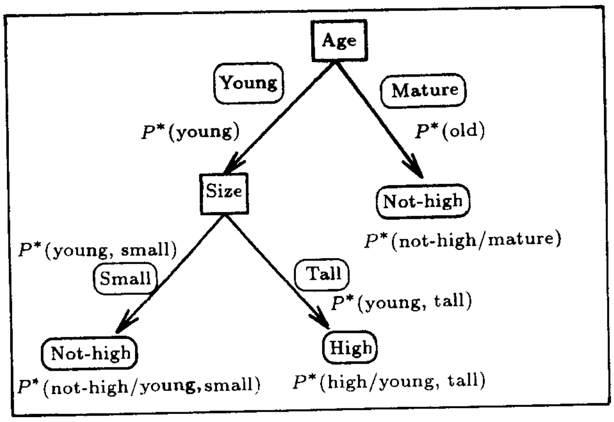

Figure 6. A fuzzy decision tree.

FUZZY CLASSIFICATION WITH DEGREES OF SATISFIABILITY

In this section, the procedure of classifying a new example by means of a fuzzy decision tree is presented.

When a new example should be classified, various kinds of values for the attributes have to be considered. Given an attribute, the example can have either a numerical value or a symbolic modality. However, numerical values do not appear in the tree and the symbolic modality can differ from the modalities used in the training step. This new modality may be given by other experts or be a training modality altered by linguistic modifiers. Therefore, the degree of satisfiability is used to match this new modality with those occurring in the tree.

A decision tree can be regarded as a rule-based system, each rule having the form if <premises> then <conclusion>. Each path of the tree provides such a rule r:

$$\text{if } < A_{l_1} = v_{l_1} > \text{ and } < A_{l_2} = v_{l_2} >$$

$$\text{and} \ldots \text{and } < A_{l_p} = v_{l_p} >$$

$$\text{then } < C = c_k > .$$

$A_{l_1}, \ldots, A_{l_p}$ are the attributes associated with the vertices of the path, chosen from the list $\{A_1, \ldots, A_N\}$; $v_{l_1}, \ldots, v_{l_p}$ are the respective characterizations of $A_{l_1}, \ldots, A_{l_p}$ associated with the edges of the path.

With respect to attributes $A_1, \ldots, A_N$, the example e, being classified, is characterized by:

$$< A_1 = \omega_1 > \text{and} < A_2 = \omega_2 >$$

$$\text{and} \ldots \text{and} < A_N = \omega_N > .$$

It is desirable to know the class c_e of e.

Each characterization $< A_{l_i} = \omega_{l_i} >$ of the example is associated with the premise $< A_{l_i} = v_{l_i} >$ of rule r with the degree of satisfiability $\text{Deg}(\omega_{l_i} \subset v_{l_i})$. This degree values how much $< A_{l_i} = \omega_{l_i} >$ matches the premise $< A_{l_i} = v_{l_i} >$. The degrees from all the premises of a rule are aggregated with the product to obtain a global degree according to this rule:

$$\Pi_{i=1\ldots p} \text{Deg}(\omega_{l_i} \subset v_{l_i}).$$

This definition of a global degree is consistant with the definition of degree of satisfiability chosen if the intersection and the Cartesian product of fuzzy sets are defined by means of the product t-norm. Namely, in this case:

$$\Pi_{i=1\ldots p} \text{Deg}(\omega_{l_i} \subset v_{l_i}) = \text{Deg}((\omega_{l_1}, \ldots, \omega_{l_p}) \subset (v_{l_1}, \ldots, v_{l_p})).$$

The t-norm product is the only t-norm which satisfies this egality for the chosen degree of satisfiability. For another degree of satisfiability, corresponding t-norms have to be chosen in order to satisfy the same kind of constraint.

The example is associated with class c_k, according to the rule r, with a final degree of satisfiability. This final degree is the global degree weighted by the fuzzy conditional probability of class c_k associated with rule r:

$$F\,\text{Deg}_r(c_k) = \Pi_{i=1\ldots p} \text{Deg}(\omega_{l_i} \subset v_{l_i}) \cdot P^*(c_k|(v_{l_1}, v_{l_2}, \ldots, v_{l_p})).$$

This degree values the fact that "c_k could be the class of the example according to the rule r".

A component of an example can have non-null satisfiability degrees with regard to several training modalities for the same attribute. Therefore, an example can result in several rules and provide several non-null final degrees to different classes. All these degrees are aggregated by means of a triangular conorm $\perp$ (e.g. maximum) to obtain one degree for each class. If n_p is the number of the rules in the decision tree, it is concluded that:

$$F\,\text{Deg}(c_k) = \perp_{r=1\ldots n_p} F\,\text{Deg}_r(c_k).$$

This represents the degree for the example to be associated with class c_k, according to the decision tree. Therefore, the example e can be associated with the class c_e corresponding to the higher degree:

$$F\,\text{Deg}(c_e) = \max_{k=1\ldots K} F\,\text{Deg}(c_k).$$

Let us consider again the athletics domain example. The age (about 32), the size (1.90 m) and the weight (80 kg) of a man are known and it is desirable to estimate the height he is able to jump.

With the help of the tree (Figure 6) and after computation of the corresponding degrees of satisfiability, Deg (about 32 $\subset$ young) = 0.95 and Deg (about 32 $\subset$ mature) = 0.33, the two following rules are found for this example:

RULE 1: if $<$ age = young $>$

and $<$ size = tall $>$

then $<$ height = high $>$,

RULE 2: if $<$ size = mature $>$

then $<$ height = not-high $>$.

This corresponds to the following global degree for the first rule: $0.95 * 1.00$. Then, the class could be $<$ height = high $>$ with the final degree $F\,\text{Deg}(\text{high}) = 0.95 * P^*(\text{high}|\text{ young, tall})$ according to the first rule.

Moreover, 0.33 is the global degree for the second rule and then the class could also be $<$ height = not-high $>$ with the final degree:

$$F\,\text{Deg}(\text{not-high}) = 0.33 * P^*(\text{not-high}|\text{mature})$$

according to the second rule.

Finally, the chosen class of the example is the one giving the higher degree:

$$\max(F\,\text{Deg}\,(\text{high}), F\,\text{Deg}\,(\text{not-high})) .$$

CONCLUSION

In this paper, two methods are presented in order to take into account numerical-symbolic data and problems derivated from these kinds of values in induction learning systems. The first method works with additive knowledge from experts that support the usual training set. This knowledge is given as fuzzy partitions of the universes of the numerical attributes. Moreover, it makes it possible to deal with fuzzy decisions. The second method is usable when no particular additive knowledge is available from experts. It is necessary to induce fuzzy partitions of the universes of the attributes to use the given algorithm. These methods are very close. They are based upon the same algorithm and use a new measure, the entropy-star, that generalizes the classical Shannon's entropy. They handle particular knowledge in data that appears in a numerical-symbolic form.

These methods have been tested on data. Some results are given in [8] for the first kind and in [17,18] for the second one.

This kind of improvement will be enhanced with another type of expert knowledge. In future developments it is desirable to add more expert knowledge in such systems: knowledge to help in building such trees and knowledge to use decision trees to classify objects.

REFERENCES

1. Quinlan, J.R. "Induction of decision trees", *Machine Learning,* **1**(1), pp 86–106 (1986).

2. Quinlan, J.R. *C4.5: Programs for Machine Learning,* Morgan Kaufmann, San Mateo, CA, USA (1993).

3. Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. *Classification and Regression Trees,* Chapman and Hall, New York, USA (1984).

4. Picard, C.F. *Graphes et Questionnaires,* Gauthier-Villars (1972).

5. Fu, K.S. *Sequential Methods in Pattern Recognition and Machine Learning,* **52**, *Mathematics in Science and Engineering,* Academic Press, New York, USA (1968).

6. Maher, P.E. and Saint-Clair, D. "Uncertain reasoning in an ID3 machine learning framework", *2nd IEEE Int. Conf. on Fuzzy Systems,* pp 7–12 (1993).

7. Ramdani, M. "Une approche floue pour traiter les valeurs numériques en apprentissage", In *Journées Francophones d'Apprentissage et d'Explication des Connaissances* (1992).

8. Ramdani, M. "Système d'induction formelle à base de connaissances imprécises", Thèse de l'Université P. et M. Curie, Rapport 94/1, LAFORIA-IBP (1994).

9. Janikow, C.Z. "Fuzzy decision trees: FIDMV", *Proc. of the Joint Conf. on Information Science*, pp 232–235 (1994).

10. Catlett, J. "On changing continuous attributes into ordered discrete attributes", *Lecture notes in Artificial Intelligence*, 482, pp 164–178 (1991).

11. Jang, J-S. R. "Structure determination in fuzzy modeling: a fuzzy CART approach", *Proc. of the 3rd IEEE Int. Conf. on Fuzzy Systems,* pp 480–485 (1994).

12. Bouchon-Meunier, B., Rifqi, M. and Bothorel, S. "Towards general measures of comparison of objects", *Fuzzy Sets and Systems,* to appear.

13. Sanchez, E. "Inverses of fuzzy relations. Application to possibility distributions and medical diagnosis", *Fuzzy Sets and Systems*, **2**(1), pp75–86 (1979).

14. Zadeh, L.A. "Probability measure of fuzzy events", *J. of Math. Anal. App.*, **23**, pp 421–427 (1968).

15. Shannon, C.E. and Weaver, W. *The Mathematical Theory of Communication,* Uni-

versity of Illinois Press, Urbana, USA (1949).

16. Janikow, C.Z. "A genetic algorithm for optimizing fuzzy decision trees" *Proc. of the 6th Int. Conf. on Genetic Algorithms*, L.J. Eshelman, Ed., Morgan Kaufman Publishers, Inc., San Francisco, CA, USA, pp 421–428 (1995).

17. Marsala, C. "Arbres de décision et sous-ensembles flous", Rapport 94/21, LAFORIA-IBP (1994).

18. Bouchon-Meunier, B., Marsala, C. and Ramdani, M. "Arbres de décision et théorie des sous-ensembles flous", *Actes des 5èmes Journées du PRC-GDR d'Intelligence Artificielle,* pp 50–53 (1995).

19. Marsala, C. "Fuzzy partition inference over a set of numerical values" Report 95/22, LAFORIA-IBP (1995).

20. Serra, J. *Image Analysis and Mathematical Morphology*, Academic Press, New York, USA (1982).

21. Coster, M. and Chermant, J.L. *Précis d'Analyse d'Images,* Presses du CNRS (1989).

22. Ginsburg, S. *The Mathematical Theory of Context Free Languages,* McGraw-Hill, New York, USA (1966).

23. Autebert, J.M. *Langages Algébriques,* Masson, Paris, France (1987).

24. Fu, K.S. *Syntactic Methods in Pattern Recognition*, **112**, *Mathematics in Science and Engineering*, Academic Press, New York, USA (1974).

25. Bloch, I. and Maître, H. "Mathematical morphology on fuzzy sets", *Mathematical Morphology and its Application to Signal Processing*, Barcelona, Spain (1993).