Gradual Fuzzy Decision Trees to Help Medical Diagnosis

Christophe Marsala
LIP6, Université Pierre et Marie Curie, Paris 6,
4 place Jussieu, 75005 Paris, France
Email: Christophe.Marsala@lip6.fr

***Abstract*—In this paper, we consider the problem of the construction of fuzzy decision trees when there exists a graduality between the values of attributes and values of the class. We propose a new measure, extended from the measure of classification ambiguity, that takes into account both discrimination power and graduality with regards to the class. To highlight the importance of that kinds of measures, Medical applications is presented in which often the values of the class are symbolic and ordered and in which the discovery of gradual links between descriptive attributes and the class are seek for.**

I. Introduction

Fuzzy decision trees (FDT) are a very popular machine learning tools. They propose a summarized view of a set of data. Moreover, there exist several inductive approaches to construct them from a training data set. Basic approaches are based on a *Top Down Induction of Decision Tree* (TDIDT) method. A tree is built from its root to its leaves, by successive partitioning of the training set into subsets. An attribute is selected thanks to a *measure of discrimination* H (in classical decision tree, the Shannon entropy is generally use [1], [2]) that ranks the attributes according to their discriminating power with regard to the class. The attribute with the highest discriminating power is selected to split the training set. Methods to construct decision trees differ mainly in their choice of H [3], [4].

In the fuzzy settings, these measures of discrimination have been studied in previous work and a hierarchical model of validation of such a measure has been presented [3], [5], [4]. This model points out the properties required for a given measure to be considered as a measure of discrimination in order to be used in the process of construction of a (fuzzy) decision tree. Moreover, a comparative study of these measures has been conducted [6] that leads to a better understanding of their main properties.

At each step of the construction of a FDT, the measure of discrimination H is used to value the power of discrimination of each attribute with regard to the class. Thus, it will produced a ranking of all the attributes according to this value, and the *winner* attribute will be the one that is ranked first (ie. the one that has the lowest value). As a consequence the whole ranking is not interesting in this process (only the first one is selected).

Thus, FDT have been extensively used in the past years as a powerful knowledge extraction tool, and nowadays it is still an active domain of researches and applications [7], [8], [9], [10], [11], [12]. Very recent works have also shown that FDT can be used too in ranking applications where it is more useful to associate test example with a degree of classification rather than a crisp class [13].

There exist a lot of real-world applications problems, and in particular in Medical application, where the values of the class are symbolic and ordered. In that kind of problems, it appears that finding attributes that are *gradually* linked with the class could be more valuable in order to explain the decision process done by means of that tree. For instance, *the older the patient, the most vulnerable to disease*.

However, classical measures of discrimination take into account only the informative properties of attributes with regards to the class and forget to handle the graduality that could link their values.

In this paper, we propose a new measure that extend the measure of ambiguity and take into account both the discriminative power of an attribute, and the graduality that links it to the class.

The paper is composed as follow. In Section , we recall the method to construct fuzzy decision trees and the classical measures that are used to rank attributes and select them to construct the tree. In Section , the problem of measuring both discrimination and graduality is set and a new measure is proposed that handles them. In Section , a presentation of the Medical domain and a related application that benefits the use of the proposed measure is done. Finally, we conclude and present some future work.

II. Construction of Fuzzy Decision Trees

Construction of a fuzzy decision tree is based on a machine learning process. A *training set* $\mathcal{E} = \{e_1, ..., e_N\}$ is provided, composed of examples (or cases, or observations) of a given phenomenon. The examples are associated with a *description* which is a N_A-tuple of attribute-value pairs (A_j, v_{jl}), where each attribute A_j, from a set of attributes $\mathcal{A} = \{A_1, ..., A_{N_A}\}$, can take a (fuzzy, numerical, or symbolic) value v_{jl} in a set $\{v_{j1}, ..., v_{jm_j}\}$ of possible values. A value v_{jl} is associated with a membership function $\mu_{v_{jl}}$. We consider supervised learning, where a set of classes $C = \{c_1, ..., c_K\}$ is also given. Each c_k is supposed to be associated with a membership function μ_{c_k}. In the training set, the description of an example e_i is linked with a particular class c_k from C.

Most algorithms to construct decision trees build a tree from the root to the leaves, by successive partitioning of the training set into subsets. Each partition is done by means of the values of an attribute and leads to the definition of a node of the tree. An attribute is selected by means of a *measure of discrimination* H (in classical decision tree, the Shannon entropy is generally use [2], it comes from Information Theory [14]). Such a measure ranks the attributes from $\mathcal{A}$ according to their discriminating power with regard to the class. The attribute with the highest discriminating power is selected to construct a node in the decision tree.

Various measures of discrimination H do exist and can be used to construct a FDT. The main one is the *entropy of fuzzy events* [15], [16], [17]. It corresponds to an extension of the Shannon entropy by substituting probabilities of fuzzy events [18] to classical probabilities.

The second one deals with methods based on another family of fuzzy measures [19], [20], [21], [22], [23]. In this second family of measures, a very different kind of works is presented in [19]. A method is introduced to construct a FDT by means of a *measure of classification ambiguity* as a measure of discrimination. This measure is defined from both a measure of fuzzy subsethood and a measure of non-specificity.

Another measure has been studied in [1] for the construction of decision trees, the Gini index of diversity. This measure is based on probabilities and can easily be extended to a Fuzzy Gini index of diversity to handle fuzzy values and, thus, enable the construction of FDT. For instance, it has been used by [24].

In previous works, the seminal properties of a measure of discrimination have been studied and a hierarchical model of validation of such a measure has been presented [3]. This model points out the properties required for a given measure to be considered as a measure of discrimination in order to be used to rank attributes in the process of construction of a (fuzzy) decision tree. While the model is based on fuzzy values of attribute, it was only set for crisp values and class in the description of a training example. An extension of this model was proposed to handle also fuzzy values in the description of the example [5]. However, this extension was set for particular and very restrictive hypotheses: the definition of the inclusion of fuzzy sets and the t-norm used for the intersection of fuzzy sets were very narrowed.

A. Measures of Information for Attribute Ranking

We place ourselves in inductive learning where examples from a training set $\mathcal{E} = \{e_1, ..., e_N\}$ are associated with a *description* which is a N_A-tuple of attribute-value pairs (A_j, v_{jl}). Each attribute A_j, from $\mathcal{A} = \{A_1, ..., A_{N_A}\}$, can take a fuzzy, or symbolic value v_{jl} from $\{v_{j1}, ..., v_{jm_j}\}$. A value v_{jl} can be associated with a membership function $\mu_{v_{jl}}$ from $\mathcal{E}$ to $[0, 1]$. Moreover, a set of classes $C = \{c_1, ..., c_K\}$ is also considered. Each c_k is can be associated with a membership function μ_{c_k} from $\mathcal{E}$ to $[0, 1]$. In the training set, each description is linked with a particular class c_k from C to make up an example e.

During the construction of a FDT, attributes are ranked according to their discriminating power related to the class C. The attribute with the better discriminating power is selected to make up a node in the tree. The discriminating power $H(C|A_j)$ of attribute A_j with regard to C is valued by means of a measure of discrimination H. In the following the three main measures of discrimination used to construct fuzzy decision trees are presented: the entropy of fuzzy events H_E, the fuzzy index of Gini H_G, and the measure of ambiguity H_Y.

1) Entropy of fuzzy events: The discriminating power $H_E(C|A_j)$ is valued by means of the entropy of fuzzy events which is an extension of the Shannon entropy to fuzzy events. $H_E(C|A_j)$ can be written [6]:

$$H_E(C|A_j) = \sum_{l=1}^{m_j} p^*(v_l) \cdot G_E(v_l),$$

with

$$G_E(v_l) \quad = \quad -\sum_{k=1}^{K} p^*(c_k|v_l) \log p^*(c_k|v_l). \qquad (1)$$

with $p^*(a|b) = \frac{p^*(a \cap b)}{p^*(a)}$. The conditional probability $p^*(a|b)$ of fuzzy events a and b has been defined by [18] with, for all $x \in X$, $\mu_{a \cap b}(x) = \mu_a(x) \cdot \mu_b(x)$.

2) Fuzzy Index of Gini: An *index of diversity* has been introduced by Gini in statistical inference and by Simpson in biology [25]. Its use in the construction of decision trees has been mentioned in [1]. In recent literature, a use of the fuzzy extension of this index to construct fuzzy decision trees can be found in [24]. The fuzzy index of Gini $H_G(C|A_j)$ can be written as follows:

$$H_G(C|A_j) = \sum_{l=1}^{m_j} p^*(v_l) \cdot G_G(v_l), \qquad (2)$$

with $G_G(v_l) = 1 - \sum_{k=1}^{K} p^*(c_k|v_l)^2$.

3) Measure of ambiguity: The *measure of ambiguity* $H_Y(C|A_j)$ was introduced by [19] to measure the discriminating power of attributes for the construction of a fuzzy decision tree. It is defined as:

$$H_Y(C|A_j) \quad = \quad \sum_{l=1}^{m_j} w(v_l) \cdot G_Y(v_l) \qquad (3)$$

with

$$w(v_l) \quad = \quad \frac{M(v_l)}{\sum_{l=1}^{m_j} M(v_l)}$$

where M is the sigma-count [18]: $M(v_l) = \sum_{x \in X} \mu_{v_l}(x)$.

It has been prove that this measure can be rewritten by means of the probability of fuzzy events [6], and thus, Equation (3) can be rewritten:

$$H_Y(C|A_j) \quad = \quad \sum_{l=1}^{m_j} p^*(v_l) \cdot G_Y(v_l).$$

In [19], it is considered that $G_Y(v_l) = g(\Pi(C|v_l))$ where g is a *non-specificity measure*, and $\Pi(C|v_l) = \{\pi(c_k|v_l),\ k = 1,..,K\}$ is the possibility distribution of c_k related to value v_l of attribute A_j. $\Pi(C|v_l)$ is determined as follows. Each $\pi(c_k|v_l)$ is defined from a *fuzzy subsethood* measure S as:

$$\pi(c_k|v_l) = \frac{S(v_l, c_k)}{\max_{i=1,..,K} S(v_l, c_i)}$$

where the fuzzy subsethood S of value w related to value v is defined as:

$$S(w, v) = \frac{\sum_{x\in X} \mu_{w\cap v}(x)}{\sum_{x\in X} \mu_w(x)}.$$

We denote hereafter $\pi(c_k|v_l)$ simply π_k. It has been prove that this function can be rewritten [6]:

$$G_Y(v_l) = \sum_{i=2}^{K} \frac{p^*(c_{\rho(i)}|v_l)}{p^*(c_+^l|v_l)} \log(1 + \frac{1}{i-1})$$

with c_+^l the majority class, *ie.* $c_+^l = \text{argmax}_{c_i} S(v_l, c_i)$ and ρ is a permutation such that for all $i = 1,..,K$, $\pi_i^* \geq \pi_{i+1}^*$, with $\pi_i^* = \pi_{\rho(i)}$.

B. A comparison of these measures

The presentation of these measures highlights the similarities that exists among them. First of all, it is easy to see that these three measures are all based on the same aggregation of a specific function G that differentiates them. We have:

$$H(C|A_j) = \sum_{l=1}^{m_j} p^*(v_l) \cdot G(v_l).$$

Afterwards, we can assume that for each value v_l, a ranking of the classes is done. We denote c_1^l, c_2^l,... c_K^l, the classes such that for all $i = 1,..,K$, $p^*(c_i^l|v_l) \geq p^*(c_{i+1}^l|v_l)$. This assumption has no effect for the entropy of fuzzy event or for the fuzzy index of Gini, but it simplifies the notation for the measure of ambiguity and it enables us to rewrite the three measures:

- The entropy of fuzzy events:

$$G_E(v_l) = -\sum_{k=1}^{K} p^*(c_k^l|v_l) \log p^*(c_k^l|v_l)$$

- The fuzzy index of Gini:

$$G_G(v_l) = 1 - \sum_{k=1}^{K} p^*(c_k^l|v_l)^2$$

- The measure of ambiguity:

$$G_Y(v_l) = \sum_{i=2}^{K} \frac{p^*(c_k^l|v_l)}{p^*(c_1^l|v_l)} \log(1 + \frac{1}{i-1})$$

Let $x_l = p^*(v_l)$, $l = 1, \ldots, m$, and $y_{lk} = p^*(c_k^l|v_l)$, $k = 1,..,K$. We also consider that the following properties hold:

- for all $l = 1, \ldots, m$, $x_l \in]0, 1]$ and $\sum_{l=1}^{m} x_l = 1$.
- for all $l = 1, \ldots, m$, for all $k = 1,..,K$, $y_{lk} \in [0, 1]$, $\sum_{k=1}^{K} y_{lk} = 1$, and $y_{l1} \geq y_{l2} \geq \ldots \geq y_{lK}$.

In the following, we denote $X = (x_1, \ldots, x_m)$, $Y_l = (y_{l1}, \ldots, y_{lK})$, and $Y = (Y_1, \ldots, Y_m)$. Under these assumptions, we can rewrite the three measures as functions from $[0, 1]^{m\times m\times K}$ to $[0, 1]$ where m is the number of values of the attribute and K is the number of classes. (for the sake of simplicity, we denote m rather than m_j the number of values of attribute A_j):

- the entropy of fuzzy events $H_E(C|A_j)$ can be rewritten:

$$f_E(X, Y) = -\sum_{l=1}^{m} x_l \sum_{k=1}^{K} y_{lk} \log y_{lk},$$

- the fuzzy index of Gini $H_G(C|A_j)$ can be rewritten:

$$f_G(X, Y) = 1 - \sum_{l=1}^{m} x_l \sum_{k=1}^{K} y_{lk}^2,$$

- the measure of ambiguity $H_Y(C|A_j)$ can be rewritten:

$$f_Y(X, Y) = \sum_{l=1}^{m} x_l \sum_{k=2}^{K} \frac{y_{lk}}{y_{l1}} \log(1 + \frac{1}{k-1}).$$

A more deeper study that compares these measures can be found in [4].

III. Graduality and information

As seen in the previous section, there exists various discrimination measures to select attributes during the construction of the fuzzy decision trees. However, if these measures tackle perfectly the information brought out by an attribute with regards to the class, they are always independent of the distribution of the examples.

A. Illustrative examples

For instance, in Table I and Table II, some illustrative examples of the independence of the measure with regards to the graduality of the values are shown. Let A_i, $i = 1, \ldots, 4$ be attributes with 4 ordered values $v_{il}, l = 1, \ldots, 4$ such that $v_{il} \leq v_{ij}$ for all $l \leq j$. Let C be the class to recognize with 4 values $c_k, k = 1, \ldots, 4$ such that $c_k \leq c_{k+1}$ for all $k = 1, \ldots, 3$.

Considering attributes A_1 and A_2 given in Table I, we have $H_E(C|A_1) = H_E(C|A_2) = 0$. However, looking at the distribution of the examples (10 examples from the training set have the value v_{11} for A_1 and the value c_1 for C, and so on...), a graduality could be highlighted for the values of A_1 with regards to C: the greater is A_1, the greater is C.

Considering attributes A_3 and A_4 given in Table II, we have $H_E(C|A_3) = H_E(C|A_4) = 0.588$. Here again, looking at the distribution of the examples a graduality could be highlighted: the greater is A_3, the greater is C.

In that last case, the graduality is more predominant for attribute A_1 than for attribute A_3 due to the loss of discrimination power if that attribute, but this graduality do exist.

TABLE I
NULL ENTROPY AND GRADUALITY (LEFT) OR NO GRADUALITY (RIGHT)

A_1	v_{11}	v_{12}	v_{13}	v_{14}
c_1	10	0	0	0
c_2	0	10	0	0
c_3	0	0	10	0
c_4	0	0	0	10

A_2	v_{21}	v_{22}	v_{23}	v_{24}
c_1	0	0	10	0
c_2	10	0	0	0
c_3	0	0	0	10
c_4	0	10	0	0

TABLE II
SAME ENTROPY AND GRADUALITY (LEFT) OR NO GRADUALITY (RIGHT)

A_3	v_{31}	v_{32}	v_{33}	v_{34}
c_1	7	3	0	0
c_2	2	6	0	2
c_3	1	0	7	2
c_4	0	1	3	6

A_4	v_{41}	v_{42}	v_{43}	v_{44}
c_1	0	7	0	3
c_2	0	2	2	6
c_3	7	1	2	0
c_4	3	0	6	1

Such consideration can be done with the two other measures of discrimination, the Gini index and the measure of ambiguity.

Our aim is thus to introduce a measure to enable the selection of attribute gradually linked to the class. For instance, here, we would prefer attribute A_1 to attribute A_2, and we would prefer attribute A_3 to attribute A_4, as attributes that should appear in the node of the built fuzzy decision tree.

It is important to note here that the choice of the attribute can not be only solved by means of a order comparison measure as, for instance, the Spearman's rank or the Kendall's rank correlation coefficients. The main reasons are on one hand, the fact these correlation coefficients measures only the distribution of the values and are not able to take into account the information power associated with the values. On the other hand, the complexity of their valuation could be seen as too expensive in a data mining process (the complexity of their use in the process of construction of a fuzzy decision tree could not often be afforded).

Moreover, the use of two different measures (one for the discriminative power, and the other one for the graduality detection) could generally be not convenient because an aggregation problem could arise here: how to aggregate the values of such two measures in order to rank the attributes and enable us to select one?

B. *A new measure to value graduality and information*

An alternative solution we propose in this paper is to construct a single measure that could combine discrimination and graduality measurements.

As explained in the previous section, the discrimination power of attribute A_i is measured thanks to the conditional probability $p^*(c_k|v_{il})$ of each of its value v_{il}. The graduality of the values of A_i with regards to C could be also valued by taking into account the relation between all the values $p^*(c|v)$ of the class and of the attribute. The main linked attribute A_i and class C will have a distribution of their values such that $p^*(c_k|v_{ik})$ will be optimal and $p^*(c_k|v_{il})$, for $k \neq l$, will be non optimal.

In order to value that, we propose to use an adapted measure of ambiguity. In spite of taking into account the $p^*(c_k|v_{ik})$ values to value $H_G(C|A_i)$ after ordering them, that will produce the information of their ranking with regards to both the values of C and pf A_i, we propose to keep the order of that conditional values.

Moreover, as we need to penalize a non gradual order of the distribution of the values, we propose to introduce the value of $k - j$ in the valuation of the discrimination power.

Let $H_O(C|A_i)$ be the measure:

$$H_O(C|A_i) = \sum_{l=1}^{m_i} p^*(v_l) \cdot G_O(v_l),$$

with

$$G_O(v_l) \quad = \quad \sum_{k=1}^{K} \frac{p^*(c_k|v_l)}{p^*(c_+|v_l)} \log(1 + \frac{1}{\max(m_i, K) - |k-l|}).$$

with $p^*(c_+|v_l) = \max_{k=1,\ldots,K} p^*(c_k|v_l)$.

By means of $H_O(C|A_i)$, the measure of the discrimination power of A_i with regards to C will take into account a gradual order between the values of A_i and those of C.

For instance, with the examples given in Table I and Table II, we have $H_O(C|A_1) = 0.097$, $H_O(C|A_2) = 0.15$, $H_O(C|A_3) = 0.173$, and $H_O(C|A_4) = 0.254$. Here we can see that the measure H_O favors not only attribute with a good discrimination power but also attributes with a high gradual relation with the class. For instance, A_1 will be preferred to A_2 clearly, and A_3 will be preferred to A_4.

IV. GRADUALITY FOR MEDICAL DIAGNOSIS

A. *Assessing asthma severity*

We describe here a data mining application for assessing asthma severity. This work has been conducted jointly with the physician Dr. Alain Lurie[1] and more details can be found in [26].

1) Study on observance: The aim of this study has been to identify variables and decision pathways patients use to determine the severity of their asthma (perceived severity of the asthma). Thus, the identified variables were compared to those involved in the assessment of asthma severity according to the National Asthma Education and Prevention Program (NAEPP) Guidelines (objective severity of the asthma).

The database is composed of a set of 113 outpatients (51 men, 62 women), with a percentage of patients with mild intermittent (6.2), mild persistent (15.9), moderate (65.5) and severe (12.4) asthma.

A questionnaire was filled by each patient. It is composed of several information, including the patients' sociodemographic characteristics and their asthma characteristics. For the latter, two parts are to be distinguished, respectively assessed by the patients and by the doctors: the patients had to assess their perceived asthma severity (rated as mild intermittent,

[1]Service de Pneumologie, Hpital Cochin, Assistance Publique-Hpitaux de Paris, 27 rue du Faubourg Saint-Jacques, F75679 Paris Cedex 14, France.

mild persistent, moderate or severe), the response to treatment (perceived treatment efficacy), the quality of their life, valued by means of the asthma-related quality of life questionnaire (AQLQ), and the rating of its medical adherence by the patients.

The doctor part concerned the objective asthma severity (also rated as mild intermittent, mild persistent, moderate or severe) derived from medical criteria and the valuation of the respiratory functions. All these variables were pooled, and considered as potential variables patients might use to determine the perceived severity of their asthma.

In order to analyze the relationships between the features and the perceived severity of the asthma, a fuzzy decision tree was constructed and enables to highlight variables and decision pathways patients use to determine themselves the severity of their asthma.

In this Medical application, as often in such kind of application, the main aim is to characterize relationships between the features and the class with regards to the patients' perception. The values of the class are symbolic and ordered: from mild intermittent to severe asthma. Thus, in this kind of application, it will be valuable to discover gradual relations between attributes and class in order to offer a more understandable set of rules that will help the Physician to consider the diagnosis proposed by the fuzzy decision tree.

2) Experiment: In order to study the proposed measure, a cross validation has been done on this medical data set.

The set of 113 patients has been split 10 times by selecting 75% of patients (86 patients) to set a training set, and the other 25% of patients (27 patients) to set a test set. Each patient is described by means of 94 attributes (age, AQLQ answers, and so on.). The class associated with each patient is the valuation of the severity of the asthma expressed by the physician, and ranges gradually from 1 (mild) to 4 (severe). The main aim here is to study the attributes involved in the decision.

We compare 3 measures to select attributes during the construction of the fuzzy decision tree: entropy of fuzzy events (H_E), Yuan and Shaw's measure (H_Y), and the measure we proposed (H_O). The results presented here are only provided to compare the H_O to other fuzzy measures used to construct fuzzy decision trees.

TABLE III
COMPARISON OF MEASURES

Measure	Accuracy (avg.)	Max. depth (avg.)	Avg. Depth (avg.)
H_E	0.833 $\pm$0.07	5.9 $\pm$1.1	3.9 $\pm$0.27
H_Y	0.833 $\pm$0.07	5 $\pm$0	3.84 $\pm$0.14
H_O	0.837 $\pm$0.10	5 $\pm$0.66	3.81 $\pm$0.23

Results are presented in Table III. In this table, we present the average accuracy values obtained from the 10 fuzzy decision trees (FDT) constructed as explained on the 10 random training samples, and the corresponding standard deviation. Moreover, we present the average depth of the constructed FDT (in number of nodes on pathes from the root to leaves) that is a measure of the complexity of the trees (the smaller, the better).

It can be seen that the accuracy of the trees is very high. The FDT constructed by means of the H_O measure have a good accuracy and outperform slightly the FDT constructed by other measures but with a higher standard deviation that highlight a lower robustness.

Concerning the size of the obtained FDT, if the FDT constructed by means of the H_O measure are, in average, smaller that those constructed by means of the other measures. As a consequence, they have a better complexity in depth (e.g. reduced number of nodes) than the FDT constructed by means of other measures.

Thus, to summarize, the H_O measure brings out a good performance both in accuracy and in complexity, with regards to those constructed by means of the two other well-known measures.

B. Cardio-vascular diseases

In such applications, the main aim is to build predictions that should help medical scientists to detect and prevent cardio-vascular diseases for hypertensive patients.

For instance, the INDANA (INdividual Data ANalysis of Antihypertensive intervention) database [27] has been used by several teams for such application. Thus, in [28], this database was used to study cardio-vascular risks for patients.

For each patient, a set of classical features was combined with a set of medical measurements on several years and lead to the conclusion for this patient (death or not). The particular database used for this experiment was composed of 20 features (ident, sex, age, height, weight, medical measures, and the class either death or not) for 2230 patients. These features were measured during a long period of time and the class associated to each patient indicated if the patient died or not during the period. In this set, 107 patient died of a cardio-vascular disease during the period, and 2123 were alive at the end of the period.

As in the previous Medical application, here, the detection and the highlight of gradual relations between attributes and the class will be of great benefit. For instance, inferring rules such that *the more cigarettes smoked, the higher the cardio-vascular diseases risk* could bring a lot of information in order to understand better the relation between the attributes and the class.

V. CONCLUSION

In this paper, we have proposed a new measure, extended from the measure of classification ambiguity, to take into account both discrimination power and graduality with regards to the class.

In future work, the proposed measure will be used to construct fuzzy decision trees for Medical applications in order to highlight and to discover gradual links between descriptive attributes and the diagnosis class. Moreover, a deeper study of the properties of that measure will be conducted by means of the hierarchical model of measures of discrimination [6], [4], in order to ensure their good properties to be used in a fuzzy decision tree construction process.

REFERENCES

[1] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification And Regression Trees*. New York: Chapman and Hall, 1984.

[2] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, Ca: Morgan Kaufmann, 1993.

[3] C. Marsala, B. Bouchon-Meunier, and A. Ramer, "Hierarchical model for discrimination measures," in *Proc. of the IFSA'99 World Congress*, Taiwan, 1999, pp. 339–343.

[4] C. Marsala and B. Bouchon-Meunier, "Quality of measures for attribute selection in fuzzy decision trees," in *Proceedings of the International Conference on Fuzzy Systems (WCCI-2010)*, Barcelona, Spain, July 2010.

[5] T.-H. Dang and C. Marsala, "Extension of hierarchical model for fuzzy measures of discrimination," in *Proc. of the 11th IPMU'06 Conf.*, Paris, France, July 2006, pp. 1284–1291.

[6] C. Marsala and B. Bouchon-Meunier, "Ranking attributes to build fuzzy decision trees: a comparative study of measures," in *Proc. of the IEEE Int. Conf. on Fuzzy Systems, FUZZ-IEEE*, Vancouver, Canada, July 2006, pp. 1777–1783.

[7] X. Wang and C. Borgelt, "Information measures in fuzzy decision trees," in *Proc. of the IEEE International Conference on Fuzzy Systems*, vol. 1, July 2004, pp. 85–90.

[8] W. Pedrycz and Z. Sosnowski, "Genetically optimized fuzzy decision trees," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 35, no. 3, pp. 633–641, June 2005.

[9] X. Liu and W. Pedrycz, "The development of fuzzy decision trees in the framework of axiomatic fuzzy set logic," *Applied Soft Computing*, vol. 7, pp. 325–342, 2007.

[10] K. Crockett, Z. Bandar, and D. Mclean, "On the optimization of t-norm parameters within fuzzy decision trees," in *Proceedings of the IEEE International Conference on Fuzzy Systems*, July 2007, pp. 1–6.

[11] L.-C. Dong, D. Wang, and X.-Z. Wang, "An improved sample selection algorithm in fuzzy decision tree induction," in *Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernectics*, San Antonio, TX (USA), October 2009, pp. 629–634.

[12] N. Abu-Halaweh and R. Harrison, "Practical fuzzy decision trees," in *Proc. of the IEEE Symposium on Computational Intelligence and Data Mining (CIDM '09)*, Nashville, (USA), March 2009, pp. 211 – 216.

[13] E. Hüllermeier and S. Vanderlooy, "Why fuzzy decision trees are good rankers," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 6, pp. 1233–1244, December 2009.

[14] J. Aczel, "Entropies, characterizations, applications and some history," in *Modern Information Processing, from Theory to applications*. Elsevier, 2006.

[15] C. Z. Janikow, "Fuzzy decision trees: Issues and methods," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 28, no. 1, pp. 1–14, 1998.

[16] M. Ramdani, "Une approche floue pour traiter les valeurs numériques en apprentissage," in *Journées Francophones d'apprentissage et d'explication des connaissances*, 1992.

[17] R. Weber, "Fuzzy-ID3: A class of methods for automatic knowledge acquisition," in *IIZUKA'92 Proceedings of the 2nd International Conference on Fuzzy Logic*, 1992, pp. 265–268.

[18] L. Zadeh, "Probability measures of fuzzy events," *Journal Math. Anal. Applic.*, vol. 23, 1968, 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. 45–51.

[19] Y. Yuan and M. Shaw, "Induction of fuzzy decision trees," *Fuzzy Sets and systems*, vol. 69, pp. 125–139, 1995.

[20] C. Olaru and L. Wehenkel, "A complete fuzzy decision tree technique," *Fuzzy Sets and Systems*, vol. 138, no. 2, pp. 221–254, September 2003.

[21] X. Boyen and L. Wehenkel, "Automatic induction of fuzzy decision tree and its application to power system security assessmen," *Fuzzy Sets and Systems*, vol. 102, no. 1, pp. 3–19, 1999.

[22] K. Cios and L. Sztandera, "Continuous ID3 algorithm with fuzzy entropy measures," in *Proceedings of the first International IEEE Conference on Fuzzy Systems*, San Diego, 1992.

[23] X. Wang, B. Chen, G. Qian, and F. Ye, "On the optimization of fuzzy decision trees," *Fuzzy Sets and Systems*, vol. 112, no. 1, pp. 117–125, May 2000.

[24] B. Chandra and P. Paul Varghese, "Fuzzifying Gini index based decision trees," *Expert Systems with Applications*, vol. 36, pp. 8549–8559, 2009.

[25] R. C. Guiaşu and S. Guiaşu, "Conditional and weighted measures of ecological diversity," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 11, no. 3, pp. 283–300, 2003.

[26] A. Lurie, C. Marsala, S. Hartley, B. Bouchon-Meunier, and D. Dusser, "Patients' perception of asthma severity," *Respiratory Medecine*, vol. 101, no. 10, pp. 2145–2152, October 2007.

[27] F. Gueyffier, F. Boutitie, J. P. Boissel, J. Coope, J. Cutler, T. Ekbom, R. Fagard, L. Friedman, H. M. Perry, and S. Pocock, "INDANA: a meta-analysis on individual patient data in hypertension. protocol and preliminary results." *Thérapie*, vol. 50, no. 4, pp. 353–62, 1995.

[28] C. Marsala, "A fuzzy decision tree based approach to characterize medical data," in *Proceedings of the IEEE International Conference on Fuzzy Systems (Fuzz-IEEE'2009)*, Jeju Island, Korea, August 2009, pp. 1332–1337.