International Journal of Uncertainty,
Fuzziness and Knowledge-Based Systems
Vol. 20, Suppl. 2 (October 2012) 1–30
© World Scientific Publishing Company
DOI: 10.1142/S0218488512400132

World Scientific
www.worldscientific.com

IIVFDT: IGNORANCE FUNCTIONS BASED INTERVAL-VALUED FUZZY DECISION TREE WITH GENETIC TUNING

J. SANZ* and H. BUSTINCE†

Department of Automática and Computación, Universidad Publica de Navarra,
Pamplona, Spain 31006, Spain
**joseantonio.sanz@unavarra.es*
†bustince@unavarra.es

A. FERNÁNDEZ

Department of Computer Science, University of Jaén,
Jaén, Spain 23071, Spain
alberto.fernandez@ujaen.es

F. HERRERA

Department of Computer Science and Artificial Intelligence,
CITIC-UGR (Research Center on Information and Communications Technology),
University of Granada, Granada, Spain 18071, Spain
herrera@decsai.ugr.es

Received 5 October 2011
Revised 3 May 2012

The choice of membership functions plays an essential role in the success of fuzzy systems. This is a complex problem due to the possible lack of knowledge when assigning punctual values as membership degrees. To face this handicap, we propose a methodology called *Ignorance functions based Interval-Valued Fuzzy Decision Tree with genetic tuning*, IIVFDT for short, which allows to improve the performance of fuzzy decision trees by taking into account the ignorance degree. This ignorance degree is the result of a weak ignorance function applied to the punctual value set as membership degree.

Our IIVFDT proposal is composed of four steps: (1) the base fuzzy decision tree is generated using the fuzzy ID3 algorithm; (2) the linguistic labels are modeled with Interval-Valued Fuzzy Sets. To do so, a new parametrized construction method of Interval-Valued Fuzzy Sets is defined, whose length represents such ignorance degree; (3) the fuzzy reasoning method is extended to work with this representation of the linguistic terms; (4) an evolutionary tuning step is applied for computing the optimal ignorance degree for each Interval-Valued Fuzzy Set.

The experimental study shows that the IIVFDT method allows the results provided by the initial fuzzy ID3 with and without Interval-Valued Fuzzy Sets to be outperformed. The suitability of the proposed methodology is shown with respect to both several state-of-the-art fuzzy decision trees and C4.5. Furthermore, we analyze the quality of our approach versus two methods that learn the fuzzy decision tree using genetic algorithms.

Finally, we show that a superior performance can be achieved by means of the positive synergy obtained when applying the well known genetic tuning of the lateral position after the application of the IIVFDT method.

Keywords: Linguistic fuzzy rule-based classification systems; interval-valued fuzzy sets; ignorance functions; tuning; fuzzy decision trees; classification.

1. Introduction

Classification is one of the most studied problems in machine learning and data mining.[1,2] In many classification problems there is a quantity of complex information that any human user can process in a natural way but which is difficult to represent and to process in a classifier. Consequently, in order to design an interpretable and accurate classifier it is necessary to draw upon a suitable tool to handle this information.[3]

The hybridization of fuzzy sets[4] with decision trees[5] naturally enhances the representative power of decision trees with the knowledge component inherent in fuzzy logic, leading to greater robustness and applicability in uncertain or imprecise domains.[6] Numerous techniques have been proposed in the specialized literature for designing Fuzzy Decision Trees (FDTs).[6–8] Specifically, fuzzy ID3 algorithm and its variants[9–11] are popular and efficient methods for inducing FDTs.[12]

The use of linguistic labels enables the acquisition of interpretable knowledge systems and, in this manner, the choice of the membership function plays an essential role in their success. The punctual value set as membership degree is usually defined either by means of expert knowledge or homogeneously over the input space. In both cases, there can be a lack of knowledge associated with their assignment. To face it, one solution is to employ Interval-Valued Fuzzy Sets (IVFSs),[13] whose length represent the degree of ignorance when assigning punctual values as membership degrees.[14,15] In order to compute the ignorance degree we use the concept of ignorance function,[14] which is completely different to the ignorance of an event defined in the possibility theory.[16] IVFSs have been applied successfully to numerous topics such as classification,[17] image processing,[18,19] multiple criteria analysis[20] and computing with words,[21] among others.

In addition to the previous issue, the amount of available information to define the membership functions associated with the different linguistic terms may not be the same. Consequently, the ignorance degree that their corresponding IVFSs represents can vary for each of them. For this reason, it seems necessary to carry out a tuning step to compute the best ignorance degree for each IVFS. Genetic Algorithms (GAs) have been applied successfully to compute the optimal values of the membership functions' parameters[22–25] due to the fact that they consider many points of the search space simultaneously and, therefore, they reduce the chances of converging to local optima.[26]

In this paper, we aim to improve the performance of FDTs exploiting the suitable features of both IVFSs and GAs to face the previous problems. To do so, we present a new methodology called IIVFDT, which is short for *Ignorance functions* *based Interval-Valued Fuzzy Decision Tree with genetic tuning*. The IIVFDT method involves the following four steps:

(1) The induction of the base FDT using the fuzzy ID3 algorithm.[11]
(2) A new modeling of the linguistic labels of the classifier by means of IVFSs. With this aim, we define a novel construction method of IVFSs starting from the fuzzy sets used by the learning algorithm and using weak ignorance functions[27] to measure the degree of ignorance when assigning punctual values as membership degrees. The parametrization employed in the construction allows us to: (a) set the position of the initial fuzzy set within the IVFS; (b) weight the ignorance degree in order to determine the length of the IVFS.
(3) The extension of the Fuzzy Reasoning Method (FRM) exploiting the full power of IVFSs in the inference process. To do so, in every step of the FRM we make the computation using intervals and, in this manner, we take into account the ignorance degree throughout the whole process.
(4) The definition of an evolutionary tuning 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.

We must stress that after all these steps, the linguistic structure of the base FDT is not modified at all, maintaining the original interpretability of this kind of model.

The suitability of the IIVFDT method is evaluated in the framework of standard classification. Specifically, our new method is tested on 20 data-sets selected from the KEEL data-set repository[28,29] (`http://www.keel.es/dataset.php`) and it is supported by a proper statistical analysis, as suggested in the literature.[30–32] Firstly, we will determine the goodness of our methodology analysing the differences in performance achieved with respect to both the initial fuzzy ID3 algorithm with and without IVFSs. We will also compare our new approach with our previous construction scheme of IVFSs based on weak ignorance functions with adjusted parameters for modifying the ignorance degree.[27] Furthermore, we will study the behaviour of the IIVFDT method in comparison with four state-of-the-art FDTs selected in this paper, namely the simple pattern tree,[8] a look-ahead approach,[9] the FDT proposed by Janikow in Ref. 6 and a method which fuzzifies the Gini index[7] and also with the C4.5 decision tree,[33] since it is considered a very robust approach in machine learning.[34] We will also compare our new method with two proposals for learning the best FDT by means of GAs, which were defined by Kim and Ryu[35] and Chang et al.;[36] thus, we will show the goodness of our new approach when compared not only with state-of-the-art decision trees, but also with respect to genetically learnt FDTs. Finally, we will study the usefulness of the cooperation between our new approach and the genetic tuning of the lateral position of the membership functions.[37]

This paper is set out as follows. Section 2 describes the use of IVFSs for dealing with classification tasks according to our previous works and also provides the description of the fuzzy ID3 induction algorithm. Then, in Sec. 3 we introduce the IIVFDT method describing in detail each step of our new methodology for working with IVFSs in FDTs. Finally, the experimental framework along with the respective experimental analysis are presented in Secs. 4 and 5 respectively. We summarize the paper with the main concluding remarks in Sec. 6.

2. Preliminaries

In this section, we will recall some preliminary concepts of IVFSs together with the description of our previous model to work with IVFSs in linguistic fuzzy rule-based classification systems[27] (Sec. 2.1) and then, we also describe the FDT generation algorithm considered in this paper, that is, the fuzzy ID3 algorithm[11] (Sec. 2.2).

2.1. *Interval-valued fuzzy sets in classification*

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 $1_L = [1,1]$. Using the order relationship given by Xu and Yager,[38] it is easy to prove that 0_L and 1_L are the smallest and the largest element of $L([0,1])$ respectively.

Definition 1. [39,40] An interval-valued fuzzy set (IVFS) (or interval type 2 fuzzy set) A on the universe $U \neq \emptyset$ is a mapping $A_{IV} : U \rightarrow L([0,1])$, such that

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

We must point out that in this paper we will use t-representable IV t-norms without zero divisors, which will be denoted $\mathbf{T}_{T_a,T_b}$, to model conjunction operators. Furthermore, we present the interval arithmetic that we will use to to be able to extend the FRM on IVFSs.

Let $[\underline{x}, \overline{x}]$, $[\underline{y}, \overline{y}]$ be two intervals in $\mathbb{R}^+$, the rules of interval arithmetic 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: $[\min(\min(\frac{\underline{x}}{\underline{y}}, \frac{\overline{x}}{\overline{y}}), 1), \min(\max(\frac{\underline{x}}{\underline{y}}, \frac{\overline{x}}{\overline{y}}), 1)]$.

A deep study about interval-valued fuzzy logic operators can be found in Refs. 41 and 42 and about interval arithmetic in Ref. 43.

We denote by $\mathcal{L}$ the length of the interval under consideration, that is

$$\mathcal{L}(A_{IV}(u_i)) = \overline{A}(u_i) - \underline{A}(u_i).$$

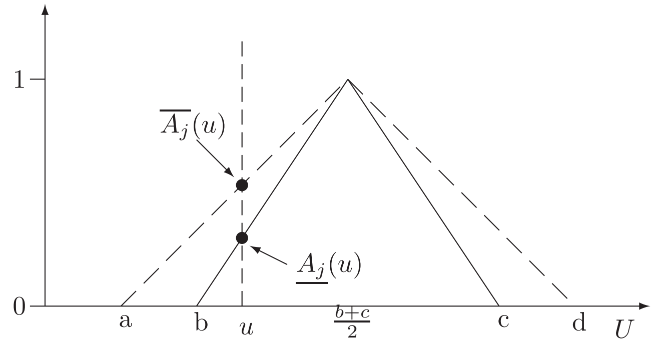


Fig. 1. Example of an IVFS considered in our previous works. The solid line is the initial fuzzy set, which in turn is the lower bound of the IVFS. The dashed line is the upper bound of the IVFS.

The length of the IVFSs can be seen as a representation of the ignorance when assigning punctual values as membership degrees.[15] In order to measure the ignorance degree, in our previous work on the topic we defined the concept of weak ignorance functions,[27] which are a particular case of ignorance functions depending on a single variable and demanding a less number of properties.

Definition 2.[27] A weak ignorance function 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.

In our previous work,[27] we constructed IVFSs starting from given fuzzy sets and applying weak ignorance functions. Then, we used the resulting IVFSs (like the one depicted in Fig. 1) in the fuzzy rule-based classification system generated by the Fuzzy Hybrid Genetics-Based Machine Learning algorithm.[44] Another construction method of intervals can be found in Ref. 45.

We must remark that the amplitude of the support of the upper bound of the IVFSs, consequently the ignorance degree that each IVFSs represents, is defined by the points a and d as it is shown in Fig. 1. We computed both points using parametrized equations based on the points defining the lower bound, that is, $a = b - W \cdot (c - b)$ and $d = c + W \cdot (c - b)$. In this way, we could vary the amount of ignorance each IVFS represents by modifying the value of the parameter W.

In fact, in order to improve the system behaviour, we proposed a tuning approach called weak ignorance tuning.[27] By means of this tuning approach we performed changes to the amplitude of the support of the upper bound of each IVFSs by varying the value of the parameter W and, consequently, we allowed the ignorance degree that each IVFSs represents to be modified. As a result, we could find the best set-up of the linguistic labels modeled by means of IVFSs and, in this way, we provided the system with a good capability of uncertainty management leading to an enhancement of the system performance.

In addition, the modeling of the linguistic labels by means of IVFSs led us to perform simple modifications to the FRM in order to work with this representation. The original FRM[46] is composed of four steps: to compute the matching degree, to compute the association degree, to compute the pattern classification soundness degree for all classes and the classification step. In our previous approach, we modified the two first steps in the following way:

- *Matching degree*: we apply a t-norm to the lower and upper bounds of the interval membership degrees of the elements to the IVFSs composing the antecedent of the rules.
- *Association degree*: we take the mean between the product of the matching degree by the RW associated with the lower bound and the product of the matching degree by the RW associated with the upper bound.

At this point we already have a single number associated with the class and we apply the remaining steps as in the original FRM. For more details about both the IVFSs construction method, the extended FRM and the weak ignorance tuning, please refer to Ref. 27.

2.2. *Fuzzy ID3 induction process*

FDTs aim at high comprehensibility, attributed to decision trees, with the gradual and graceful behaviour attributed to fuzzy systems. Thus, they extend the symbolic decision trees procedures using fuzzy sets and approximate reasoning both for the tree building and the inference mechanism. At the same time, they borrow the rich existing decision tree methodologies for dealing with incomplete or imprecise information, extended to use new wealth of information available in fuzzy representation.[6]

In the specialized literature there are many techniques to design FDTs.[6–8] Among them, one of the most widely used is the fuzzy ID3 algorithm as it is shown through its numerous variants, i.e. the ones given in Refs. 9–11 and its application to several real problems.[47–49] Furthermore, it provides a good trade-off between interpretability and accuracy with a small computation-effort.[12]

In the remainder of the section we describe in detail the fuzzy ID3 induction process, which involves two main steps, namely the selection of the expanded attribute and the FDT generation process.

Consider there are N labeled patterns and n attributes $A = A^{(1)}, \ldots, A^{(n)}$. For each $k(1 \leq k \leq n)$, the attribute $A^{(k)}$ takes m_k values of fuzzy subsets, $(A_1^{(k)}, A_2^{(k)}, \ldots, A_{m_k}^{(k)})$, $A^{(n+1)}$ denotes the classification (decision) attribute, taking m values $C_1, C_2, \ldots, C_m$. We use the symbol M($\cdot$) to denote the cardinality of a given fuzzy set, that is, the sum of all membership values of the fuzzy set.

The key to the fuzzy ID3 algorithm is to select the expanded attribute, which can be performed in the following steps:[12]

(1) For each linguistic label $A_i^{(k)}, (i = 1, 2, \ldots, m_k)$, compute its relative frequencies with respect to class $C_j, (j = 1, 2, \ldots, m)$.

$$p_i^{(k)}(j) = \frac{M(A_i^{(k)} \cap C_j)}{M(A_i^{(k)})} \,. \tag{1}$$

(2) For each linguistic label $A_i^{(k)}, (i = 1, 2, \ldots, m_k)$, compute its fuzzy classification entropy.

$$Entr_i^{(k)} = \sum_{j=1}^{m} -p_i^{(k)}(j) \log(p_i^{(k)}(j)) \,. \tag{2}$$

(3) Compute the average fuzzy classification entropy of each attribute.

$$E_k = \sum_{i=1}^{m_k} \frac{M(A_i^{(k)})}{\sum_{j=1}^{m_k} M(A_j^{(k)})} Entr_i^{(k)} \,. \tag{3}$$

(4) Select the attribute that minimizes the average classification entropy.

$$Atr = \arg\min_{1 \leq k \leq n}(E_k) \,. \tag{4}$$

Next, we briefly describe the induction based on the fuzzy ID3 algorithm.

With a given evidence significance level α, a truth level threshold β and A being the set of attributes of the problem, the induction process consists of the following steps:[11]

(1) Calculate the α-cut over the set of fuzzified patterns with the evidence significance level α.
(2) Select the attribute with the minimum average fuzzy classification entropy (Eq. (4)) as the root decision node and add the linguistic labels as candidate branches of the tree.
(3) Select one branch to analyse. Delete the branch if it is empty. If the branch is non-empty, compute the relative frequencies (Eq. (1)) of all objects within the branch into each class. If the relative frequency of one class is above the given threshold β or all the attributes have been expanded for this branch terminate the branch as a leaf. Otherwise, select the attribute, from among those which have not been expanded yet in this branch, with the smallest average fuzzy classification entropy (Eq. (4)) as a new decision node for the branch and add

its linguistic labels as candidate branches of the tree. At each leaf, each class will have its relative frequency.

(4) Repeat step 3 while there are branches to analyse. If there are no candidate branches the decision tree is complete.

3. IIVFDT: Ignorance Functions Based Interval-Valued Fuzzy Decision Tree

The aim of this section is to describe the IIVFDT method, that is, our new methodology for working with IVFSs using the fuzzy ID3 algorithm to generate the initial FDT. To do so, in first place we present our new parametrized construction method of IVFSs from given fuzzy sets. Then, we show how to model the linguistic labels by means of IVFSs making use of the previously defined construction method. The modeling of the linguistic labels by means of IVFSs implies the extension of the FRM in order to fully exploit the whole power of the application of IVFSs for this approach. To do so, we perform simple modifications to the original FRM using some of the concepts mentioned in Sec. 2.1.

Furthermore, in the initial construction of the IVFSs we consider the same ignorance degree for all IVFSs used by the system. This fact may imply a deficit on the system accuracy, since the degree of ignorance related to the definition of the different linguistic labels can vary. To deal with this problem, we define an evolutionary tuning approach in which we modify the values of the parameters of the IVFSs construction method in order to look for the best amount of ignorance that each IVFS represents.

3.1. *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 satisfies 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*

$$\begin{aligned} &G : [0,1]^4 \to L([0,1]) \text{ given by} \\ &G(x,y,\delta,\gamma) = [x \cdot (1-\delta \cdot y), x \cdot (1-\delta \cdot y) + \gamma \cdot y)] \end{aligned} \tag{5}$$

satisfies the following properties:

(1) $x \in G(x,y,\delta,\gamma)$;
(2) $W(G(x,y,\delta,\gamma)) = \gamma \cdot y$;
(3) If $x = 0$, *then* $G(0,y,\delta,\gamma) = [0, \gamma \cdot y]$;
(4) If $y = 0$, *then* $G(x,0,\delta,\gamma) = x$;

(5) If $\delta = \gamma$, then:

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

Proof. Direct. □

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 $g(\mu_A(u_i))$, $\delta(u_i)$, $\gamma(u_i) \in [0, 1]$, then the set*

$$\begin{aligned} A_{IV} &= \{(u_i, A_{IV}(u_i)) | u_i \in U\} \text{ where} \\ A_{IV}(u_i) &= G(\mu_A(u_i), g(\mu_A(u_i)), \delta(u_i), \gamma(u_i)) \end{aligned} \tag{6}$$

is an IVFS on U.

Proof. Direct. □

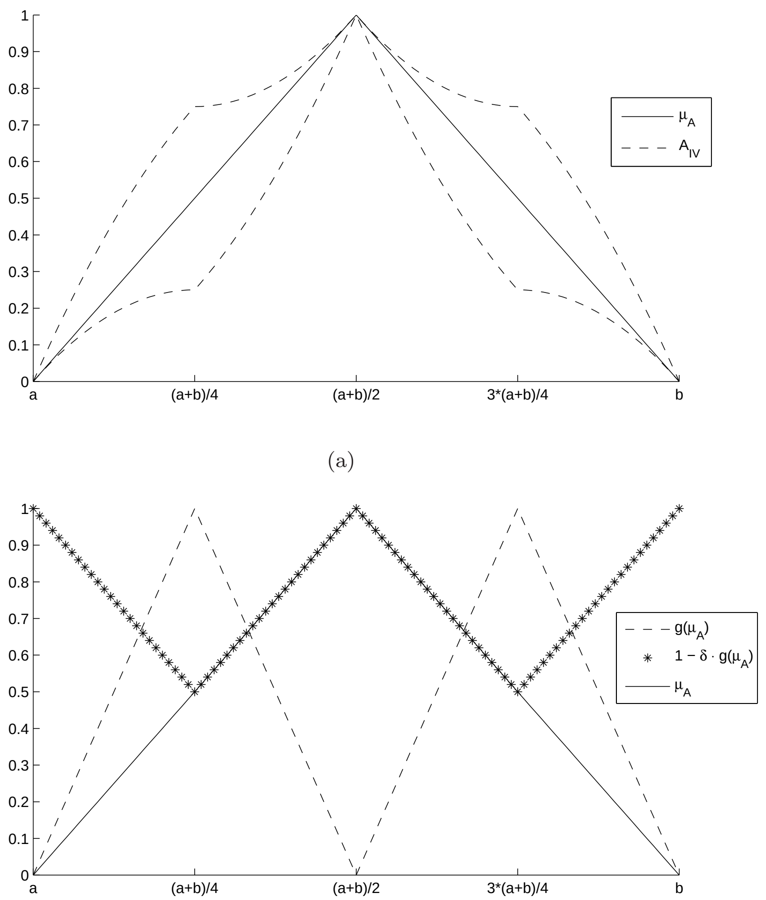
3.2. *Modeling the linguistic labels by means of interval-valued fuzzy sets*

In this section, we present how to model the linguistic by means of IVFSs. To do so, we begin describing the initial membership functions that are the starting point to apply the IVFSs construction method presented in Sec. 3.1.

We consider fuzzy sets that are represented by triangular membership functions, which are widely used in the specialized literature. Furthermore, they can be defined using only 3 points (a, $\frac{a+b}{2}$, b):

$$\mu_A(u_i) = \begin{cases} 0, & \text{if } u_i \leq a\,, \\ \dfrac{2}{b-a}(u_i - a), & \text{if } a \leq u_i \leq \dfrac{a+b}{2} \text{ and } b \neq a\,, \\ \dfrac{2}{a-b}(u_i - b), & \text{if } \dfrac{a+b}{2} \leq u_i \leq b \text{ and } b \neq a\,, \\ 0, & \text{if } b \leq u_i\,. \end{cases} \tag{7}$$

The solid line in both subfigures of Fig. 2 depicts the membership function given in Eq. (7). To construct the IVFSs, we apply Theorem 1 using the initial fuzzy sets. For the initial construction of the IVFSs, we consider the average 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. As a result, the initially constructed IVFSs are as follows: for all $u_i \in U$,



(a)

(b)

Fig. 2. (a) Solid line: initial fuzzy set; Dashed line: initial IVFS. (b) Solid line: initial fuzzy set; Dashed line: weak ignorance function; Star line: membership function weighting factor.

$$
\begin{aligned}
A_{IV}(u_i) &= G(\mu_A(u_i), g(\mu_A(u_i)), 0.5, 0.5) \\
&= [\mu_A(u_i) \cdot (1 - 0.5 \cdot g(\mu_A(u_i))), \mu_A(u_i) \cdot (1 - 0.5 \cdot g(\mu_A(u_i))) \\
&\quad + 0.5 \cdot g(\mu_A(u_i))] \,.
\end{aligned}
$$

Figure 2(a) depicts the initial construction of an IVFSs. As we can observe, both the lower and the upper bounds are represented by piecewise functions. This fact is due to the weak ignorance function (dashed line in Fig. 2(b)) returns the maximum value when the membership degree is 0.5. Therefore, it presents two maximum values (points $\frac{a+b}{4}$ and $\frac{3\cdot(a+b)}{4}$) when considering a membership function like the one given in Eq. (7). Therefore, the ignorance weighting factor, $1 - 0.5 \cdot g(\mu_A)$, presents a double minimum at the same points (star line in Fig. 2(b)). As consequence, when the ignorance weighting factor is decreasing (intervals $[a, \frac{a+b}{4}]$ and $[\frac{a+b}{2}, \frac{3\cdot(a+b)}{4}]$) the smaller the value of the ignorance weighting factor the greater the distance from the lower bound to the initial membership function is. On the other hand, when the ignorance weighting factor is increasing (intervals $[\frac{a+b}{4}, \frac{a+b}{2}]$ and $[\frac{3\cdot(a+b)}{4}, b]$) the greater the ignorance weighting factor the lower the distance from the lower bound to the initial membership function is. Similarly, we obtain the shape of the upper bound.

3.3. *Fuzzy decision trees with interval-valued fuzzy sets: a new fuzzy reasoning method*

The modeling of the linguistic labels by means of IVFSs in FDTs implies that the relative frequencies of the classes in each leaf must be recalculated. In order to do this, we follow the procedure explained in the FDT induction (Sec. 2.2) and we apply Eq. (1) using the interval arithmetic described in Sec. 2.1. In this manner, the relative frequencies of the classes will be elements of $L([0,1])$.

The use of IVFSs also imply that we need to modify the FRM to predict the classification of new examples.

For an FDT, each connection from the root node to a leaf is called a path. It is clear that each path corresponds to a different leaf, so there are as many paths as leaves in the FDT.

Suppose that the FDT contains l leaves ($Path^i, i = 1, 2, \ldots, l$) and 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 point out 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 $k = 1, \ldots, N_i$. The *interval-valued fuzzy reasoning mechanism* follows the following 4 key steps:

(1) 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 the corresponding path.

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

(2) 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.

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

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

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

(4) Classify the example in the class which maximizes the total certainty.

In order to decide the maximum interval, which is necessary to perform the last step of the FRM presented above, we will use the following relationship based on the score and accuracy functions given in Ref. 38 (see Subsec. 2.1). For any interval $[\underline{x}, \overline{x}]$, $[\underline{y}, \overline{y}]$ on $\mathbb{R}$ and 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 $[\underline{x}, \overline{x}] < [\underline{y}, \overline{y}]$.

3.4. *Genetic tuning of the ignorance weighting factor*

The definition of membership functions is usually performed homogeneously over the input space or by means of expert knowledge. In both cases, there can be some unknown amount of ignorance when assigning punctual values as membership degrees. In the former, this ignorance degree can be due to the ad-hoc construction of fuzzy partitions while in the latter it is associated with the possible lack of information suffered by the expert. As we have pointed out, we use IVFSs to deal with this problem due to their length can be seen as a representation of the ignorance degree. In the initial construction of all the IVFSs (Sec. 3.2) we have considered the average degree of ignorance, that is, we initially fix the values of δ and γ to 0.5. However, the ignorance degree can vary depending on the linguistic label because the available information can differ.

In order to look for a good management of the semantic uncertainties of the classifier, we propose the application of an evolutionary tuning step (afterwards the generation of the initial FDT with IVFSs) in which we adapt the parameters δ and γ (see Theorem 1) keeping the restriction $\delta \geq \gamma \geq \delta \cdot x$. As a result, the ignorance degree associated with the definition of each fuzzy set will be weighted depending on the suitability of the membership function to the specific problem we are dealing with, which can lead to an improvement of the system accuracy. An example of the behaviour of this evolutionary tuning approach is depicted in Fig. 3 where the final IVFS (dark gray IVFS) is embedded in the initial one (light gray IVFS) since both δ and γ are lower than the initial ones.

To accomplish this tuning process we follow the CHC evolutionary algorithm,[50] which has provided good results in this topic,[37,51] with the same scheme described

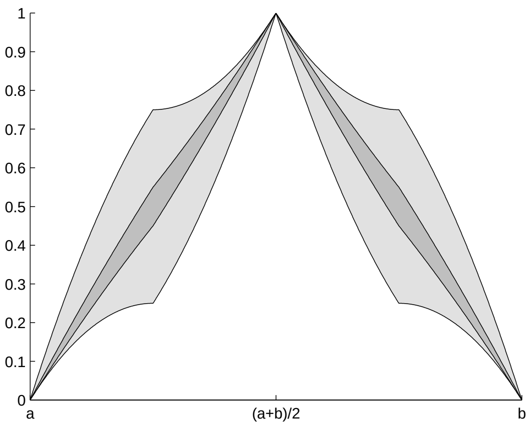


Fig. 3. Genetic tuning of the ignorance weighting factor. The final values of the parameters are $\delta = 0.1$ and $\gamma = 0.1$.

in our previous work.[27] In the remainder of this section, we present the specific features of our new evolutionary tuning approach, which involves the specification of the representation of the solutions, the definition of the fitness function and the initialization of the population of solutions.

(1) *Representation*: We consider a real coding scheme, where each pair of genes, $(\delta, \gamma) \in [0, 1]$, represents the modification of the parameters used to weight the ignorance degree when assigning punctual values as membership degrees. The form of the chromosome is:

$$C_{IWF} = (\delta_{11}, \gamma_{11}, \ldots, \delta_{1m^1}, \gamma_{1m^1}, \delta_{21}, \gamma_{21}, \ldots, \delta_{2m^2}, \gamma_{2m^2}, \ldots, \delta_{n1}, \gamma_{n1}, \ldots, \delta_{nm^n}, \gamma_{nm^n}),$$

being $(m^1, m^2, \ldots, m^n)$ the number of labels per variable and n the number of variables. Therefore, the chromosome length is twice the number of labels times the number of variables.

(2) *Fitness function*: We employ the most common metric for classification, i.e. the classification rate.

(3) *Initial Gene Pool*: we initialize the first individual having all the genes with a value of 0.5 (the values considered for the initial construction of all the IVFSs). The second and third individuals have all genes with values of 0 and 1 respectively, whereas the remaining individuals are randomly generated in [0, 1].

For details about the remainder features of the optimization process, please refer to Ref. 27.

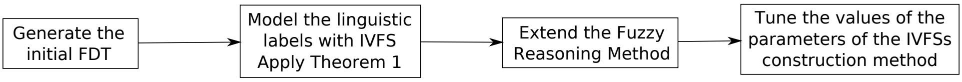


Fig. 4. Flow diagram of the IIVFDT method.

3.5. *Summarizing the IIVFDT method*

For the sake of clarity, we summarize the IIVFDT method by means of a flow diagram (Fig. 4). We can observe that the whole process is composed of four steps.

The first step consists on inducing the initial FDT from which to apply our approach. As pointed out, we have designed the IIVFDT method using the fuzzy ID3 algorithm to accomplish the learning process (Sec. 2.2). To carry out the learning process we pre-fuzzify the data using triangular membership functions, which are obtained by performing a homogeneous partition of the input space and whose expressions are like the one described in Eq. (7).

In the second step, we model the linguistic labels by means of IVFSs as explained in Sec. 3.2. To this aim, we apply the new parametrized construction method given in Theorem 1 to the membership functions (defined in the first step), which are used in the learning process. Consequently, the interpretability of the initial FDT is not modified, since both the number of rules and the linguistic terms in each of them are the same ones than those of the initial FDT.

The modeling of the linguistic labels by means of IVFSs implies the extension of the FRM on IVFSs, leading to the new interval-valued fuzzy reasoning method introduced in Sec. 3.3 (third step). In the last step, we attempt to further improve the performance of the fuzzy classifier. In order to achieve this goal, we apply the new evolutionary tuning approach in which we look for the shape of each IVFS that best represents the ignorance degree for each IVFS (Sec. 3.4).

4. Experimental Framework

In this section, we firstly present the real world classification data-sets selected for the experimental study. Next, we briefly describe the different FDTs that we will use in the experimental analysis and we also provide the values assigned to the FTDs' parameters. Finally, we introduce the statistical tests carried out in order to compare the results achieved throughout the experimental study.

4.1. *Data-sets*

We have selected a wide benchmark of 20 numerical data-sets selected from the KEEL data-set repository,[28,29] which are publicly available on the corresponding web page[a] including general information about them, partitions for the validation of the experimental results and so on. Table 1 summarizes the properties of the

[a]`http://www.keel.es/dataset.php`

Table 1. 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	1,484	8	10

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.

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, in order to avoid failed convergences of the evolutionary tuning, the process was repeated 3 times for each partition, using three different seeds, implying the achievement of a sample of 15 results which have been averaged to obtain the mean accuracy for each data-set.

4.2. *Fuzzy decision trees for comparison*

In this paper we have selected six FDTs (four approaches without using GAs and the remainder two ones using GAs) in order to compare our methodology with respect to different methods in the literature. Their descriptions are as follows:

- The first approach is the *fuzzy Gini index*.[7] This method uses the SLIQ decision tree as the base algorithm, which uses the Gini index as the split measure, and fuzzifyies its decision boundaries. Unlike the remaining FDTs considered in this paper, this method fuzzifies the decision boundaries, depending on the standard deviation of the attributes, during the decision tree construction instead of using pre-fuzzified data.
- The second approach is the *look-ahead method*.[9] This method attempts to establish a decision node by analysing the classifiability of instances that are split along branches of the node. Dong and Kothari propose the evaluation of the classifiability by means of a co-occurrence matrix. Finally, they select the node that optimizes an objective function which considers both a split measure and classifiability.
- The third proposal is the *Janikow's FDT*.[6] This FDT imposes the fuzzy sets defining the fuzzy terms used for building the tree and uses a splitting criteria based on fuzzy restrictions. Janikow adapts norms used in fuzzy logic to deal with conjunctions of fuzzy propositions in order to compute the number of examples falling in a node.
- The fourth approach we have selected is the *simple pattern tree* algorithm.[8] This method, instead of constructing one tree, whose leaves have a probability distribution expressing the membership of each class, constructs one tree per class. In order to do so, small pattern trees are aggregated to complex ones taking into account the similarity between the tree and the class represented by that tree.
- The first approach for optimizing the generation of FDTs that we have selected was defined by Kim and Ryu.[35] In this case, they consider triangular membership functions and they induce the best possible FDT by learning the most suitable fuzzy partition for each variable by means of a GA. To this aim, they apply a classic genetic approach in which they modify the values of the three points which define each triangular membership function.
- The second approach we have considered to generate FDTs using GAs was proposed by Chang *et al.*[36] In this proposal, authors use the fuzzy ID3 induction process and they use a GA to learn the best FDT by optimizing both the mean and the standard deviation of Gaussian membership functions and also to select the most appropriate thresholds for the two stopping criteria that they consider. Then, they carry out a pruning step and finally, they tune the membership functions' parameters using again a GA.

In Table 2 there are specified the configurations for the previously described FDTs without GAs together with the configuration for the fuzzy ID3 with and without IVFSs. These configurations have been fitted experimentally in the terms recommended by authors, since they make up the configuration with the best performance for each FDT.

Table 2. 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
IIVFDT method without the tuning step
Ignorance function: $g(x) = 2 \cdot \min(x, 1 - x)$ Conjunction operator: product IV t-norm $\delta = 0.5$ $\gamma = 0.5$
Fuzzy Gini Index
Number of labels per variable: 3 labels Threshold = 0.85 Number of examples threshold = 0.05 Maximum depth: number of attributes
Look-Ahead Method
Number of labels per variable: 3 labels Split measure: information gain Number of neighbours: 3 Weighting factor: 1
Janikow's FDT
Number of labels per variable: 3 labels Conjunction operator: product t-norm
Simple Pattern Tree
Number of labels per variable: 3 labels Aggregation functions: min and max operators Similarity function: the similarity related with the root mean square error

The configuration of the FDTs which are learnt using GAs is introduced in Table 3. In this case, we have used the configuration suggested by Kim and Ryu in Ref. 35, since it provides a solution in a feasible amount of time.

Finally, we indicate the values that have been considered for the parameters of the evolutionary tuning of our IIVFDT proposal:

- Population Size: 50 individuals.
- Number of evaluations: 5,000 · d.
- Bits per gene for the Gray codification (for incest prevention): 30 bits.

where d stands for the dimensionality of the problem (number of variables).

4.3. *Statistical tests for performance comparison*

In this paper, we use some hypothesis validation techniques in order to give statistical support to the analysis of the results.[52,53] We will use non-parametric

Table 3. Parameter specification for the FDTs optimized with GAs.

Kim and Ryu's GAFDT
Number of labels per variable: 3 labels Conjunction operator: minimum t-norm Truth level threshold = 1 Number of examples threshold = 0.02 w = 0.98 Number of generations = 100 Population size = 20 Crossover probability = 0.9 Mutation probability = 0.1
Chang's GAFDT
Number of labels per variable: 3 labels Conjunction operator: product t-norm Number of generations = 100 Population size = 20 Crossover probability = 0.9 Mutation probability = 0.1

tests because the initial conditions that guarantee the reliability of the parametric tests cannot be fulfilled, which imply that the statistical analysis loses credibility with these parametric tests.[30]

Specifically, we employ the Wilcoxon rank test[54] as a non-parametric statistical procedure for making pairwise comparisons between two algorithms. For multiple comparisons, we use the Friedman aligned ranks test[55] to detect statistical differences among a group of results and the Holm post-hoc test[56] to find the algorithms that reject the equality hypothesis with respect to a selected control method.

The post-hoc procedure allows us to know whether a hypothesis of comparison of means could be rejected at a specified level of significance α. Furthermore, we compute the adjusted p-value (APV) in order to take into account that multiple tests are conducted. In this manner, we can compare directly the APV with respect to the level of significance α in order to be able to reject the null hypothesis.

In addition, we consider the method of aligned ranks of the algorithms in order to show graphically how good a method is with respect to its partners. The first step to compute this ranking is to obtain the average performance of the algorithms in each data set. Next, we compute the subtractions between the accuracy of each algorithm minus the average value for each data-set. Then, we rank all these differences in a descending way and, finally, we average the rankings obtained by each algorithm. In this manner, the algorithm which achieves the lowest average ranking is the best one.

These tests are suggested in the studies presented in Refs. 30–32 and 52, where it is recommended their use in the field of machine learning. A complete description of these tests and software for their use can be found on the website: http://sci2s.ugr.es/sicidm/.

5. Analyzing the Usefulness of the IIVFDT Method

In order to show the suitability of the IIVFDT method, we have divided our study in this way:

(1) We analyse the capacity for enhancement of our new approach not only with respect to the basic fuzzy ID3 algorithm with and without IVFSs but also with our previous tuning approach, that is, the tuning of the weak ignorance[27] (Sec. 5.1).
(2) We aim to show the goodness of the IIVFDT method through its comparison with four state-of-the-art FDTs and C4.5 (Sec. 5.2).
(3) We determine the quality of our new methodology through its comparison with two proposals of generation of FDTs using GAs (Sec. 5.3).
(4) We study whether the use of the genetic tuning of the lateral position[37] afterwards the application of our new methodology allows the results provided by the latter to be outperformed (Sec. 5.4).

This experimental study is carried out in the following four sections.

5.1. *Study of the behaviour of the IIVFDT method*

In this section, we analyse the suitability of the IIVFDT proposal. To do so, we show empirically whether our methodology enhances the results of both the initial fuzzy ID3 algorithm and the IIVFDT method without the tuning step. Furthermore, we will compare the results of our new approach versus the results achieved by our previous proposal for tuning the weak ignorance.[27]

Table 4 shows the classification accuracy of the different approaches applied to the fuzzy ID3 algorithm, specifically:

- F-ID3: the standard fuzzy ID3 algorithm without IVFSs.
- F-ID3_IVFS_WI: our previous approach for tuning the weak ignorance degree[27] applied to the fuzzy ID3 algorithm, that is, its linguistic labels are modeled using triangular shaped IVFSs, whose ignorance degrees are genetically optimized.
- IIVFDT-0.5: the IIVFDT method without the tuning step, that is, with $\delta = \gamma = 0.5$.
- IIVFDT: the complete approach proposed in this paper.

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

We observe from the results of Table 4 the good behaviour of the IIVFDT method, since it enhances the performance of both the initial fuzzy ID3 algorithm and the IIVFDT proposal without the tuning step. Furthermore, our methodology

Table 4. Results in Train (Tr.) and Test (Tst) achieved by the different approaches applied to the fuzzy ID3 algorithm.

Data Set	F-ID3		F-ID3_IVFS_WI		IIVFDT-0.5		IIVFDT	
	Tr.	*Tst*	*Tr.*	*Tst*	*Tr.*	*Tst*	*Tr.*	*Tst*
App	90.57	84.94	91.98	85.84	88.45	**86.84**	93.16	86.80
Bal	91.72	90.08	93.24	90.56	92.08	90.08	96.60	**90.72**
Bup	61.52	58.26	67.10	58.84	60.00	58.55	79.93	**67.25**
Cle	87.29	53.87	92.09	55.55	87.21	**56.55**	93.18	55.88
Eco	80.21	76.80	81.62	77.09	77.83	75.59	82.14	**77.69**
Gla	60.17	52.80	75.24	**62.66**	66.02	58.86	77.57	60.80
Hab	75.08	71.90	77.45	71.89	74.26	**72.88**	79.90	72.87
Hea	93.52	79.26	95.09	78.89	91.02	**78.89**	94.81	78.52
Iri	93.07	95.33	98.00	95.33	93.17	92.67	98.83	**96.00**
Mag	95.50	78.28	83.10	**79.60**	79.46	78.76	82.75	78.91
New	79.80	91.16	98.26	**95.81**	90.35	90.23	97.79	94.88
Pag	92.67	92.15	93.57	93.24	92.75	92.15	95.03	**94.16**
Pim	92.84	75.77	82.36	76.30	78.87	75.65	83.56	**76.43**
Rin	79.95	49.59	49.73	49.59	52.13	51.76	97.09	**90.81**
Shu	98.74	90.57	93.37	93.20	83.45	83.31	97.74	**97.98**
Tae	63.40	54.99	69.87	55.63	64.74	53.61	70.86	**58.95**
Tit	78.33	**78.33**	78.33	**78.33**	78.33	**78.33**	78.33	**78.33**
Win	99.59	97.75	100.00	97.75	98.88	96.60	100.00	**98.87**
Wis	90.63	94.58	98.32	**96.63**	96.71	94.73	98.28	96.04
Yea	95.90	55.46	62.62	56.87	49.61	47.98	63.01	**57.41**
Mean	81.02	76.09	84.07	77.48	79.76	75.70	88.03	**80.46**

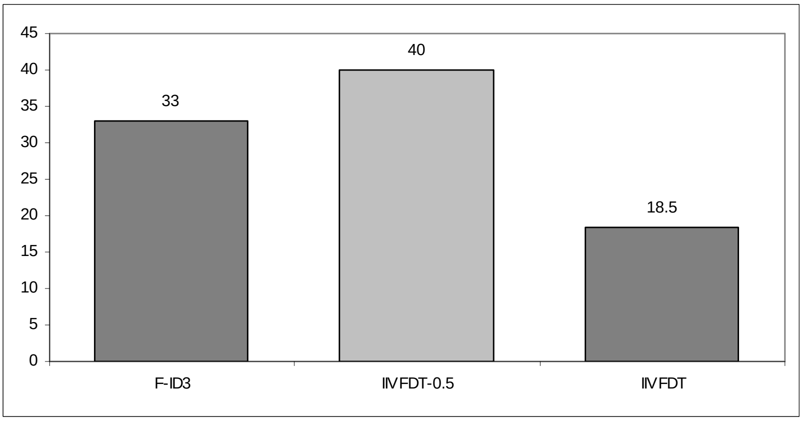


Fig. 5. Rankings of the approaches with the fuzzy ID3 as base algorithm.

provides the best mean value in test and achieves the best performance in most of the data-sets considered in the study. This situation is confirmed in Fig. 5 where it is shown that the best ranking is reached by the IIVFDT approach.

Table 5. Holm test to compare three methodologies with the fuzzy ID3 as base algorithm. The IIVFDT method is used as the control method.

i	Algorithm	APV	Hypothesis
2	IIVFDT-0.5	1.30E-5	Rejected for IIVFDT
1	F-ID3	3.55E-5	Rejected for IIVFDT

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 (7.73E-4), 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 approaches. Specifically, a Holm test is applied, which is presented in Table 5. The statistical analysis reflects that the IIVFDT method outperforms the two remainder approaches with a high level of confidence.

In order to strengthen to goodness of our new methodology, we compare the results provided by the IIVFDT method with respect to the ones obtained afterwards the application of our previous approach with IVFSs and tuning of the weak ignorance degree[27] using the fuzzy ID3 algorithm to construct the initial fuzzy system. We can observe from results on Table 4 that our new proposal enhances notably the mean test result of our previous model. The statistical analysis, which is carried out by means of a Wilcoxon test (Table 6), clearly reflects the superiority of our new methodology with a low p-value.

Table 6. Wilcoxon Test to compare the fuzzy ID3 algorithm with IVFS and the weak ignorance tuning model (R^+) against the IIVFDT method (R^-).

Comparison	R^+	R^-	Hypothesis	p-value
F-ID3_IVFS_WI vs. IIVFDT	50.5	159.5	Rejected for IIVFDT	0.044

5.2. *Comparing the IIVFDT method versus state-of-the-art decision trees*

We present in Table 7 the results achieved in training and test provided by our methodology, along with the four FDTs considered and C4.5. For the sake of clarity we present the notation of each approach:

- 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.

Table 7. Results in Train (Tr.) and Test (Tst) achieved by the different decision trees.

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

- C4.5: the well known decision tree.
- IIVFDT: the complete approach proposed in this paper.

The best global result for each data-set is highlighted in **bold-face**.

From the results of Table 7, the capacity of our methodology for improvement with respect to the results obtained by the decision trees is clearly shown, as it achieves the best global performance and provides the best performance in half of the data sets. This situation is confirmed in Fig. 6, where the rankings of the different approaches are presented, showing that the best ranking is achieved by our new proposal.

In order to strengthen the previous findings we apply a Friedman aligned rank test. The p-value is 0.004 which implies that there are statistical differences among the studied approaches with a high level of significance. In this manner, we apply a Holm post-hoc test (Table 8) in order to compare our methodology with the remaining approaches. The statistical analysis shows that all of the FDTs, which have been considered in this study, and also the C4.5 decision tree are notably enhanced by the IIVFDT method.

We must point out that the initial fuzzy ID3 algorithm does not outperform any of the approaches for comparison. In this manner, it is clearly shown the robustness

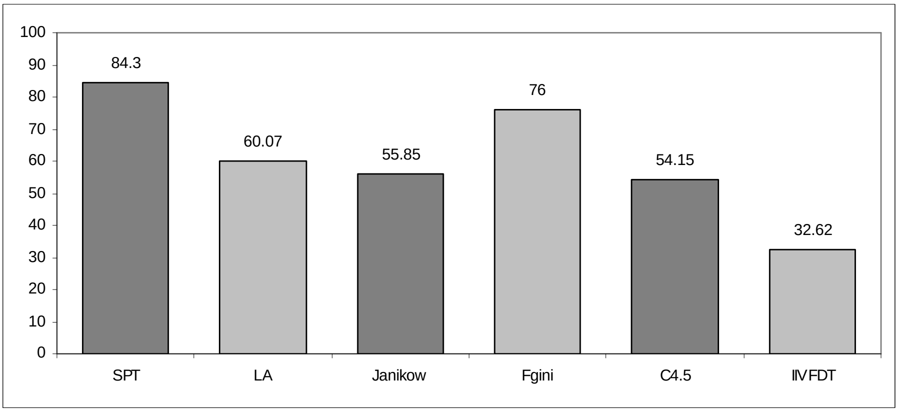


Fig. 6. Rankings of the different fuzzy decision trees and C4.5.

Table 8. Holm test to compare the IIVFDT proposal (used as control method) with all the remaining decision trees.

i	Algorithm	APV	Hypothesis
4	SPT	1.31E-5	Rejected for IIVFDT
3	Fgini	3.21E-4	Rejected for IIVFDT
2	LA	0.037	Rejected for IIVFDT
1	Janikow	0.068	Rejected for IIVFDT
1	C4.5	0.068	Rejected for IIVFDT

of our new method, since it allows the initial results provided by this algorithm to be enhanced in such a way that they outperform the results provided by several state-of-the-art FDTs and also the ones obtained by C4.5.

5.3. *On the comparison with fuzzy decision trees constructed using genetic algorithms*

Table 9 shows the results achieved by the different approaches that use GAs, both in training and in test in each data-set. In first place, there are presented the results of the methods for generating FDTs with GAs defined by Kim[35] (GAFDT_Kim) and Chang[36] (GAFDT_Chang) and then, there are shown the results obtained when applying our methodology. The best global result for each data-set is stressed in **bold-face**.

From the results of Table 9, we must highlight both the average improvement of our proposal with respect to the remainder ones and the achievement of the best result in eleven out of twenty data-sets. This situation is confirmed in Fig. 7, where the rankings of the three approaches are presented, showing that the best ranking is provided by our evolutionary tuning method.

Table 9. Results in Train (Tr.) and Test (Tst) achieved by the IIVFDT method and the two proposals of FDT generation using GAs.

Data Set	GAFDT_Kim		GAFDT_Chang		IIVFDT	
	Tr.	*Tst*	*Tr.*	*Tst*	*Tr.*	*Tst*
App	91.03	85.84	92.92	84.94	93.16	**86.80**
Bal	84.80	82.72	86.76	86.40	96.60	**90.72**
Bup	67.97	62.03	71.23	62.90	79.93	**67.25**
Cle	62.21	**57.59**	84.43	53.53	93.18	55.88
Eco	89.14	74.13	90.40	73.81	82.14	**77.69**
Gla	67.41	61.22	73.25	**62.65**	77.57	60.80
Hab	78.02	72.22	78.02	**72.89**	79.90	72.87
Hea	85.09	76.30	92.96	**81.48**	94.81	78.52
Iri	98.67	**96.67**	97.67	92.67	98.83	96.00
Mag	78.89	77.65	81.69	**79.65**	82.75	78.91
New	95.35	**95.35**	97.33	**95.35**	97.79	94.88
Pag	93.52	93.06	94.39	93.61	95.03	**94.16**
Pim	77.08	73.83	79.33	75.13	83.56	**76.43**
Rin	83.34	81.62	88.61	85.81	97.09	**90.81**
Shu	94.56	94.34	85.95	85.47	97.74	**97.98**
Tae	65.73	49.74	67.72	56.34	70.86	**58.95**
Tit	79.05	**79.06**	78.90	78.33	78.33	78.33
Win	96.49	90.40	97.89	95.49	100.00	**98.87**
Wis	97.11	95.90	97.99	**96.49**	98.28	96.04
Yea	49.88	48.38	43.48	44.95	63.01	**57.41**
Mean	81.77	77.40	84.05	77.89	88.03	**80.46**

We have used the Friedman aligned ranks test in order to find out whether significant differences exist among all the mean values. This test obtains a p-value near to zero (5.64E-4), which implies that there are significant differences between the results. We now apply Holm's test to compare the best ranking method (IIVFDT approach) with the remaining methods. Table 10 presents these results. In this table, the methods are ordered with respect to the APV obtained. Holm's test rejects the hypothesis of equality with the rest of the methods.

Table 10. Holm test to compare the IIVFDT proposal (used as control method) with the FDTs using genetic algorithms.

i	Algorithm	APV	Hypothesis
2	GAFDT_Kim	2.57E-4	Rejected for IIVFDT
1	GAFDT_Chang	0.005	Rejected for IIVFDT

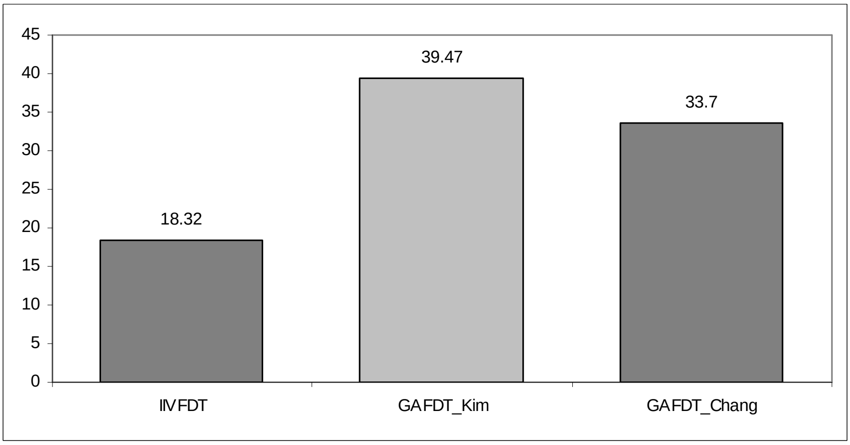


Fig. 7. Rankings of the different proposals using GAs.

Therefore, analyzing the results presented in Table 9 and the statistical study shown in Table 10 we can conclude that our new method outperforms the performance obtained with two recent proposals for constructing FDTs in an optimal way by means of the use of GAs. In this manner, it is strengthened the synergy produced when combining IVFSs with the evolutionary tuning step to face classification problems.

5.4. *Using the tuning of the lateral position of the membership functions*

The genetic tuning of the lateral position of the linguistic labels[37] has proved to provide very accurate models.[51,57,58] This tuning approach, which is based on the linguistic 2-tuples representation,[59] allows the lateral displacement of the labels considering only one parameter (slight displacements to the left/right of the original membership functions). Therefore, it seems natural to extend our proposal by applying the genetic tuning of the lateral position afterwards the application of our new methodology.

Table 11 shows the mean results in training and testing achieved by both the IIVFDT proposal and the sequential application of our new methodology and the genetic tuning of the lateral position (IIVFDT+Lat). The best result is highlighted in **bold-face**.

From the results in Table 11, it is observed that the synergy between both approaches allows to achieve a higher classification accuracy in most of the data-sets of the study. In order to compare both methods, we apply a Wilcoxon test (Table 12). The statistical analysis allows us to asseverate with a high level of confidence that the sequential application of both approaches allows to improve the results obtained by our new methodology. This enhancement is due to the lateral tuning faces the possible lack of adaptation of the membership functions to the the

Table 11. Results in Train (Tr.) and Test (Tst) achieved by the IIVFDT method with and without lateral tuning.

Data Set	IIVFDT		IIVFDT+Lat	
	Tr.	*Tst*	*Tr.*	*Tst*
App	93.16	86.80	94.58	**87.75**
Bal	96.60	90.72	96.62	**92.00**
Bup	79.93	67.25	80.65	**72.46**
Cle	93.18	**55.88**	96.04	54.55
Eco	82.14	**77.69**	83.18	77.40
Gla	77.57	60.80	80.14	**64.04**
Hab	79.90	72.87	80.31	**73.19**
Hea	94.81	**78.52**	98.33	**78.52**
Iri	98.83	**96.00**	99.83	**96.00**
Mag	82.75	78.91	84.83	**80.39**
New	97.79	94.88	99.77	**96.28**
Pag	95.03	**94.16**	95.30	93.97
Pim	83.56	**76.43**	84.18	76.42
Rin	97.09	90.81	99.29	**96.22**
Shu	97.74	97.98	99.31	**99.13**
Tae	70.86	58.95	74.34	**60.26**
Tit	78.33	78.33	79.07	**78.87**
Win	100.00	**98.87**	100.00	97.73
Wis	98.28	**96.04**	98.98	95.75
Yea	63.01	57.41	65.43	**59.50**
Mean	88.03	80.46	89.51	**81.49**

Table 12. Wilcoxon Test to compare the the IIVFDT method (R^+) versus its cooperation with the tuning of the lateral position (R^-).

Comparison	R^+	R^-	Hypothesis	p-value
IIVFDT vs. IIVFDT+Lat	43.5	166.5	Rejected for IIVFDT+Lat	0.016

context of each variable and as consequence, it improves the system accuracy by properly suiting the linguistic labels to each specific variable of the problem.

6. Concluding Remarks

In this paper we have presented the IIVFDT method, a proposal to improve the performance of FDTs using IVFSs. In order to do so, we have developed a new IVFS construction method based on weak ignorance functions which starts from the fuzzy sets used in the induction process of the initial FDT. The final shape of the IVFSs is set by two parameters which weight the degree of ignorance related to the assignment of punctual values as membership degrees.

We have extended the FRM performing natural modifications in order to be able to work with interval-valued membership functions. Furthermore, we have introduced an evolutionary tuning approach in which we optimize the parameters of the construction method of IVFSs. In this manner, we compute the best degree of ignorance that each IVFS represents, leading to an enhancement of the system performance.

We have developed the IIVFDT method using the fuzzy ID3 algorithm in order to generate the initial FDT with which to apply our new methodology. Along the experimental study, we have reached several lessons learned:

(1) The IIVFDT method allows to improve the results of the initial FDT with and without IVFSs and also the results provided by our previous methodology applied to FDTs.
(2) Our new methodology enhances the behaviour of some state-of-the-art FDTs.
(3) The accuracy results of C4.5, a well-known machine learning algorithm, which is considered in order to strengthen our results, are outperformed by our approach.
(4) Our new approach notably enhances the results obtained with two FDTs that have been learnt using GAs and, in this way, it is stressed the quality of our new evolutionary tuning proposal.
(5) We achieve a positive synergy when applying sequentially the IIVFDT method and the lateral tuning.

These results allow us to conclude that our new methodology is a suitable solution to confront classification problems dealing with the ignorance degree when assigning punctual values as membership degrees and a fine evolutionary tuning of the fuzzy partitions.

Acknowledgements

This work was supported in part by the Spanish Ministry of Science and Technology under projects TIN2011-28488 and TIN2010-15055.

References

1. R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification*, 2nd edn. (John Wiley, 2001).
2. J. Han and M. Kamber, Data Mining — Concepts and Technique, in *The Morgan Kaufmann Series in Data Management Systems*, 2nd edn. (Morgan-Kaufmann, San Mateo, CA, 2006).
3. L. A. Zadeh, Fuzzy logic and approximate reasoning, *Fuzzy Sets Syst.* **99**(3) (1975) 283–290.
4. L. A. Zadeh, Fuzzy sets, *Information and Control* **8**(3) (1965) 338–353.
5. R. Kothari and M. Dong, *Decision Trees for Classification: A Review and Some New Results* (World Scientific, Singapore, 2000).
6. C. Z. Janikow, Fuzzy decision trees: issues and methods, *IEEE Trans. Syst. Man Cybern. B* **28**(1) (1998) 1–14.

7. B. Chandra and P. P. Varghese, Fuzzifying gini index based decision trees, *Expert Syst. Appl.* **36**(4) (2009) 8549–8559.
8. Z. Huang, T. Gedeon and M. Nikravesh, Pattern trees induction: A new machine learning method, *IEEE Trans. Fuzzy Syst.* **16**(4) (2008) 958–970.
9. M. Dong and R. Kothari, Look-ahead based fuzzy decision tree induction, *IEEE Trans. Fuzzy Syst.* **9**(3) (2001) 461–468.
10. H. Ichihashi, T. Shirai, K. Nagasaka and T. Miyoshi, Neuro-fuzzy ID3: a method of inducing fuzzy decision trees with linear programming for maximizing entropy and an algebraic method for incremental learning, *Fuzzy Sets Syst.* **81**(1) (1996) 157–167.
11. Y. Yuan and M. J. Shaw, Induction of fuzzy decision trees, *Fuzzy Sets Syst.* **69**(2) (1995) 125–139.
12. X.-Z. Wang, D. S. Yeung and E. C. C. Tsang, A comparative study on heuristic algorithms for generating fuzzy decision trees, *IEEE Trans. Syst. Man Cybern. B* **31**(2) (2001) 215–226.
13. R. Sambuc, Function ϕ-flous, application a l'aide au diagnostic en pathologie thyroidienne, Ph.D. thesis, University of Marseille, 1975.
14. 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 Syst.* **161**(1) (2010) 20–36.
15. D. Dubois and H. Prade, An introduction to bipolar representations of information and preference, *Int. J. Intell. Syst.* **23**(8) (2008) 866–877.
16. D. Dubois and H. Prade, *Possibility Theory* (Plenum Press, New York, 1988).
17. 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, *Inform. Sciences* **180**(19) (2010) 3674–3685.
18. M. Galar, J. Fernandez, G. Beliakov and H. Bustince, Interval-valued fuzzy sets applied to stereo matching of color images, *IEEE Trans. Image Process.* **20**(7) (2011) 1949 – 1961.
19. H. Bustince, E. Barrenechea, M. Pagola, J. Fernandez and J. Sanz, Comment on: Image thresholding using type II fuzzy sets. Importance of this method, *Pattern Recogn.* **43**(9) (2010) 3188–3192.
20. T.-Y. Chen, Optimistic and pessimistic decision making with dissonance reduction using interval-valued fuzzy sets, *Inform. Sciences* **181**(3) (2011) 479–502.
21. J. M. Mendel, Computing with words and its relationships with fuzzistics, *Inform. Sciences* **177**(4) (2007) 988–1006.
22. O. Cordón, F. Gomide, F. Herrera, F. Hoffmann, and L. Magdalena, Ten years of genetic fuzzy systems: Current framework and new trends, *Fuzzy Sets Syst.* **141**(1) (2004) 5–31.
23. C. b. Das and S. K. Roy, Fuzzy based GA for entropy bimatrix goal game, *Int. J. Uncertain. Fuzz.* **18**(8) (2010) 779–799.
24. D. G. Stavrakoudis, J. B. Theocharis and G. C. Zalidis, A multistage genetic fuzzy classifier for land cover classification from satellite imagery, *Soft Comput.* **15**(12) (2011) 2355–2374.
25. F. Chávez, F. Fernández, M. J. Gacto and R. Alcalá, Automatic laser pointer detection algorithm for environment control device systems based on template matching and genetic tuning of fuzzy rule-based systems, *Int. J. Comput. Int. Syst.* **5**(2) (2012) 368–386.
26. C. Karr, Genetic algorithms for fuzzy controllers, *AI Expert* **6**(2) (1991) 26–33.
27. 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, *Int. J. Approx. Reason.* **52**(6) (2011) 751–766.
28. J. Alcalá-Fdez, L. Sánchez, S. García, M. J. del Jesus, S. Ventura, J. Garrell, J. Otero, C. Romero, J. Bacardit, V. Rivas, J. C. Fernández and F. Herrera, KEEL: A software tool to assess evolutionary algorithms for data mining problems, *Soft Comput.* **13**(3) (2009) 307–318.
29. 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, *J. Mult.-Valued Log. Soft Comput.* **17**(2–3) (2011) 255–287.
30. J. Demšar, Statistical comparisons of classifiers over multiple data sets, *J. Mach. Learn. Res.* **7** (2006) 1–30.
31. S. García and F. Herrera, An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons, *J. Mach. Learn. Res.* **9** (2008) 2677–2694.
32. S. García, A. Fernández, J. Luengo and F. Herrera, Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power, *Inform. Sciences* **180**(10) (2010) 2044–2064.
33. J. Quinlan, *C4.5: Programs for Machine Learning* (Morgan Kaufmann Publishers, San Mateo, California, 1993).
34. X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. McLachlan, A. Ng, B. Liu, P. Yu, Z.-H. Zhou, M. Steinbach, D. Hand and D. Steinberg, Top 10 algorithms in data mining, *Knowl. Inf. Syst.* **14** (2008) 1–37.
35. M. Kim and J. Ryu, Optimized fuzzy decision tree using genetic algorithm in *Neural Information Processing*, Lecture Notes in Computer Science, Vol. 4234 (Springer, Berlin/Heidelberg, 2006), pp. 797–806.
36. J.-Y. Chang, C.-W. Cho, S.-H. Hsieh and S.-T. Chen, Genetic algorithm based fuzzy ID3 algorithm, in *Neural Information Processing*, Lecture Notes in Computer Science, Vol. 3316 (Springer, Berlin/Heidelberg, 2004), pp. 989–995.
37. R. Alcalá, J. Alcalá-Fdez, and F. Herrera, A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection, *IEEE Trans. Fuzzy Syst.* **15**(4) (2007) 616–635.
38. Z. Xu and R. Yager, Some geometric aggregation operators based on intuitionistic fuzzy sets, *Int. J. Gen. Syst.* **35**(4) (2006) 417–433.
39. 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, *Int. J. Intell. Syst.* **23**(6) (2008) 680–714.
40. J. M. Mendel, R. I. John and F. Liu, Interval type-2 fuzzy logic systems made simple, *IEEE Trans. Fuzzy Syst.* **14**(6) (2006) 808–821.
41. G. Deschrijver, C. Cornelis and E. Kerre, On the representation of intuitionistic fuzzy t-norms and t-conorms, *IEEE Trans. Fuzzy Syst.* **12**(1) (2004) 45–61.
42. B. C. Bedregal, On interval fuzzy negations, *Fuzzy Sets Syst.* **161**(17) (2010) 2290–2313.
43. G. Deschrijver, Generalized arithmetic operators and their relationship to t-norms in interval-valued fuzzy set theory, *Fuzzy Sets Syst.* **160**(21) (2009) 3080–3102.
44. H. Ishibuchi, T. Yamamoto and T. Nakashima, Hybridization of fuzzy GBML approaches for pattern classification problems, *IEEE Trans. Syst. Man Cybern. B* **35**(2) (2005) 359–365.

45. H. Bustince, E. Barrenechea, M. Pagola, J. Fernandez, C. Guerra, P. Couto and P. Melo-Pinto, Generalized Atanassov's intuitionistic fuzzy index. Construction of Atanassov's fuzzy entropy from fuzzy implication operators, *Int. J. Uncertain. Fuzz.* **19**(1) (2011) 51–69.
46. O. Cordón, M. J. del Jesus and F. Herrera, A proposal on reasoning methods in fuzzy rule-based classification systems, *Int. J. Approx. Reason.* **20**(1) (1999) 21–45.
47. I. Bajla, I. Hollander, D. Czedik-Heiss and R. Granec, Classification of image objects in epo doping control using fuzzy decision tree, *Pattern Anal. Appl.* **12** (2009) 285–300.
48. K. Park, M. G. H. Bell, I. Kaparias and H. Belzner, Soft discretization in a classification model for modeling adaptive route choice with a fuzzy id3 algorithm, *Transport. Res. Rec.* **2076** (2008) 20–28.
49. F. Aymerich, J. Alonso, M. Cabaas, M. Comabella, P. Sobrevilla and A. Rovira, Decision tree based fuzzy classifier of magnetic resonance spectra from cerebrospinal fluid samples, *Fuzzy Sets Syst.* **170**(1) (2011) 43–63.
50. L. Eshelman, The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination, in *Foundations of Genetic Algorithms* (Morgan Kaufman, 1991), pp. 265–283.
51. A. Fernández, M. J. del Jesus and F. Herrera, On the 2-tuples based genetic tuning performance for fuzzy rule based classification systems in imbalanced data-sets, *Inform. Sciences* **180**(8) (2010) 1268–1291.
52. S. García, A. Fernández, J. Luengo and F. Herrera, A study of statistical techniques and performance measures for genetics–based machine learning: Accuracy and interpretability, *Soft Comput.* **13**(10) (2009) 959–977.
53. D. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, 2nd edn. (Chapman & Hall/CRC, 2006).
54. F. Wilcoxon, Individual comparisons by ranking methods, *Biometrics* **1** (1945) 80–83.
55. J. L. Hodges and E. L. Lehmann, Ranks methods for combination of independent experiments in analysis of variance, *Annals of Mathematical Statistics* **33** (1962) 482–497.
56. S. Holm, A simple sequentially rejective multiple test procedure, *Scand. J. Stat.* **6** (1979) 65–70.
57. R. Alcalá, Y. Nojima, F. Herrera and H. Ishibuchi, Multiobjective genetic fuzzy rule selection of single granularity-based fuzzy classification rules and its interaction with the lateral tuning of membership functions, *Soft Comput.* **15**(12) (2011) 2303–2318.
58. C. J. Carmona, P. González, M. J. Gacto and M. J. del Jesus, Genetic lateral tuning for subgroup discovery with fuzzy rules using the algorithm NMEEF-SD, *Int. J. Comput. Int. Sys.* **5**(2) (2012) 355–367.
59. F. Herrera and L. Martínez, A 2-tuple fuzzy linguistic representation model for computing with words, *IEEE Trans. Fuzzy Syst.* **6**(8) (2000) 746–752.