2011 IEEE International Conference on Fuzzy Systems
June 27-30, 2011, Taipei, Taiwan

On the Cooperation of Interval-Valued Fuzzy Sets and Genetic Tuning to Improve the Performance of Fuzzy Decision Trees.

José Antonio Sanz and Humberto Bustince
Department of Automatica y Computacion
Universidad Publica de Navarra
Pamplona, 31006 Spain
Emails: {joseantonio.sanz, bustince}@unavarra.es

Alberto Fernández
Department of Computer Science
University of Jaén
Jaén, 23071 Spain
Email: alberto.fernandez@ujaen.es

Francisco Herrera
Department of Computer Science
and Artificial Intelligence
University of Granada
Granada, 18071 Spain
Email: herrera@decsai.ugr.es

***Abstract*—Fuzzy decision trees are widely employed to face classification problems since they combine the high interpretability given by the decision tree and the capability of management of the uncertainty inherent to fuzzy logic.**

However, the success of fuzzy systems in general depends, to a large degree, on the choice of the membership functions. For this reason, we propose to model the linguistic labels by means of Interval-Valued Fuzzy Sets to take into account the ignorance related to their definition. On the other hand, we define an evolutionary method to tune the shape of the Interval-Valued Fuzzy Sets looking for the best ignorance degree that each Interval-Valued Fuzzy Set represents.

In this contribution, we will make use of the fuzzy ID3 algorithm as a base technique from which to apply our methodology. The experimental study shows how our methodology enhances the performance of the base fuzzy decision tree. Furthermore, we compare our approach with respect to four state-of-the-art fuzzy decision trees and C4.5 as a representative algorithm for crisp decision trees. The goodness of our proposal is tested on a large collection of data-sets and it is supported by an exhaustive statistical analysis.

***Keywords*—Linguistic Fuzzy Rule-Based Classification Systems, Interval-Valued Fuzzy Sets, Ignorance functions, Tuning, Fuzzy Decision Tree, Classification.**

I. Introduction

Classification problem is one of the most studied problems in machine learning and data mining [1]. In order to face this task, a classifier needs to be induced by means of a learning algorithm. The classifier is a model encoding a set of criteria that allows a data instance to be assigned to a particular class depending on the value of certain variables. Different kinds of models, such as fuzzy rule-based classification systems or Fuzzy Decision Trees (FDTs), can be used to represent classifiers.

Specifically, FDTs are a suitable solution to face classification problems due to the good synergy between the handling of the uncertainty provided by fuzzy logic [2] and the good interpretability given by decision trees [3]. Among the large number of techniques proposed in the specialized literature for designing FDTs [4], [5], the fuzzy ID3 algorithm [6] is widely employed since it is a popular and efficient method to construct a FDT [7].

In fuzzy systems, like FDTs, the definition of the membership functions used to represent the linguistic labels is truly significant. When defining the fuzzy partitions, either through expert knowledge or homogeneously over the input space, it is possible to have an ignorance degree (lack of information) which makes their definition difficult. The theory of Interval-Valued Fuzzy Sets (IVFSs) [8] has proved to be a suitable tool to model this ignorance [9]. Specifically, we have shown that the use of IVFSs is useful to improve the performance of fuzzy rule-based classification systems [10].

In this contribution, we propose to model the linguistic labels of the classifier by means of IVFSs. To do so, we use a new parametrized construction method of IVFSs that starts from the fuzzy sets composing the knowledge base of the system. The length of each IVFS is proportional to the degree of ignorance related to the definition of the corresponding membership function, which is measured using weak ignorance functions. We must point out that we introduce the IVFSs after the induction process and we study their influence in the fuzzy reasoning mechanism, which has to be modified to take into account this modeling of the linguistic labels. As a result, a good management of the uncertainty is provided for the system while its interpretability is maintained.

Additionally, we propose the definition of an evolutionary methodology that allows to compute the optimal ignorance degree that each IVFS represents. To do so, we modify the parameters used in the IVFSs construction method to weight the degree of ignorance and consequently, we tune the length of the IVFSs. In this manner, the membership functions will be fine-tuned leading to an improvement of the performance of the system in a general framework.

The aim of our study is to show the goodness of our proposed methodology using the fuzzy ID3 algorithm as the base technique. To do so, we will compare our approach with respect to both the base FDT and with four state-of-the-art FDTs, namely the simple pattern tree [11], a look-ahead approach [12], the FDT proposed by Janikow in [13] and a method which fuzzifies the Gini index [14]. Furthermore, we will also compare our methodology with C4.5 [15]

978-1-4244-7316-8/11/$26.00

as a reference algorithm of crisp decision trees. In order to obtain well-founded conclusions, we have considered 20 numerical data-sets from the KEEL data-set repository [16] (http://www.keel.es/dataset.php). The measure of performance is based on accuracy rate and the study is supported by the proper statistical analysis, as suggested in the literature [17], [18], [19].

The work is arranged as follows: Section II provides some preliminary concepts of IVFSs and also describes the generation of IVFSs employing weak ignorance functions. The induction process of the fuzzy ID3 algorithm together with the modifications carried out in the reasoning mechanism, due to the use of IVFSs, are shown in Section III. In Section IV we describe in detail our methodology to tune the membership functions by means of GAs and the evolutionary method is also explained. Then, the experimental study is presented in Section V and we finish the work with some concluding remarks in Section VI.

II. Linguistic Labels Modeled by means of Interval-Valued Fuzzy Sets

In this work we propose to model the linguistic labels by means of IVFSs in order to take into account the ignorance related to the definition of the membership functions.

In this section, we remind some theoretical concepts which are necessary to understand the remainder of this contribution. Then, we define the construction method of IVFSs starting from given fuzzy sets in which we employ weak ignorance functions to measure the ignorance associated with the definition of the membership functions. Finally, we present our new methodology to model the linguistic labels by means of IVFSs.

A. Preliminaries

Let us denote by $L([0,1])$ the set of all closed subintervals in $[0,1]$, that is,

$$L([0,1]) = \{\mathbf{x} = [\underline{x}, \overline{x}] | (\underline{x}, \overline{x}) \in [0,1]^2 \text{ and } \underline{x} \leq \overline{x}\}.$$

We also denote $0_L = [0,0]$ and $1_L = [1,1]$.

Definition 1: An interval-valued fuzzy set (IVFS) [20], [21] (or interval type 2 fuzzy set) A on the universe $U \neq \emptyset$ is a mapping $A : U \to L([0,1])$, such that

$$A(u_i) = [\underline{A}(u_i), \overline{A}(u_i)] \in L([0,1]), \text{ for all } u_i \in U.$$

We denote by $\mathcal{FS}(U)$ the set of all fuzzy sets on U and by $\mathcal{IVFS}(U)$ the set of all IVFSs on U. We also denote by W the length of the interval under consideration, that is

$$W(A(u_i)) = \overline{A}(u_i) - \underline{A}(u_i).$$

To order the IVFSs we use the order relation defined by Xu and Yager [22]. Using this relationship on $L([0,1])$ is easy to see that the smallest element of $L([0,1])$ is 0_L and the largest is 1_L.

We will now recall the extension of t-norms and t-conorms in $[0,1]$ (see [23]) to IVFSs [20], [24].

Definition 2: A function $\mathbf{T} : (L([0,1]))^2 \to L([0,1])$ is said to be an *interval-valued t-norm* (IV t-norm) if it is commutative, associative, increasing in both arguments (with respect to the order $\mathbf{x} \leq_L \mathbf{y}$ if and only $\underline{x} \leq \underline{y}$ and $\overline{x} \leq \overline{y}$), and has the neutral element 1_L. In the same way, a function $\mathbf{S} : (L([0,1]))^2 \to L([0,1])$ is said to be an *interval-valued t-conorm* (IV t-conorm) if it is commutative, associative, increasing, and has the neutral element 0_L.

Definition 3: a) An IV t-norm is said to be *t-representable* if there are two t-norms T_a and T_b in $[0,1]$ so that $\mathbf{T}(\mathbf{x},\mathbf{y}) = [T_a(\underline{x},\underline{y}), T_b(\overline{x},\overline{y})]$ for all $\mathbf{x}, \mathbf{y} \in L([0,1])$.
b) An IV t-conorm is said to be *s-representable* if there are two t-conorms S_a and S_b in $[0,1]$ so that $\mathbf{S}(\mathbf{x},\mathbf{y}) = [S_a(\underline{x},\underline{y}), S_b(\overline{x},\overline{y})]$ for all $\mathbf{x}, \mathbf{y} \in L([0,1])$.

We will use $\mathbf{T}_{T_a,T_b}$ to denote the t-representable IV t-norm that can be represented by T_a and T_b as defined above. Similarly, a specific s-representable IV t-conorm will be denoted $\mathbf{S}_{S_a,S_b}$.

In this work we only use t-representable IV t-norms and s-representable IV t-conorms.

Let $[\underline{x}, \overline{x}]$, $[\underline{y}, \overline{y}]$ be two intervals in $\mathbb{R}$ so that $\underline{x} \leq \underline{y}$ and $\overline{x} \leq \overline{y}$, the rules of interval arithmetic employed in this contribution are as follows:

- Addition: $[\underline{x}, \overline{x}] + [\underline{y}, \overline{y}] = [\underline{x} + \underline{y}, \overline{x} + \overline{y}]$.
- Subtraction: $[\underline{x}, \overline{x}] - [\underline{y}, \overline{y}] = [\underline{x} - \overline{y}, \overline{x} - \underline{y}]$.
- Multiplication: $[\underline{x}, \overline{x}] * [\underline{y}, \overline{y}] = [\underline{x} * \underline{y}, \overline{x} * \overline{y}]$.
- Division: $[\underline{x}, \overline{x}] / [\underline{y}, \overline{y}] = [\underline{x}/\overline{y}, \overline{x}/\underline{y}]$ assuming $0 < \underline{y}$.

We defined a weak ignorance function from the concept of ignorance function given in [9] as follows:

Definition 4: A weak ignorance function [25] is a mapping

$$g : [0,1] \to [0,1]$$

that satisfies:

(g1) $g(x) = g(1-x)$ for all $x \in [0,1]$;
(g2) $g(x) = 0$ if and only if $x = 0$ or $x = 1$;
(g3) $g(0.5) = 1$.

Example 1: $g(x) = 2 \cdot \min(x, 1-x)$ is a weak ignorance function.

B. Construction of Interval-Valued Fuzzy Sets of Fixed Length From a Fuzzy Set

Our aim in this section is to construct an IVFS starting from any given fuzzy set. To do so, we define a function G parametrized by δ and γ, which satisfy a determined set of properties. These properties allow to obtain intervals in such a way that their length is proportional to the ignorance degree and the initial membership degree is within the interval.

Proposition 1: Let $\delta, \gamma \in [0,1]$ with $\delta \geq \gamma \geq \delta \cdot x$. The function

$$\begin{aligned} &G : [0,1]^4 \to L([0,1]) \text{ given by} \\ &G(x,y,\delta,\gamma) = [x \cdot (1 - \delta \cdot y), x \cdot (1 - \delta \cdot y) + \gamma \cdot y)] \end{aligned} \tag{1}$$

satisfies the following properties:

1) $x \in G(x,y,\delta,\gamma)$;
2) $W(G(x,y,\delta,\gamma)) = \gamma \cdot y$;

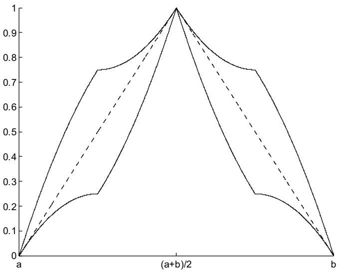

Fig. 1: Initial constructed IVFS.

3) If $x = 0$, then $G(0, y, \delta, \gamma) = [0, \gamma \cdot y]$;
4) If $y = 0$, then $G(x, 0, \delta, \gamma) = [x, x]$;
5) If $\delta = \gamma$, then:

$$\underline{G}(x, y, \delta, \delta) + \overline{G}(1 - x, y, \delta, \delta) = 1$$
$$\overline{G}(x, y, \delta, \delta) + \underline{G}(1 - x, y, \delta, \delta) = 1.$$

According to the previous proposition, we define Theorem 1 as the new construction of IVFSs.

Theorem 1: Let $A \in \mathcal{FS}(U)$. If for each $u_i \in U$ we take $g(A(u_i))$, $\delta(u_i)$, $\gamma(u_i) \in [0, 1]$, then the set

$$\begin{aligned} &A_{IV} = \{(u_i, A_{IV}(u_i)) | u_i \in U\} \text{ where} \\ &A_{IV}(u_i) = G(A(u_i), g(A(u_i)), \delta(u_i), \gamma(u_i)) \end{aligned} \tag{2}$$

is an IVFS on U.

C. Modeling Linguistic Labels by means of Interval-Valued Fuzzy Sets

In this section, we describe our methodology to generate linguistic labels modeled by means of IVFSs starting from given fuzzy sets.

We use the initial fuzzy sets (dashed line in Fig. 1 whose expression is shown in Eq. (3)) as the base to construct the IVFSs (solid line in Fig. 1) by means of Theorem 1. In the initial construction of the IVFSs, we consider the middle degree of ignorance. Therefore, we initialize $\delta(u_i) = \gamma(u_i) = 0.5$ for all $u_i \in U$ since, according to Theorem 1, the minimum value of both parameters is 0 and the maximum is 1.

$$\mu_A(u_i) = \begin{cases} 0, & \text{if } u_i \leq a, \\ \frac{2}{b-a}(u_i - a), & \text{if } a \leq u_i \leq \frac{a+b}{2} \text{ and } b \neq a, \\ \frac{2}{a-b}(u_i - b), & \text{if } \frac{a+b}{2} \leq u_i \leq b \text{ and } b \neq a, \\ 0, & \text{if } b \leq u_i, \end{cases} \tag{3}$$

III. Fuzzy ID3 with Interval-valued Fuzzy Sets

In this section, we describe the induction process of the fuzzy ID3 algorithm and then we introduce our methodology, that is, the construction of the IVFSs and the modifications carried out both in the computation of the relative frequencies and in the reasoning mechanism.

A. Fuzzy ID3 Induction

Let the universe of objects be described by n attributes $A = A^{(1)}, ..., A^{(n)}$. An attribute $A^{(k)}$ takes m_k values of fuzzy subsets $A_1^{(k)}, A_2^{(k)}, ..., A_{m_k}^{(k)}$. There are a total of N training instances. Based on the attributes, an object is classified into *m* classes $C_1, C_2, ..., C_m$. We use the symbol M($\cdot$) to denote the cardinality of a given fuzzy set. The key to the fuzzy ID3 algorithm is to select the expanded attribute, which can be performed in the following steps [7]:

(1) For each attribute value (fuzzy subset) $A_i^{(k)}, (i = 1, 2, ..., m_k)$, to compute its relative frequencies with respect to class $C_j, (j = 1, 2, ..., m)$:

$$p_i^{(k)}(j) = \frac{M(A_i^{(k)} \cap C_j)}{M(A_i^{(k)})} \tag{4}$$

(2) For each attribute value (fuzzy subset) $A_i^{(k)}, (i = 1, 2, ..., m_k)$, to compute its fuzzy classification entropy:

$$Entr_i^{(k)} = \sum_{j=1}^{m} -p_i^{(k)}(j) log(p_i^{(k)}(j)), \tag{5}$$

(3) To compute the average fuzzy classification entropy of the *k*th attribute:

$$E_k = \sum_{i=1}^{m_k} \frac{M(A_i^{(k)})}{\sum_{j=1}^{m_k} M(A_j^{(k)})} Entr_i^{(k)}. \tag{6}$$

(4) Select such an integer k_0, that

$$E_{k_0} = \arg \min_{1 \leq k \leq n} (E_k) \tag{7}$$

With a given evidence significance level α, a truth level threshold β and being A the set of attributes of the problem, the induction process consists of the following steps [6]:

(1) Execute the α-cut of the set of fuzzified examples with the evidence significant level.
(2) Measure the average fuzzy classification entropy associated with each attribute and select the attribute with the smallest average fuzzy classification entropy as the root decision node.
(3) Delete all empty branches of the decision node. For each nonempty branch of the decision node, compute the relative frequencies of all objects within the branch into each class. If the relative frequency of one class is above the given threshold β, terminate the branch as a leaf. Otherwise, do further research if an additional attribute will further partition the branch (i.e. generate more than one nonempty branch). If yes, select the attribute with the smallest average fuzzy classification entropy as a new decision node from the branch. If not, terminate this branch as a leaf. At the leaf, each class will has its relative frequency.
(4) Repeat step 3 for all newly generated decision nodes until no further growth is possible (i.e. all the attributes have been used), the decision tree then is complete.

B. Fuzzy ID3 with Interval-Valued Fuzzy Sets: A New Fuzzy Reasoning Method

The fuzzy sets composing the data base of the FDT, generated by the algorithm previously described, are created homogeneously over the input space. These fuzzy sets are employed as the starting point to construct the IVFSs using the construction method presented in Section II-C.

The modeling of the linguistic labels by means of IVFSs implies two main modifications. The relative frequencies of the classes in each leaf must be elements of $L([0,1])$. To do so, we compute the relative frequencies (Eq. (4)) associated with the lower and the upper bounds respectively. On the other hand, the reasoning mechanism, used to classify novel examples, is adapted as follows:

In a FDT, a path is the connection from the root node to each leaf. Suppose that the FDT contains l leaves, consequently l paths $(Path^i, i = 1, 2, ..., l)$. For each leaf, the path i has N_i nodes ($Path^i_1$, $Path^i_2$, ...,$Path^i_{N_i}$). Let e be an example to be classified in one of the m classes. We must recall that each node is modeled by means of an IVFS, therefore, $Path^i_k(e_k) = [\underline{Path^i_k}(e_k), \overline{Path^i_k}(e_k)] \in L([0,1])$ with $k = 1, ..., N_i$. The *interval-valued fuzzy reasoning mechanism* follows the following 4 key steps:

- To compute the matching degree between each path and the new example. We apply a t-representable IV t-norm as the conjunction among the nodes in each path.

$$M_i = \mathbf{T}_{T_a,T_b}([\underline{Path^i_1}(e_1), \overline{Path^i_1}(e_1)], ...,$$
$$[\underline{Path^i_{N_i}}(e_{N_i}), \overline{Path^i_{N_i}}(e_{N_i})]) =$$
$$[T_a(\underline{Path^i_1}(e_1), ..., \underline{Path^i_{N_i}}(e_{N_i})),$$
$$T_b(\overline{Path^i_1}(e_1), ..., \overline{Path^i_{N_i}}(e_{N_i}))], i = 1, 2, ..., l$$

- To compute the certainty of each class in each leaf. We apply a t-representable IV t-norm to weight the matching degree of the example along the paths with the relative frequencies of the leaves.

$$Cert^i_j = \mathbf{T}_{T_c,T_d}([\underline{M_i}, \overline{M_i}], [\underline{p_i}(j), \overline{p_i}(j)]) =$$
$$[T_c(\underline{M_i}, \underline{p_i}(j)), T_d(\overline{M_i}, \overline{p_i}(j))], j = 1, ..., m, i = 1, ..., l$$

- To compute the total certainty of each class. We employ the interval addition to aggregate certainties of the same class in all the paths.

$$Total\_Certainty_j = \sum_{i=1}^{l} [\underline{Cert^i_j}, \overline{Cert^i_j}], j = 1, ..., m$$

- Classify the example in the class which maximizes the total certainty following the relationship presented below.

For any interval $[\underline{x}, \overline{x}]$, $[\underline{y}, \overline{y}]$ on $\mathbb{R}$, we use the following relationship based on the score and accuracy functions given in [22] (see Subsection II-A): let $s([\underline{x}, \overline{x}]) = \underline{x} + \overline{x}$ and $s([\underline{y}, \overline{y}]) = \underline{y} + \overline{y}$ be the scores of $[\underline{x}, \overline{x}]$ and $[\underline{y}, \overline{y}]$ respectively. Let $h([\underline{x}, \overline{x}]) = \overline{x} - \underline{x}$ and $h([\underline{y}, \overline{y}]) = \overline{y} - \underline{y}$ be the accuracy degrees of $[\underline{x}, \overline{x}]$ and $[\underline{y}, \overline{y}]$ respectively. Then

- If $s([\underline{x}, \overline{x}]) < s([\underline{y}, \overline{y}])$, then $[\underline{x}, \overline{x}] < [\underline{y}, \overline{y}]$;
- If $s([\underline{x}, \overline{x}]) = s([\underline{y}, \overline{y}])$, then
  a) If $h([\underline{x}, \overline{x}]) = h([\underline{y}, \overline{y}])$, then $[\underline{x}, \overline{x}] = [\underline{y}, \overline{y}]$;
  b) if $h([\underline{x}, \overline{x}]) < h([\underline{y}, \overline{y}])$, then $[\underline{x}, \overline{x}] < [\underline{y}, \overline{y}]$.

IV. Genetic Tuning of the Linguistic Labels

Membership functions, which are usually obtained by normalization process or defined by experts, remains fixed during the induction process. Consequently, they could not be perfectly adapted to the context of each problem. To solve this problem, it can be appropriate to carry out a post-processing tuning step in which membership functions are tuned leading to a better adaptation of the system to the specific problem we are dealing with.

The optimization post-processing step is often performed by means of a GA. The hybridization between fuzzy logic and GAs leads to Genetic Fuzzy Systems (GFSs) [26], which are basically fuzzy systems augmented by learning processes based on evolutionary computation. GFS approaches are usually divided into two processes, tuning and learning. Specifically, the genetic tuning process consists of automatically selecting the best system parameters for improving the performance of the initial model without modifying the rule base.

In the remainder of this section we will first introduce the tuning approach of the linguistic labels and then we will describe in detail the evolutionary model for tuning the membership functions.

A. Ignorance Weighting Factor Tuning

The initial set-up of the linguistic labels, which are modeled by IVFSs, could not be well-suited for the specific problem we are dealing with. Therefore, it seems necessary to tune the shape of each IVFS, which is determined by the parameters employed in the IVFSs construction method (Section II-B).

As we have introduced in Subsection II-C, we consider the middle degree of ignorance for the initial construction of all the IVFSs ($\delta = \gamma = 0.5$). However, the ignorance degree can vary depending on the linguistic label, since the amount of available information to define each fuzzy term may not be the same.

Therefore, we propose the application of a genetic tuning step in which we adapt the parameters δ and γ (see Theorem 1), keeping the restriction $\delta \geq \gamma \geq \delta \cdot x$. Consequently, the amount of ignorance that each IVFS represents will vary depending on how the membership function suits the problem.

Fig. 2 depicts the behaviour of the proposed tuning approach. The original IVFS is shown in light gray whereas its final shape (afterwards the genetic tuning) is depicted in dark gray. We observe that the shape of the final IVFS is thinner than that of the initial IVFS.

B. Evolutionary Model

In order to apply the genetic tuning, we will consider the use of CHC algorithm [27]. This algorithm is a classical evolutionary model which presents a good trade-off between

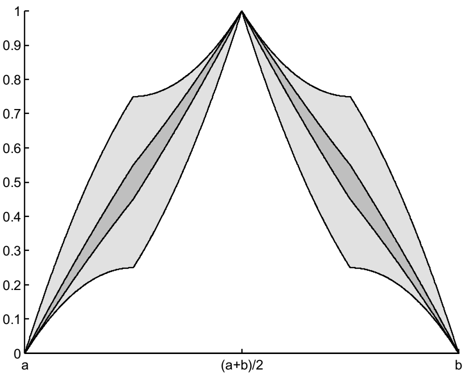

Fig. 2: Ignorance weighting factor genetic tuning. The final values of the parameters are: $\delta = 0.1$, $\gamma = 0.1$.

diversity and convergence, being a good choice in complex problems. The components needed to design this process are explained below:

1) *Coding Scheme:* A real coding is considered. Let us consider the following number of labels per variable: $(m^1, m^2, ..., m^n)$, with n being the number of variables. δ and $\gamma \in [0, 1]$ represent the genes used to encode the parameters which weight the degree of ignorance. Then, the representation of the chromosome is as follows:

$$\begin{aligned} C_{ET} = (&\delta_{11}, \gamma_{11}, ..., \delta_{1m^1}, \gamma_{1m^1}, \\ &\delta_{21}, \gamma_{21}, ..., \delta_{2m^2}, \gamma_{2m^2}, ..., \\ &\delta_{n1}, \gamma_{n1}, ..., \delta_{nm^n}, \gamma_{nm^n}) \end{aligned}$$

   Therefore, the chromosome length is equal to two times the number of variables times the number of linguistic labels.
2) *Chromosome Evaluation:* The fitness function is the classification accuracy.
3) *Initial Gene Pool:* The initial pool is obtained initializing three individuals having all of the genes with values of 0, 0.5 and 1, corresponding with the situations of minimum, middle and maximum degree of ignorance respectively. The remaining individuals will have initialized all the genes randomly in their respective domains.
4) *Crossover Operator:* We consider the Parent Centric BLX (PCBLX) operator, which is based on the BLX-α. We consider the incest prevention mechanism, checking and modifying an initial threshold, in order to apply the PCBLX operator.
5) *Restarting approach:* When the threshold value is lower than zero, all the chromosomes are regenerated at random in their respective domains. Furthermore, the best global solution found is included in the population to increase the convergence of the algorithm as in the elitist scheme.

V. Experimental Study

This study is oriented towards analysing the goodness of our evolutionary tuning method applied to the FDT generated by the fuzzy ID3 algorithm. To do so, we develop an experimental analysis in which we compare our methodology with respect to several state-of-the-art FDTs and also with C4.5.

In this section, we present the experimental set-up in first place. Next, we briefly describe the FDTs for the comparison together with the parameters employed in the study. Then we introduce the statistical tests used and finally, we present the results together with the corresponding analysis.

A. Experimental Set-Up

We have selected a wide benchmark of 20 numerical data-sets selected from the KEEL data-set repository [16], which are publicly available on the corresponding web page [1] including general information about them, partitions for the validation of the experimental results and so on. Table I summarizes the properties of the selected data-sets, showing for each data-set the number of examples (#Ex.), the number of attributes (#Atts.) and the number of classes (#Class.). We must point out that the *magic, page-blocks, ring* and *shuttle* data-sets have been stratified sampled at 10% in order to reduce their size for training. In the case of missing values, (*cleveland* and *wisconsin*), those instances have been removed from the data-set.

TABLE I: Summary Description for the employed data-sets.

Id.	Data-set	#Ex.	#Atts.	#Class.
app	Appendicitis	106	7	2
bal	Balance	625	4	3
bup	Bupa	345	6	2
cle	Cleveland	297	13	5
eco	Ecoli	336	7	8
gla	Glass	214	9	6
hab	Haberman	306	3	2
hea	Heart	270	13	2
iri	Iris	150	4	3
mag	Magic	1,902	10	2
new	New-Thyroid	215	5	3
pag	Page-blocks	548	10	5
pim	Pima	768	8	2
rin	Ring	740	20	2
shu	Shuttle	2,175	9	7
tae	Tae	151	5	3
tit	Titanic	2,201	3	2
win	Wine	178	13	3
wis	Wisconsin	683	9	2
yea	Yeast	1484	8	10

A *5-folder cross-validation model* was considered in order to carry out the different experiments. That is, we split the data-set into 5 random partitions of data, each one with 20% of the patterns, and we employ a combination of 4 of them (80%) to train the system and the remaining one to test it. Furthermore, the process was repeated 3 times using a different seed to obtain a sample of 15 results, which have been averaged for each data-set.

[1] http://www.keel.es/dataset.php

B. Fuzzy Decision Trees for Comparison and Parameters Set-Up

In this work, we have considered four FDTs in order to compare our methodology with respect to different approaches of the literature. Next, we briefly describe each FDT together with their specific configuration values:

- The first approach fuzzifies the Gini index [14]. This method uses the SLIQ decision tree as the base algorithm, which uses the Gini index as the split measure, and fuzzifies its decision boundaries during the tree induction. The configuration is as follows:
  - Number of labels per variable: 3 labels.
  - Threshold = 0.85.
  - Maximum number of patterns in a leaf: 5% of the number of patterns.
  - Maximum depth: number of attributes.
- The second approach is the look-ahead method [12]. This method attempts to establish a decision node by analyzing both a split measure and the classifiability of instances that are split along branches of the node. The specific values of its configuration are:
  - Number of labels per variable: 3 labels.
  - Split measure: Information Gain.
  - Number of neighbours: 3.
  - Weighting factor: 1.
- The third proposal is the FDT proposed by Janikow in [13]. This is a classical FDT widely employed in the specialized literature whose configuration is as follows:
  - Number of labels per variable: 3 labels.
  - Conjunction operator: product T-norm.
- Finally, we have selected the simple pattern tree algorithm [11]. This approach, instead of constructing one tree, whose leaves have a probability distribution expressing the membership of each class, constructs one tree per class using aggregation and similarity functions. We have chosen the following configuration:
  - Aggregation functions: min and max operators.
  - Similarity function: the similarity related with the root mean square error.
  - Number of labels per variable: 3 labels.

The specific configuration for the fuzzy ID3 algorithm with and without IVFSs (presented in Table II) and the ones of the selected FDTs in this work, have been selected experimentally since they provide the best performance for each method. Regarding the genetic tuning process presented in Section IV, it employs populations composed by 50 individuals, 30 bits per gen in order to perform the *gray* codification and the number of evaluations is 5.000 times the number of attributes.

C. Statistical Tests for Performance Comparison

In this work, we use some hypothesis validation techniques in order to give a statistical support to the result analysis. We will use a non parametric test, because the initial conditions that guarantee the reliability of the parametric tests cannot be fulfilled [17], [18]. We employ the Wilcoxon rank test [28] as non parametric statistical procedure to make comparisons between two algorithms; we use the Friedman aligned ranks test [29] to detect statistical differences among a group of results and the Holm post-hoc test [30] to find the algorithms that reject the equality hypothesis with respect to a selected control method.

TABLE II: Parameter specification for the FDTs.

Fuzzy ID3
Number of labels per variable: 3 labels Conjunction operator: product t-norm Evidence significance level = 0.4 Truth level threshold = 0.95
Fuzzy ID3 with IVFSs
Ignorance function: $2 \cdot \min(x, 1-x)$ Conjunction operator: product IV t-norm δ = 0.5 γ = 0.5

These tests are suggested in the studies presented in [17], [18], [19], where its shown that their use in the field of machine learning is highly recommended. A complete description of the test and software for its use can be found on the website: http://sci2s.ugr.es/sicidm/.

D. Analysis of the Behaviour of the Evolutionary Tuning Method.

Table III shows the classification accuracy provided by our methodology and the different approaches considered in this work, specifically:

- F-ID3_IVFS_ET: the fuzzy ID3 algorithm with IVFSs and the evolutionary tuning method.
- F-ID3: the standard fuzzy ID3 algorithm.
- SPT: the simple pattern tree algorithm.
- LA: the look-ahead approach with the information gain heuristic.
- Fgini: the approach in which authors fuzzify the Gini index.
- Janikow: the classical FDT.
- C4.5: the well known decision tree.

Results are grouped in pairs for training and test, where the best global result for each data-set is stressed in **bold-face**.

From the results of Table III, it can be observed that our methodology enhances the performance of the fuzzy ID3 algorithm in most of the data-sets. Furthermore, the goodness of our evolutionary tuning model is shown since it increases the initial performance of the fuzzy ID3 algorithm in such a way that the results obtained by the remaining FDTs are clearly outperformed, as it achieves the best global performance and provides the best performance in 11 of the 17 data sets. This situation is confirmed in Fig. 3, where the rankings of the different approaches are presented, showing that the best ranking is provided by our evolutionary tuning method.

In order to detect significant differences among the results of the different approaches, we carry out the Friedman aligned rank test. This test obtains a p-value near to zero (0.004),

TABLE III: Results in Train (Tr.) and Test (Tst) achieved by the different decision trees.

Data	F-ID3_IVFS_ET		F-ID3		SPT		LA		Janikow		Fgini		C4.5	
Set	$Tr.$	Tst	$Tr.$	Tst	$Tr.$	Tst	$Tr.$	Tst	$Tr.$	Tst	$Tr.$	Tst	$Tr.$	Tst
App	93.16	86.80	90.57	84.94	80.19	80.22	91.27	84.85	89.39	**86.84**	80.19	80.22	90.09	84.98
Bal	96.60	**90.72**	91.72	90.08	78.96	77.92	90.32	88.32	91.72	90.08	89.40	88.64	89.72	77.28
Bup	79.93	**67.25**	61.52	58.26	63.99	60.29	79.78	61.74	60.65	58.84	88.62	66.67	83.84	66.09
Cle	93.18	**55.88**	87.29	53.87	53.87	53.88	94.69	49.83	97.64	50.50	89.06	53.19	83.41	51.82
Eco	82.14	77.69	80.21	76.80	67.11	66.67	80.73	75.90	77.98	76.79	89.88	77.39	91.74	**78.28**
Gla	77.57	60.80	60.17	52.80	60.40	57.52	85.86	68.24	75.12	64.04	46.26	43.02	91.94	**68.73**
Hab	79.90	72.87	75.08	71.90	75.98	74.17	78.43	73.52	74.51	72.88	79.90	**74.83**	76.06	72.22
Hea	94.81	78.52	93.52	**79.26**	81.57	75.19	98.43	75.93	98.52	78.52	90.93	72.96	92.96	**79.26**
Iri	98.83	96.00	93.07	95.33	97.17	**98.67**	97.50	96.00	95.67	96.00	83.67	83.33	97.83	93.33
Mag	82.75	78.91	95.50	78.28	78.12	78.23	86.40	78.44	78.52	77.55	79.98	77.18	87.22	**79.81**
New	97.79	**94.88**	79.80	91.16	91.86	91.63	96.05	93.95	86.05	85.58	93.49	93.02	98.37	91.16
Pag	95.03	94.16	92.67	92.15	92.97	91.24	95.67	93.79	93.75	92.51	91.70	91.24	98.95	**95.07**
Pim	83.56	**76.43**	92.84	75.77	76.20	74.34	88.25	73.57	77.54	74.34	85.38	75.78	85.81	74.09
Rin	97.09	**90.81**	79.95	49.59	80.78	78.11	94.76	77.03	95.00	90.14	88.55	84.86	97.13	82.70
Shu	97.74	97.98	98.74	90.57	92.72	92.64	97.72	97.33	83.31	83.31	99.46	99.49	99.66	**99.54**
Tae	70.86	**58.95**	63.40	54.99	55.30	48.99	71.36	53.61	68.38	57.61	56.28	50.34	78.15	54.99
Tit	78.33	**78.33**	78.33	**78.33**	77.64	77.78	**78.33**	**78.33**	78.33	78.33	67.37	66.65	78.48	77.78
Win	100.00	**98.87**	99.59	97.75	91.85	89.33	98.45	92.11	100.00	97.71	94.10	90.40	99.02	94.90
Wis	98.28	96.04	90.63	94.58	96.85	96.20	96.85	95.02	98.43	**96.49**	94.40	93.56	98.43	95.03
Yea	63.01	**57.41**	95.90	55.46	35.56	34.16	62.99	51.29	46.60	44.54	21.38	20.95	82.18	55.80
Mean	88.03	**80.46**	81.02	76.09	76.45	74.86	88.19	77.94	83.35	77.63	80.50	74.19	90.05	78.64

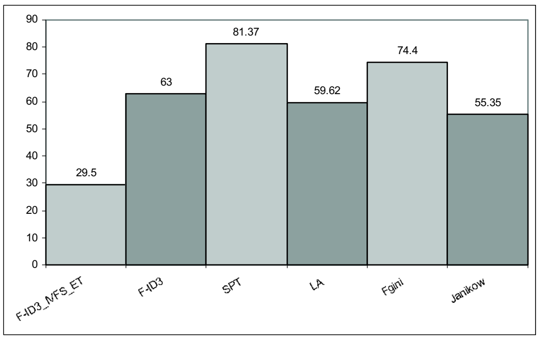

Fig. 3: Rankings of the different fuzzy decision trees.

which implies that there are significant differences between the results. For this reason, we can apply a post-hoc test to compare our methodology against the remaining FDTs. Specifically, a Holm test is applied, which is presented in Table IV. The statistical analysis reflects that all of the FDTs considered in this study are outperformed by our approach.

TABLE IV: Holm test to compare the evolutionary tuning method with all the FDTs. The evolutionary tuning model is used as the control method.

i	Algorithm	z	p	α/i	Hypothesis ($\alpha = 0.05$)
5	SPT	4.74	2.15E-6	0.01	Rejected for F-ID3_IVFS_ET
4	Fgini	4.10	4.05E-5	0.0125	Rejected for F-ID3_IVFS_ET
3	F-ID3	3.07	0.002	0.017	Rejected for F-ID3_IVFS_ET
2	LA	2.76	0.006	0.025	Rejected for F-ID3_IVFS_ET
1	Janikow	2.37	0.018	0.05	Rejected for F-ID3_IVFS_ET

Regarding C4.5, the results presented in Table III also show that our evolutionary tuning approach achieves a significant global improvement. A Wilcoxon test (Table V) will help us to make a pairwise comparison between both approaches. We can observe that the null hypothesis of equivalence is rejected in favour of the evolutionary tuning method with a high level of confidence.

TABLE V: Wilcoxon Test to compare the C4.5 algorithm (R^+) against the evolutionary tuning method (R^-).

Comparison	R^+	R^-	Hypothesis ($\alpha = 0.05$)	p-value
C4.5 vs. F-ID3_IVFS_ET	44	166	Rejected for F-ID3_IVFS_ET	0.023

VI. CONCLUSION

In this work we have proposed a methodology to improve the performance of FDTs by means of IVFSs and a post-processing genetic tuning step. To do so, we have modeled the linguistic labels by means of IVFSs to take into account the ignorance related to the definition of the membership functions. The degree of ignorance is weighted by two parameters and, in this manner, the shapes of the IVFSs are determined. This parametrization allows us to analyse the most appropriate set-up of the IVFSs partitions by means of a post-processing genetic tuning step. In this manner, we compute the best ignorance degree that each IVFS represents.

The experimental study has shown the quality of our approach since, applied to the FDT generated by the fuzzy ID3 algorithm, allows to outperform both the results of the base FDT and the ones of several state-of-the-art FDTs. We must highlight that our proposed approach also enhances the results of a reference algorithm in data mining like C4.5. The results obtained with this methodology allow us to determine, with the corresponding statistical support, that the proposed synergy is very useful to solve classification tasks.

ACKNOWLEDGMENT

This work was supported in part by the Spanish Ministry of Science and Technology under projects TIN2008-06681-C06-01 and TIN2010-15055.

REFERENCES

[1] E. Alpaydin, *Introduction to Machine Learning*, 2nd ed. The MIT Press, Cambridge, 2010.

[2] L. A. Zadeh, "Fuzzy sets," *Information and control*, vol. 8, no. 3, pp. 338–353, 1965.

[3] R. Kothari and M. Dong, *Decision Trees for Classification: A Review and Some New Results*. World Scientific, Singapore, 2000.

[4] K. A. Crockett, Z. Bandar, J. Fowdar, and J. O'Shea, "Genetic tuning of fuzzy inference within fuzzy classifier systems," *Expert Systems*, vol. 23, no. 2, pp. 63–82, 2006.

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

[6] Y. Yuan and M. J. Shaw, "Induction of fuzzy decision trees," *Fuzzy Sets and Systems*, vol. 69, no. 2, pp. 125–139, 1995.

[7] X.-Z. Wang, D. S. Yeung, and E. C. C. Tsang, "A comparative study on heuristic algorithms for generating fuzzy decision trees," *IEEE Transactions on Systems, Man and Cybernetics - Part B*, vol. 31, no. 2, pp. 215–226, 2001.

[8] R. Sambuc, "Function ϕ-flous, application a l'aide au diagnostic en pathologie thyroidienne," Ph.D. dissertation, University of Marseille, 1975.

[9] H. Bustince, M. Pagola, E. Barrenechea, J. Fernandez, P. Melo-Pinto, P. Couto, H. R. Tizhoosh, and J. Montero, "Ignorance functions. An application to the calculation of the threshold in prostate ultrasound images," *Fuzzy Sets and Systems*, vol. 161, no. 1, pp. 20–36, 2010.

[10] J. Sanz, A. Fernández, H. Bustince, and F. Herrera, "Improving the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets and genetic amplitude tuning," *Information Sciences*, vol. 180, no. 19, pp. 3674–3685, 2010.

[11] Z. Huang, T. Gedeon, and M. Nikravesh, "Pattern trees induction: A new machine learning method," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 4, pp. 958–970, 2008.

[12] M. Dong and R. Kothari, "Look-ahead based fuzzy decision tree induction," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 3, pp. 461–468, 2001.

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

[14] B. Chandra and P. P. Varghese, "Fuzzifying gini index based decision trees," *Expert Systems with Applications*, vol. 36, no. 4, pp. 8549–8559, 2009.

[15] J. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo-California: Morgan Kaufmann Publishers, 1993.

[16] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, and F. Herrera, "KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework," *Journal of Multiple-Valued Logic and Soft Computing*, vol. 17, no. 2-3, pp. 255–287, 2011.

[17] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.

[18] S. García and F. Herrera, "An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons," *Journal of Machine Learning Research*, vol. 9, pp. 2677–2694, 2008.

[19] S. García, A. Fernández, J. Luengo, and F. Herrera, "Advanced non-parametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power," *Information Sciences*, vol. 180, no. 10, pp. 2044–2064, 2010.

[20] H. Bustince, E. Barrenechea, and M. Pagola, "Generation of interval-valued fuzzy and atanassov's intuitionistic fuzzy connectives from fuzzy connectives and from k_α operators: laws for conjunctions and disjunctions, amplitude," *International Journal of Intelligent Systems*, vol. 23, no. 6, pp. 680–714, 2008.

[21] J. Mendel and R. I. John, "Type-2 fuzzy sets made simple," *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 2, pp. 117–127, 2002.

[22] Z. Xu and R. Yager, "Some geometric aggregation operators based on intuitionistic fuzzy sets," *International Journal of General Systems*, vol. 35, no. 4, pp. 417–433, 2006.

[23] G. Beliakov, A. Pradera, and T. Calvo, *Aggregation Functions: A Guide for Practitioners. What is an aggregation function*. San Mateo-California: Studies In Fuzziness and Soft Computing, Springer, 2007.

[24] G. Deschrijver, C. Cornelis, and E. Kerre, "On the representation of intuitionistic fuzzy t-norms and t-conorms," *IEEE Transactions on Fuzzy Systems*, vol. 12, no. 1, pp. 45–61, 2004.

[25] J. Sanz, A. Fernández, H. Bustince, and F. Herrera, "A genetic tuning to improve the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets: Degree of ignorance and lateral position," *International Journal of Approximate Reasoning*, vol. In press, doi: 10.1016/j.ijar.2011.01.011, 2011.

[26] F. Herrera, "Genetic fuzzy systems: Taxonomy, current research trends and prospects," *Evolutionary Intelligence*, vol. 1, no. 1, pp. 27–46, 2008.

[27] L. Eshelman, *Foundations of Genetic Algorithms*. Morgan Kaufman, 1991, ch. The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination, pp. 265–283.

[28] D. Sheskin, *Handbook of parametric and nonparametric statistical procedures*, 2nd ed. Chapman & Hall/CRC, 2006.

[29] J. L. Hodges and E. L. Lehmann, "Ranks methods for combination of independent experiments in analysis of variance," *Annals of Mathematical Statistics*, vol. 33, pp. 482–497, 1962.

[30] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, vol. 6, pp. 65–70, 1979.