COMPARING DIFFERENT STOPPING CRITERIA FOR FUZZY DECISION TREE INDUCTION THROUGH IDFID3

M. ZEINALKHANI AND M. EFTEKHARI

Abstract. Fuzzy Decision Tree (FDT) classifiers combine decision trees with approximate reasoning offered by fuzzy representation to deal with language and measurement uncertainties. When a FDT induction algorithm utilizes stopping criteria for early stopping of the tree's growth, threshold values of stopping criteria will control the number of nodes. Finding a proper threshold value for a stopping criterion is one of the greatest challenges to be faced in FDT induction. In this paper, we propose a new method named Iterative Deepening Fuzzy ID3 (IDFID3) for FDT induction that has the ability of controlling the trees growth via dynamically setting the threshold value of stopping criterion in an iterative procedure. The final FDT induced by IDFID3 and the one obtained by common FID3 are the same when the numbers of nodes of induced FDTs are equal, but our main intention for introducing IDFID3 is the comparison of different stopping criteria through this algorithm. Therefore, a new stopping criterion named Normalized Maximum fuzzy information Gain multiplied by Number of Instances (NMGNI) is proposed and IDFID3 is used for comparing it against the other stopping criteria. Generally speaking, this paper presents a method to compare different stopping criteria independent of their threshold values utilizing IDFID3. The comparison results show that FDTs induced by the proposed stopping criterion in most situations are superior to the others and number of instances stopping criterion performs better than fuzzy information gain stopping criterion in terms of complexity (i.e. number of nodes) and classification accuracy. Also, both tree depth and fuzzy information gain stopping criteria, outperform fuzzy entropy, accuracy and number of instances in terms of mean depth of generated FDTs.

1. Introduction

Supervised learning methods attempt to discover the relationship between input attributes and a target attribute. The relationship discovered is represented in a structure referred to as a model. Usually, models can be used for predicting the value of the target attribute when the values of the input attributes are available. There are two main supervised models: regression models and classification models (classifiers). Regression models map the input space into a real-valued domain while classifiers map the input space into predefined classes [25].

Received: January 2012; Revised: September 2012, November 2012 and February 2013; Accepted: May 2013

Key words and phrases: Fuzzy inference systems, Classification, Fuzzy decision tree, Stopping criteria.

Decision trees [24, 25, 34] are commonly used as classification models in data mining that their induction methods recursively partition the instance space for generating tree structured models. The recursion is completed when the stopping condition is satisfied. In these tree structures, leaves represent classifications and branches represent conjunctions of attributes that lead to those classifications. Decision trees are comprehensible classifiers that cannot handle language uncertainties and are noise sensitive.

Fuzzy sets provide bases for fuzzy representation. Fuzzy sets and fuzzy logic allow the modeling of language related uncertainties. Fuzzy sets along with fuzzy logic and approximate reasoning methods provide the ability to model fine knowledge details [14]. Accordingly, fuzzy representation is becoming increasingly popular in dealing with problems of uncertainty, noise and inexact data [18]. Approximate reasoning offered by fuzzy representation has been combined with decision trees to preserve advantages of both: uncertainty handling and gradual processing of the former with the comprehensibility, popularity and ease of application of the latter [14].

In the literature, several publications on FDTs are available and a comprehensive survey of them can be found in [8, 28]. Some of these publications have focused on the generation of the proper membership functions which lead to better FDTs. For example, Wang and Jiarong [31] discretized continuous attributes using fuzzy numbers and probability theory. Pedrycz and Sosnowski [20] employed context-based fuzzy clustering for this purpose. Other publications tried to improve the FDT induction algorithm itself. These publications can be categorized in three main groups according to their improvement. The first group is the publications that have proposed some new splitting criteria. Yuan and Shaw [32] introduced a fuzzy decision tree induction algorithm which selects the branching attribute on the basis of classification ambiguity. Jensen and Shen [15] utilized a fuzzy rough set based splitting criterion for FDT induction. Wang and Borgelt [29] discussed the problem of information gain as a splitting criterion and have proposed some improvements. The second group contains the publications that have optimized fuzzy decision tree by evolutionary algorithms or multi-objective methods. Wang et al. [30] proposed optimization principles of fuzzy decision trees based on minimizing the total number and average depth of leaves. They proved that the construction of a minimum FDT is a NP-Hard problem. Pedrycz and Sosnowski [21] described an optimization procedure for obtaining optimal fuzzy decision tree using genetic algorithm. Pulkkinen and Koivisto have applied multi-objective evolutionary algorithm to optimize FDT [22]. The last group is the publications that have combined FDT with other techniques. Chang et al. have integrated case-based reasoning techniques with fuzzy decision tree [7]. Chandra and Paul Varghese [5, 6] have modified the SLIQ decision tree algorithm to construct a binary fuzzy decision tree. Bhatt and Gopal [4] have combined neural networks with fuzzy decision trees. They have utilized back-propagation to improve the learning accuracy of FDT without compromising its comprehensibility. Bartczuk and Rutkowska [3] have introduced type-2 fuzzy decision tree induction algorithm which employs type-2 fuzzy sets. Khan et al. [16] have used rule weighting technique to improve the accuracy of fuzzy decision tree.

Sanz et al. proposed the methodology called IIVFDT (Ignorance functions based Interval-valued Fuzzy Decision Tree with genetic tuning) by using an ignorance degree [26].

One of the challenges in fuzzy decision tree induction is to develop algorithms that produce fuzzy decision trees of small size and depth. Larger fuzzy decision trees (or over-fitted FDTs) lead to poor generalization performance as well as more computational efforts they need to classify a new instance. FDT's growth controlling methods try to cope with the problem of over-fitting. They may be classified into two categories; post-pruning methods and pre-pruning methods.

Post-pruning methods allow the FDT to over-fit the training data. Then the over-fitted FDT is cut back into a smaller one by removing sub-branches that are not contributing to the generalization accuracy. The problem of these methods is that they are computationally prohibitive. Cintra et al. [9] have analyzed the impact of different pruning rates on fuzzy decision trees.

Pre-pruning methods stop the development of the FDT according to a stopping criterion before the tree is completed. Although these methods require less computation, choosing the proper stopping criterion and deciding on its threshold value is their main problem. Employing tight stopping criteria tends to create small and under-fitted fuzzy decision trees. On the other hand, using loose stopping criteria tends to generate large decision trees that are over-fitted to the training data.

The threshold value of stopping criterion controls the size of FDT. FDTs of the same size can be induced by different stopping criteria, but the structures of generated trees are not necessarily similar. This motivated us to study the effect of different stopping criteria on FDT construction. In this paper, at first, a novel FDT induction algorithm is proposed that construct the FDT by fuzzy ID3 algorithm in an iterative procedure. This algorithm has the ability of controlling the growth of FDT's size and is called Iterative Deepening Fuzzy ID3 (IDFID3). It utilizes a stopping criterion to guide the tree growth. Second, a new stopping criterion, which is named Normalized Maximum fuzzy information Gain multiplied by Number of Instances (NMGNI), is proposed. Then some previously developed stopping criteria taken from the literature are compared with our proposed one in terms of the size and accuracy of generated FDTs. Our aim in this paper is to compare different stopping criteria independent of their threshold values through IDFID3.

The organization of the paper is as follows. Section 2 reviews the crisp decision trees. Section 3 addresses the fuzzy decision trees and their building and inference procedures. Section 4 describes IDFID3 as well as a new proposed stopping criterion (NMGNI). Also a method for comparing stopping criteria independent of their threshold values is described in section 4. The experimental results obtained are presented in the 5th section. Finally, section 6 concludes the paper.

2. Decision Tree

A crisp decision tree classifies instances by sorting them down the tree from the root to some leaf nodes by which the classification of instances are provided. Each internal node in the tree specifies a test of single attribute of the instance, and each branch descending from that node corresponds to one of the possible values for this attribute. Each leaf is assigned to one class representing the most appropriate target value. An instance is classified by starting at the root node of the tree, testing the attribute specified by this node, then moving down the tree branch corresponding to the value of the attribute in the given example. This process is then repeated for the subtree rooted at the new node [19].

ID3 is a popular algorithm for decision tree induction which constructs it in a top-down manner, beginning with the question which attribute should be tested at the root node of the tree? To answer this question, each instance attribute is evaluated using a splitting criterion to determine how well it alone classifies the training examples. The best attribute is selected and used as the test at the root node of the tree. A descendant of the root node is then created for each possible value of this attribute, and the training examples are downed the branch corresponding to the examples value for this attribute. The entire process is then repeated using the training examples associated with each descendant node to select the best attribute to test at that point in the tree until a stopping criterion is satisfied [19].

Information gain is a commonly used splitting criterion for decision tree induction which measures how well a given attribute separates the training examples according to their target classification. This criterion uses entropy that characterizes the (im)purity of an arbitrary collection of examples. If the target attribute can take on m possible values, then the entropy of collection S relative to this m-wise classification is defined as:

$$Entropy(S) = \sum_{i=1}^{m} -\frac{|S_i|}{|S|} log_2 \frac{|S_i|}{|S|} = \sum_{i=1}^{m} -p_i log_2 p_i \tag{1}$$

where $|S_i|$ is the number of instances of S belonging to class i and p_i is the probability of class i. The information gain of an attribute A, relative to a collection of examples S is defined as:

$$InformainGain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{2}$$

where $Values(A)$ is the set of all possible values for attribute A, and S_v is the subset of S for which attribute A has value v (i.e. $S_v = \{s \in S | A(s) = v\}$).

3. Fuzzy Decision Tree

3.1. **Notations.** Before we define the fuzzy decision tree construction and inference procedures, let us introduce some notations.

- The set of attributes of a dataset is denoted by $\{A_1, A_2, ..., A_k, Y\}$, where A_i is the i^{th} attribute, k is the number of input attributes and Y is the target attribute (each attribute is a column in dataset).
- The set of values of target variable (classes) is $Y \in \{c_1, c_2, ..., c_m\}$, where m is the number of classes.
- The examples in the fuzzy dataset S is denoted by

$$S = \{(\mathbf{X}_1, \mu_s(\mathbf{X}_1)), (\mathbf{X}_2, \mu_s(\mathbf{X}_2)), ..., (\mathbf{X}_n, \mu_s(\mathbf{X}_n))\},$$

where $\mathbf{X}_i$ is the i^{th} example, $\mu_s(\mathbf{X}_i)$ is the membership degree of $\mathbf{X}_i$ in S, and n is the number of examples. $\mathbf{X}_i$ has written in bold face because it is a vector containing input attributes and target attribute.

- The i^{th} example denoted by $\mathbf{X}_i = [x_i^{(1)}, x_i^{(2)}, ..., x_i^{(k)}, y_i]^T$, where $x_i^{(j)}$ is the value of j^{th} attribute, and y_i is the value of target attribute.
- Fuzzy terms defined on i^{th} attribute, A_i, are denoted by $\{F_i^{(1)}, F_i^{(2)}, ..., F_i^{(r_i)}\}$, where $F_i^{(j)}$ is the j^{th} fuzzy term and r_i is the number of fuzzy terms defined on attribute A_i.
- Membership function corresponding to the fuzzy term $F_i^{(j)}$ is denoted by $\mu_{F_i}(j)$.
- The number of examples in the fuzzy dataset S denoted by $|S|$ and is defined as $|S| = \sum_{i=1}^{n} \mu_s(\mathbf{X}_i)$.
- The examples of fuzzy dataset S which belong to the class i is denoted by $S_{y=c_i}$.
- If S denotes the fuzzy dataset of parent node, then the fuzzy dataset of child node corresponding to the fuzzy term $F_i^{(j)}$ is denoted by $S[F_i^{(j)}]$. For example consider child nodes in Figure 1. In this figure, the branching attribute is A_i and the name of fuzzy dataset for each node has been written on it. In this paper the j^{th} child means the child node corresponding to the fuzzy term $F_i^{(j)}$.

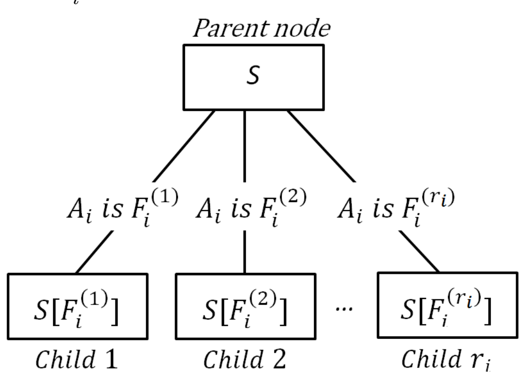


Figure 1. A Typical Parent Node with Child Nodes in Fuzzy Decision Tree

3.2. **Fuzzy Decision Tree Construction.** Fuzzy decision trees allow instances to follow down multiple branches simultaneously, with different satisfaction degrees ranged on [0, 1]. To implement these characteristics, fuzzy decision trees use fuzzy linguistic terms to specify branching condition of nodes. In FDTs, an instance may be fall into many leaves with different satisfaction degree ranged on [0, 1], because it can fall into many child nodes of parent node at any level. This fact is actually advantageous as it provides more graceful behavior, especially when dealing with noise or incomplete information. However, from a computational point of view, the

FDT induction is slower than crisp decision tree induction. This is the price paid for having a more accurate but still interpretable classifier [8, 28].

Generally speaking, fuzzy decision tree induction has two major components: a procedure for fuzzy decision tree construction, and an inference procedure for decision making (i.e. class assignment for new instances). One of the FDT building procedures is a fuzzy extension to the well-known ID3 algorithm [19] that is named Fuzzy ID3 (FID3). FID3 employs predefined fuzzy linguistic terms by which the attribute values of training data are fuzzified [27]. This method extends the information gain measure to determine branching attribute of each node expansion. Moreover, FID3 uses fuzzy dataset in which for each example a degree of membership in the dataset is included as well as crisp value of attributes (input attributes and target attribute).

Fuzzy dataset of the child nodes has all the examples of the parent node with different degree of membership. Suppose that S is the fuzzy dataset of parent node and A_i is the branching attribute with $\{F_i^{(1)}, F_i^{(2)}, ..., F_i^{(r_i)}\}$ fuzzy terms, and the fuzzy dataset of the child node corresponding to the fuzzy term $F_i^{(j)}$ is $S[F_i^{(j)}]$. The membership degree of h^{th} example, $\mathbf{X}_h = [x_h^{(1)}, x_h^{(2)}, ..., x_h^{(k)}, y_h]^T$, in $S[F_i^{(j)}]$ is defined as:

$$\mu_{S[F_i^{(j)}]}(\mathbf{X}_h) = \mu_S(\mathbf{X}_h) \times \mu_{F_i^{(j)}}(x_h^{(i)}) \tag{3}$$

where $\mu_S(\mathbf{X}_h)$ is the membership degree of $\mathbf{X}_h$ in S, and $\mu_{F_i^{(j)}}(x_h^{(i)})$ is the membership degree of $x_h^{(i)}$ in the membership function corresponding to fuzzy term $F_i^{(j)}$, namely $\mu_{F_i^{(j)}}$. In the generalized case, multiplication operator can be replaced by a t-norm operator.

FID3 selects the attribute with maximum Fuzzy Information Gain (FIG) as a branching attribute. FIG utilizes Fuzzy Entropy (FE) which is defined as:

$$FE(S) = \sum_{i=1}^{m} -\frac{|S_{y=c_i}|}{|S|} log_2 \frac{|S_{y=c_i}|}{|S|} \tag{4}$$

The FIG of attribute A_i, relative to a fuzzy dataset S is defined as:

$$FIG(S, A_i) = FE(S) - \sum_{j=1}^{r_i} w_j \times FE(S[F_i^{(j)}]) \tag{5}$$

where $FE(S)$ is the fuzzy entropy of fuzzy dataset S, $FE(S[F_i^{(j)}])$ is the fuzzy entropy of j^{th} child node, and w_j is the fraction of examples which belong to j^{th} child node. w_j is defined as:

$$w_j = \frac{|S[F_i^{(j)}]|}{\sum_{k=1}^{r_i} |S[F_i^{(k)}]|} \tag{6}$$

The first term in equation (5) is just the fuzzy entropy of S, and the second term is the expected value of the fuzzy entropy after S is partitioned using attribute A_i. $FIG(S, A_i)$ is therefore the expected reduction in fuzzy entropy caused by knowing the value of attribute A_i. There are some other methods for selecting branching attribute [29, 32]. Figure 2 summarizes the fuzzy decision tree construction procedure.

Inputs: Crisp train data and predefined membership functions on each attribute, splitting criterion, stopping criterion and the threshold value of stopping criterion.

(1) Generate the root node with a fuzzy dataset containing all crisp training data that all the membership degrees have assigned to one.
(2) For each new node N, with fuzzy dataset S do steps 3 to 7.
(3) If stopping criterion satisfied, then
(4) Make N as a leaf and assign the fraction of examples of N belonging to each class as a label of that class.
(5) If stopping criterion does not satisfied, then
(6) Calculate FIG for each attribute, and select the attribute A_{max} that maximize FIG as a branching attribute.
(7) Generate new child nodes $\{child_1, child_2, ..., child_{r_{max}}\}$, where $child_j$ is in correspondence to fuzzy term $F^{(j)}_{max}$ containing fuzzy dataset $S[F^{(j)}_{max}]$ and have all the attributes of S except A_{max}. The membership degree of each example in $S[F^{(j)}_{max}]$ is calculated by equation (3).

FIGURE 2. Fuzzy Decision Tree Induction Algorithm

3.3. **Inference for Decision Assignment.** A classical decision tree can be converted to a set of rules [23]. One can think of each leaf as one rule; the conditions leading to the leaf generate the conjunctive antecedence, and the classification of the examples in the leaf produces the consequence. In this case, a consistent set of rules is generated only when examples of every leaf have a unique classification, which may happen only when a sufficient set of attributes is used and the training data is consistent. Because in fuzzy representation a value can have a nonzero membership in more than one fuzzy set, the inconsistency problem dramatically increases. To come up with this problem, approximate reasoning methods have been used to inference for decision assignment [14].

Like classical decision trees, FDTs also can be converted to a set of fuzzy if-then rules. We can apply the approximate reasoning, to derive conclusions from these set of fuzzy if-then rules and known facts. Briefly, approximate reasoning can be divided into four steps [13]:

(1) **Degrees of compatibility**: Compare the known facts with the antecedents of fuzzy rules to find the degrees of compatibility with respect to each antecedent MF.
(2) **Firing strength**: Combine degrees of compatibility with respect to antecedent MFs in a rule to form a firing strength that indicates the degree to which the antecedent part of the rule is satisfied. In fuzzy decision tree, the antecedent MFs are connected with AND operators. Therefore, one t-norm operator can be used for this step in FDT. We adopt the multiplication operator among many alternatives.
(3) **Certainty degree of each class in each rule**: For calculating this, the firing strength of each rule should be combined with the certainty degree of the classes attached to the leaf node. A t-norm operator can be employed for this step. The multiplication operator is used in this paper. Therefore, for the above given example, each rule has two certainty degrees relating to classes C_1 and C_2.
(4) **Overall output**: Aggregate all the certainty degrees from all fuzzy if-then rules relating to the same class. One s-norm operator should be used for

aggregation. We adopt the sum from several alternatives. It maybe the total values of certainty degrees for different classes exceed the unity and thereby we normalize them.

3.4. **Stopping Criteria.** The fuzzy decision tree with the minimum number of nodes and maximum classification accuracy is the best one. Increasing the number of nodes enhance the classification accuracy. On the other hand, the tree with many nodes with respect to the number of training examples may be over-fitted. Therefore, the best FDT is the one making the best trade-off between complexity (number of nodes) and classification accuracy.

The stopping criterion, one of the growth control methods, controls the structure of FDT and its threshold determines the complexity (number of nodes) of FDT. Two different stopping criteria can generate two FDTs which have the equal size in terms of the number of nodes, but have different structures. Figure 3 shows two trees with equal size but with different structures.

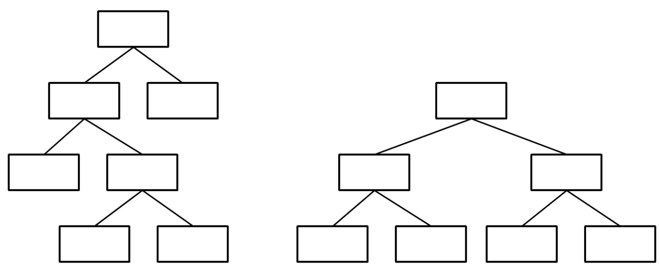


Figure 3. Trees with Equal Size But with Different Structures

Three factors impact on the structure of FDT; stopping criteria, splitting criteria and the number of membership functions defined on each attribute's domain. The number of membership functions of attribute A_i determines the number of child nodes that should be generated if A_i has been selected as a branching attribute. Splitting criteria determine the branching node and consequently the number of child nodes for each branch. Stopping criteria determine which nodes should be expanded and also its threshold determines when the FDT growth should be stopped. Having the same splitting criterion and same membership functions on each attributes domain, the stopping criterion is the only effective parameter that controls the structure of FDT.

The structure of FDT has influence on its classification accuracy. Therefore, the type of stopping criterion used in FDT induction, regardless its threshold value, has impact on the classification accuracy of FDT. The threshold value of stopping criterion also has influence on the classification of FDT. Since number of nodes in FDT has been controlled by the threshold value and the number of nodes has great impact on the accuracy of FDT. Briefly, both types of stopping criterion and its threshold value have influence on the classification accuracy of FDT; stopping criterion by controlling the structure of FDT and its threshold by controlling the number of nodes.

There are two types of stopping criteria that are named type I and type II in this study. In type I, the FDT construction algorithm continues while there is a leaf (with at least one attribute) for which the value of stopping criterion is less than a predefined threshold value. The accuracy of node is an example of a stopping criterion of type I. The threshold value 0.8 for this stopping criterion means that the expansion of nodes must be performed while there is a leaf with the accuracy less than 80 percent; otherwise nodes expansion should be stopped. On the other hand, a type II stopping criterion is one that its value must be greater than a predefined threshold value to continue node expansion. The number of instances covered by a node is an example of this type of stopping criteria. The threshold value 10 for this stopping criterion means that nodes with more than 10 instances must be expanded until each leaf node has less than 10 instances.

FDT can also utilize the stopping criteria used in crisp decision trees. Common stopping criteria for crisp decision trees reported in the literature are as follows [25]:

(1) Accuracy of node
(2) Number of instances of node
(3) Tree depth
(4) Entropy
(5) Information Gain

In the literature, the first and second criteria have been extended to the FDT induction [23]. To employ entropy and information gain criteria for FDT induction, their fuzzy versions must be used (these criteria have been introduced in the equations (4) and (5) respectively).

4. **Proposed Methods**

4.1. **Iterative Deepening FID3.** The main problem of stopping criteria is to determine a proper threshold value for them. Considering a specific threshold value, the number of nodes of induced FDT is not predictable and thereby obtaining a predefined FDT in terms of the number of nodes is impossible. In this section, we are going to introduce a new fuzzy decision tree construction method named Iterative Deepening FID3 (IDFID3) to overcome the problem of controlling FDT size (i.e. number of nodes) via stopping criteria. IDFID3 employs the same splitting and stopping criteria and the inference procedure of FID3.

IDFID3 uses a predefined Minimum Number of Nodes (MNN) of FDT instead of threshold value of stopping criterion and by dynamically setting the threshold value of stopping criterion controls the trees growth. The idea used in IDFID3 is that a fixed predefined value of threshold for a stopping criterion is not suitable for controlling the size of tree. Therefore, IDFID3 varies the threshold value dynamically in its iterations in order to have a better control on the tree size.

IDFID3 categorizes the nodes, which are not expanded, into two groups: 1- expandable, 2- not expandable. When the fuzzy dataset of a node has at least one attribute to be used as a branching attribute, it is expandable, otherwise it is not expandable. Then, IDFID3 determines the expansion preference of each expandable node based on the type and the value of stopping criterion for it. Then, determines the node with highest expansion preference (HEP) and continue the tree growth until all expandable nodes have expansion preference less than HEP. This process iteratively repeated until induced FDT has at least MNN nodes. IDFID3 induces the same FDT as FID3, when the numbers of nodes of induced FDTs are equal. Figure 4 shows the IDFID3 algorithm. Because the node expansion process in IDFID3 is the same as node expansion in FID3 (step 6 and 7 in Figure 2), in this figure, details of node expansion have been removed to make its understanding easier.

In each iteration of IDFID3s while loop, the algorithm tries to add a minimum number of nodes to FDT by changing the threshold value. In other words, the node with highest expansion preference should be found and expanded in each iteration. Expansion preference of one node depends on the type of stopping criterion employed in FDT induction. In type I stopping criteria, increasing the threshold value leads to increase in the number of nodes; therefore, the lower the threshold value for a node, the higher is the priority for expansion. After the step four, where the new threshold value was determined, all expandable nodes with the stopping criterion value less than or equal to the new threshold will be expanded. On the other hand, for stopping criteria of type II, decreasing the threshold value increases the number of nodes; therefore, the higher values of threshold imply the higher priority for expansion. Namely, after step four, all expandable nodes with the stopping criterion value greater than or equal to the new threshold will be expanded.

Inputs: Crisp train data, predefined membership functions on each attribute, splitting criterion, stopping criterion and the Minimum Number of Nodes (MNN).

(1) Generate a root node with a fuzzy dataset containing all crisp training data and set all membership degrees to one.
(2) While (minimum number of nodes has not generated) do steps 3-6
(3) Find an expandable node N with highest expansion preference
(4) Set threshold value to the value of stopping criterion for N
(5) Expand the node N
(6) While there is an expandable node in FDT which its expansion preference is higher than or equal to the expansion preference of N, expand it.
(7) Make all not expanded nodes as a leaf and assign the fraction of examples of N belonging to each class as a label of that class.

Figure 4. Iterative Deepening FID3 (IDFID3) Algorithm

A new threshold value is found in step 4 may lead to expansion of more than one node and thereby make estimating the number of new generated nodes difficult. IDFID3 tries to add minimum number of new nodes to FDT in each iteration. This is achieved by iteratively growing and deepening FDT by expansion of a subtree rooted at the node with the highest expansion preference in each iteration. The number of new generated nodes in each iteration depends on the distribution of data and the employed stopping criterion. We have introduced a growth control capability (GCC) measure to determine the ability of stopping criterion to control the trees growth [33].

FID3 construction algorithm adds new child nodes to FDT similar to a greedy Hill-climbing search in which the algorithm selects the best node for inserting in the tree based on FIG and never backtracks to reconsider earlier choices. The proposed IDFID3 selects nodes based on FIG for inserting in the tree, but the order of expansion of these selected nodes is not in DFS (Depth First Search) manner. The IDFID3 behaves like the IDA* (Iterative Deepening A*) search method [17], which its heuristic function is a stopping criterion. In other words, the algorithm makes a preference for expanding the candidate nodes by means of the value of stopping criterion for them. Therefore, the growth of tree in the proposed IDFID3 is performed similar to the combination of Breadth first search and DFS algorithms; that is along the width and depth of the tree.

Figure 5 shows a completely expanded FDT constructed on a fuzzy data set with four attributes. The value inside each node shows its accuracy. Figure 6 illustrates four iterations of IDFID3 using the accuracy stopping criterion. In this figure, new generated nodes are highlighted and the new threshold value, which should be employed in the next iteration, is written in bold face inside the node with highest expansion preference.

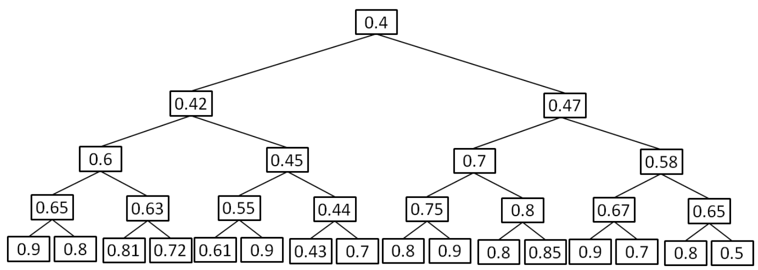
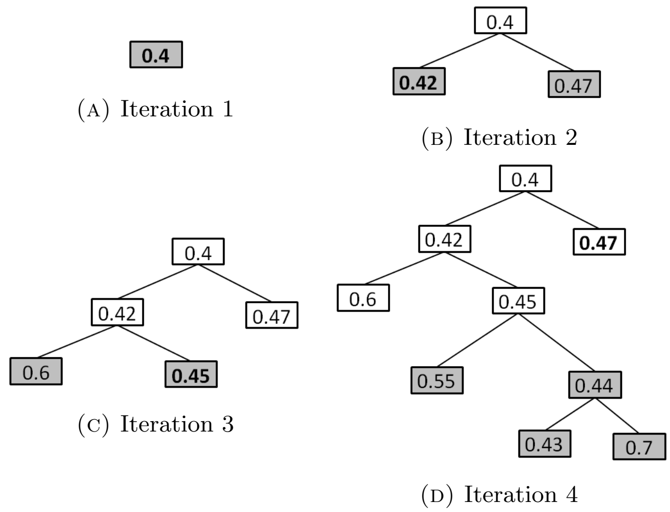


Figure 5. Complete Fuzzy Decision Tree; the Value Inside Each Node Shows Its Accuracy

In the first iteration, the root node is constructed and the value of threshold is set to 0.4, that is, the accuracy of the root node. In the second iteration two new nodes are generated that their accuracy values (i.e. stopping criterion values) are 0.42 and 0.47. The expansion process terminates in the second iteration because both 0.42 and 0.47 values violate the threshold value determined in the first iteration (i.e. 0.4). In this iteration the minimum value of accuracies (i.e. 0.42) is used as a threshold value for the next iteration. In the third iteration, the node that has the lowest value of accuracy (that is the node with accuracy of 0.42 in the previous iteration) has the highest preference for expansion. In this step, two new nodes with accuracies 0.6 and 0.45 are produced by the expansion of this node. Both of these accuracy values exceed from the threshold value for this step. Thus, the algorithm stops the expansion in this iteration and set the 0.45 (minimum value among 0.6, 0.45 and 0.47) as a threshold value for the next iteration. In the fourth iteration, algorithm starts at node with accuracy 0.45 because it has the highest priority for expansion. Its expansion generates two child nodes with the accuracy values of 0.55 and 0.44. The first child (i.e. the node with accuracy 0.55) exceeds the threshold value, but the second one, which has the accuracy 0.44, does not exceed the threshold value and so this iteration should be continued with its expansion.



(A) Iteration 1

(B) Iteration 2

(C) Iteration 3

(D) Iteration 4

FIGURE 6. Four Iterations of IDFID3

The expansion of this node produces two new child nodes with the accuracies 0.43 and 0.7. None of these new generated child nodes are expandable, because the fuzzy dataset has four attributes and all of them have been used in the parent nodes as a branching attribute. Therefore, this iteration should be terminated. Finally, the threshold value of the next iteration is set to 0.47, because 0.47 is the minimum accuracy value among the accuracy values of expandable nodes (i.e. among 0.6, 0.55 and 0.47). If the number of generated nodes is less than MNN (which is specified by user), the iterations should be continued and it should be stopped otherwise. As can be seen from the above example, in each iteration at least one node will be expanded.

4.2. **Proposed Stopping Criterion.** IDFID3 induction method utilizes the greedy algorithm to construct the fuzzy decision tree. Greedy algorithms make a locally optimal choice in the hope that this choice will lead to a globally optimal solution. IDFID3 induction method utilizes the greedy approach in two steps. The first step is the selection of an attribute for branching that selects the one with maximum FIG (i.e. splitting criterion). The second step is the selection of node for expansion that selects a node with highest expansion preference based on a stopping criterion. In other words, the greedy approach of IDFID3 is controlled by two heuristics; splitting criterion and stopping criterion. In this paper, a fix method for splitting criterion (i.e. fuzzy information gain) is used. This section describes the new proposed stopping criterion which leads to better FDTs.

Which node should be selected for expansion in the IDFID3 iterations? The best choice, according to greedy approach, is the node which its expansion leads to maximum increase in the accuracy of FDT. Calculating the precise amount of increase in the accuracy of FDT for expansion of a certain node requires the evaluation of whole FDT. Since the evaluation of FDT demands high computational efforts, heuristic functions that approximate the accuracy are good alternatives. Such a heuristic can be utilized as a stopping criterion that its value for tree nodes provides a suitable guide for IDFID3. Therefore, in the greedy approach the best stopping criterion is the one that in each iteration assigns the highest expansion preference to the node that its expansion results in maximum increase in the accuracy of FDT.

Assuming that the process of FDT construction is in an arbitrary node and the fuzzy data set related to that node is S_c (c stands for current). Then, one attribute that has the highest value of fuzzy information gain must be selected as a branching attribute. This attribute is named A_M and is obtained as the following:

$$A_M = \arg\max_{A_i} FIG(S_c, A_i) \tag{7}$$

$FIG(S_c, A_i)$ is the fuzzy information gain of attribute A_i relative to S_c which is described in section 3.2. Expansion of current node according to the attribute A_M makes the maximum difference between information of current node and its child nodes and is possibly the most promising attribute for improving the accuracy of classification. The higher value of FIG related to A_M means the more contribution of this attribute in improving the total classification accuracy. Considering the attribute A_M, the proposed criterion, that is named Normalized Maximum fuzzy information Gain multiplied by Number of Instances ($NMGNI$), is defined as follows:

$$NMGNI(S_c) = |S_c| \times \frac{FIG(S_c, A_M)}{log_2 m} \tag{8}$$

where $FIG(S_c, A_M)$ is the fuzzy information gain of attribute A_M related to S_c, m is the number of classes and $|S_c|$ is the number of instances of S_c. Having m classes, FIG can get a value in the range $[0, log_2 m]$. Thus, in order to normalize the FIG value and taking a value in range $[0, 1]$, it is divided by $log_2 m$. The higher the value of term $\frac{FIG(S_c,A_M)}{log_2 m}$, the higher the possible contribution of the current node for final classification in FDT. Multiplying this term by $|S_c|$ indicates the proportion of fuzzy dataset S_c that may be correctly classified after the expansion of current node. Therefore, the proposed $NMGNI$ heuristic seems to be suitable in terms of predicting the usefulness of a node for contributing in classification. In the following sections, we are going to justify this hypothesis by performing several experiments and comparisons.

4.3. **Comparison Method for Various Stopping Criteria.** This section presents a method to compare two stopping criteria independent of their threshold values. This comparison method considers the effectiveness of a stopping criterion for FDT induction using the IDFID3 in terms of accuracy and complexity (number of nodes). The best stopping criterion is the one leads to FDTs with the minimum number of nodes and maximum accuracy. Several significant parameters affect the number of nodes and accuracy of induced FDT such as stopping criterion, its threshold value, splitting criterion, type of membership functions and number of membership functions. In order to study the effectiveness of different stopping criteria, splitting criterion and type of membership functions are set to FIG and triangular membership functions respectively. By using different number of membership functions and averaging on the accuracies of induced FDTs, it is tried to include the effect of the number of membership functions in the comparison results. In addition, the effect of different threshold values are included in the comparison results by averaging on accuracies of the FDTs induced in different iterations of IDFID3. Considering different number of membership functions and making the comparison result independent of threshold value of stopping criterion makes the obtained results more robust and reliable. Proposed comparison method considers the classification accuracy and size of induced FDT simultaneously.

As discussed in section 3.4, stopping criterion and its threshold value have impact on both of accuracy and the size of FDT. One stopping criterion that results in constructing more accurate FDTs for each arbitrary size is better than the others. Figure 7 shows the error and size of FDTs generated in different iterations of IDFID3 for a specific stopping criterion. In this figure, square markers correspond to generated FDTs in iterations of IDFID3. For example, in 11th iteration, an FDT with 52 nodes and in the next iteration an FDT with 60 nodes have been generated. As it is apparent from Figure 7, in different iterations of IDFID3 various FDTs in terms of size are produced. Therefore, the line that connects two square points is used to interpolate the error of FDTs between two sequential iterations. For example, in Figure 7 between the 11^th and 12^th iterations, FDTs with sizes 53 to 59 are not produced and their accuracies are obtained by interpolation.

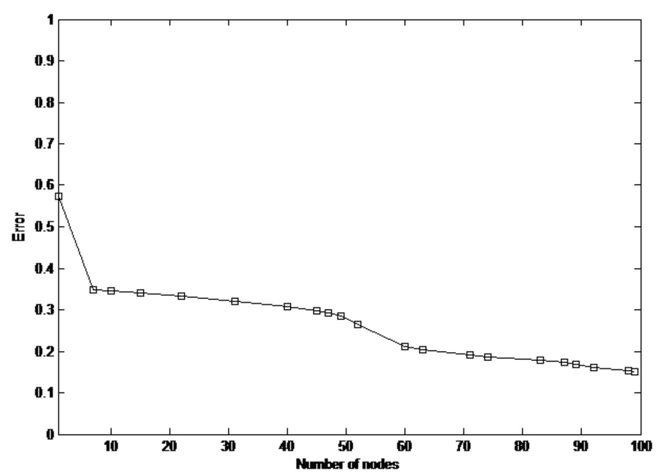


Figure 7. The Error and the Size of FDTs That are Generated in Different Iterations of IDFID3

Induction and inference in FDTs with large number of nodes requires high computational efforts. Moreover, they are probably over-fitted to train data and incorporating them in the comparison may lead to wrong conclusions. For these reasons, only the FDTs with less than 100 nodes are considered for comparison. For a specific stopping criterion, calculating the average error for all the FDTs of size 1 to 100 indicates the ineffectiveness rate of that criterion. Briefly, first, for a specific number of membership functions, the accuracy of different FDTs of arbitrary sizes is calculated (or interpolated if that size is not generated on the iterations of IDFID3). Then, the same process is repeated for different number of membership functions. Then, for each size of tree, an accuracy value is calculated by averaging on calculated accuracies for that size for different number of membership functions. Finally, ineffectiveness rate is calculated by averaging on calculated accuracies for different sizes in the previous step. The obtained ineffectiveness rate, which combines both accuracy and size of tree, is used to compare the stopping criteria.

5. Experimental Results

In experiments, different stopping criteria are employed in IDFID3 over multiple datasets to compare the effect of them on the accuracy of induced FDTs. In addition to proposed NMGNI stopping criterion, five other stopping criteria are employed: 1) accuracy, 2) number of instances, 3) fuzzy information gain, 4) fuzzy entropy and 5) tree depth (depth of node). Twenty numerical dataset have selected from UCI machine learning repository [11] and KEEL dataset repository [1] for experiments. Table 1 summarizes the properties of the selected datasets, showing for each dataset the number of attributes, the number of instances and the number of classes. On the domain of each attribute, uniformly distributed triangular membership functions are defined. Figure 8 depicts an example of five uniformly distributed triangular membership functions.

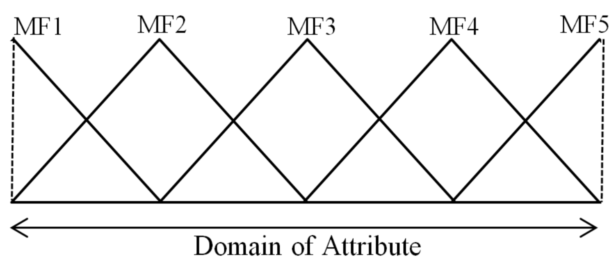


Figure 8. Five Uniformly Distributed Triangular Membership Functions

Aforementioned stopping criteria are compared using the method described in section 4.3. FDTs with more than 100 nodes are supposed to be over-fitted. The error of fuzzy decision trees is calculated based on 10-fold Cross Validation (10CV) approach. The generalization error of FDT is approximated by mean error value of FDTs induced in different folds. 10CV is repeated 10 times to make the obtained results more robust and reliable. Figure 9 shows the error rate of FDTs of different sizes induced from Ecoli dataset with four membership functions using 10CV approach. For the sake of readability only four stopping criterion are shown.

As it is apparent, the error of FDT induced by the number of instances as stopping criterion, when the tree has about 20 and also 30 nodes is less than our proposed criterion, but in the average our proposed criterion is better than all the others. By averaging on all error rates of each stopping criterion obtained in 10CV procedure for a specific number of membership functions, the ineffectiveness rate

Dataset	No. of Attributes	No. of Instances	No. of Classes
balance	4	625	3
blood transfusion	4	748	2
breast cancer	10	699	2
breast tissue	9	106	6
bupa	6	345	2
cleveland	13	297	5
contraceptive	9	1473	3
ecoli	7	336	8
glass	9	214	6
haberman	3	306	2
hayes-roth	4	160	3
heart	13	270	2
iris	4	150	3
newthyroid	5	215	3
sonar	60	280	2
tae	5	151	3
titanic	3	2201	2
vehicle	18	846	4
wine	13	178	3
yeast	8	1484	10

TABLE 1. Summary Description for the Employed Datasets

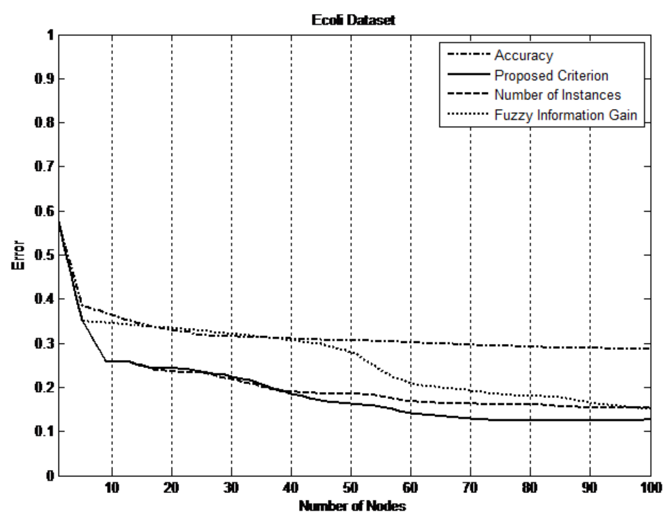


FIGURE 9. 10CV Error of FDTs Induced on Ecoli Dataset with Four MFs vs. the Size of Them

of that stopping criterion can be calculated. Lower value of ineffectiveness rate for specific stopping criterion indicates that it outperforms the others for most FDTs with less than 100 nodes. In order to make the obtained results more reliable, we calculated the ineffectiveness rate of each stopping criterion when different number of membership functions (i.e. 2, 3, ... 10 membership functions) are defined on each attribute of dataset. By averaging these nine ineffectiveness rates, the overall ineffectiveness rate of each stopping criterion for each dataset is calculated and listed in Table 2.

Dataset	Accuracy	Fuzzy Entropy	Proposed Criterion	No. of Instances	Fuzzy Information Gain	Tree Depth
balance	36.65	38.64	**35.00**	35.50	52.97	35.38
blood transfusion	77.25	77.25	77.61	77.24	77.19	**77.17**
breast cancer	7.21	7.23	**4.36**	6.10	4.96	5.48
breast tissue	33.86	33.70	**30.12**	30.26	33.93	34.41
bupa	41.38	41.41	41.39	**40.96**	42.04	41.99
cleveland	**81.99**	82.13	82.34	82.48	82.17	82.35
contraceptive	54.10	53.77	**51.47**	51.53	57.14	52.30
ecoli	30.91	29.15	**22.31**	23.79	29.09	26.32
glass	54.87	57.64	50.22	**49.01**	61.46	51.52
haberman	23.96	23.97	**22.85**	23.69	24.15	23.14
hayes-roth	44.63	42.67	**36.98**	37.60	52.45	45.16
heart	25.09	24.93	**23.57**	24.60	27.90	24.10
iris	7.05	6.95	**6.23**	6.80	6.46	7.13
newthyroid	10.65	9.35	**8.50**	9.17	8.86	9.86
sonar	27.66	27.46	**26.73**	27.19	44.54	27.55
tae	47.58	47.98	**46.33**	47.10	64.97	49.50
titanic	80.05	80.05	80.02	80.51	**78.69**	80.10
vehicle	47.86	48.58	**39.67**	39.97	63.96	42.47
wine	9.03	10.55	**6.21**	6.67	8.02	8.52
yeast	60.74	61.98	51.83	**51.67**	62.70	55.15
Average	40.13	40.27	**37.19**	37.59	44.18	38.98

Table 2. Summary Description for the Employed Datasets

As we can see from the results in Table 2, our proposed stopping criterion in most situations outperforms the other ones in terms of the overall ineffectiveness rate. Therefore, the fuzzy decision trees induced using the proposed criterion, in the average, are more compact and more accurate than FDTs constructed by the other stopping criteria. Also, the error values listed in Table 2 are statistically analyzed with the approach suggested by [10, 12] using the KEEL software [2]. Based on [10, 12], statistical comparison of classifiers over multiple datasets can be done in two steps. In the first step, the Friedman test is used to determine whether the classifiers are equivalent or not. If null-hypothesis is rejected, in the second step, a post-hoc test is used to compare the classifiers against each other. Table 3 shows the average ranking of each stopping criterion in the Friedman test. Friedman statistic which is distributed according to chi-square with 5 degrees of freedom is equal to 35.771429 For a=0.05, the Friedman test rejects the null hypothesis (that is all stopping criteria perform the same and the observed differences are merely random) and concludes that the different stopping criteria produce different classifiers.

Stopping Criteria	Ranking
NMGNI	1.65
Number of Instances	2.65
Tree Depth	3.75
Accuracy	4.2
Fuzzy Entropy	4.25
Fuzzy Information Gain	4.5

Table 3. Average Ranking of Stopping Criteria (Friedman Test) for Overall Ineffectiveness Rates

The Nemenyi's procedure is used as a post-hoc method and the obtained p-values for $\alpha = 0.05$ are shown in the Table 4. Nemenyi's procedure rejects those hypotheses that have a $p-value \leq 0.003333$. As it is apparent, the proposed stopping criterion (NMGNI) outperforms all the other stopping criteria except number of instances, and number of instances stopping criterion performs better than fuzzy information gain stopping criterion.

Stopping Criteria	p-value
NMGNI vs. Fuzzy Information Gain	0.000001
NMGNI vs. Fuzzy Entropy	0.000011
NMGNI vs. Accuracy	0.000016
NMGNI vs. Tree Depth	0.000386
Number of Instances vs. Fuzzy Information Gain	0.001766
Number of Instances vs. Fuzzy Entropy	0.006841
Number of Instances vs. Accuracy	0.008794
Number of Instances vs. Tree Depth	0.062979
NMGNI vs. Number of Instances	0.090969
Tree Depth vs. Fuzzy Information Gain	0.204894
Tree Depth vs. Fuzzy Entropy	0.398025
Tree Depth vs. Accuracy	0.446873
Accuracy vs. Fuzzy Information Gain	0.61209
Fuzzy Entropy vs. Fuzzy Information Gain	0.672604
Accuracy vs. Fuzzy Entropy	0.932647

Table 4. Post-hoc Comparison for $\alpha = 0.05$ for Overall Ineffectiveness Rates

We also analyzed the mean depth of rules generated by different stopping criteria in the same approach used for analyzing accuracy. In other words, 10CV procedure repeated 10 times for different number of membership functions (i.e. 2, 3, ..., 10) and mean depth of rules was calculated for all FDTs with less than 100 nodes. The obtained results are presented in Table 5 for different datasets.

Dataset	Accuracy	Fuzzy Entropy	Proposed Criterion	Number of Instances	Fuzzy Information Gain	Tree Depth
balance	3.61	3.36	3.23	3.19	**1.54**	2.96
blood transfusion	**3.17**	3.17	3.60	3.56	3.58	3.23
breast cancer	4.29	4.30	3.60	4.77	4.29	**3.15**
breast tissue	4.80	4.65	3.76	4.05	3.97	**3.30**
bupa	3.79	3.79	3.68	3.95	**2.01**	3.27
cleveland	**2.04**	3.46	3.84	4.08	2.69	3.41
contraceptive	3.65	4.02	3.57	3.73	**2.29**	3.24
ecoli	4.18	4.12	3.53	3.51	4.05	**3.18**
glass	4.28	4.26	3.58	4.12	3.37	**3.24**
haberman	2.54	2.54	3.18	3.28	**2.36**	2.84
hayes-roth	3.61	3.42	3.43	3.47	**2.43**	3.08
heart	4.45	4.46	4.05	4.21	4.33	**3.54**
iris	3.82	3.82	3.53	3.38	3.68	**3.08**
newthyroid	3.82	3.87	3.42	3.58	3.52	**3.07**
sonar	4.71	4.70	3.58	3.78	**1.74**	3.17
tae	3.72	3.74	3.69	3.76	**2.44**	3.15
titanic	3.20	3.20	3.13	3.16	3.14	**3.03**
vehicle	5.02	4.97	3.61	3.72	3.42	**3.22**
wine	4.45	4.61	3.52	3.55	4.13	**3.21**
yeast	4.53	4.32	3.64	3.87	3.30	**3.22**
Average	3.88	3.94	3.56	3.74	**3.11**	3.18

Table 5. Mean Depth of Rules for Different Stopping Criteria and Datasets

As it can be seen from Table 5, mean depth of rules generated by tree depth stopping criteria are less than the others in most cases. Friedman statistic, which is distributed according to chi-square with 5 degrees of freedom, is equal to 48.292857. Table 6 shows the average ranking of each stopping criteria and Table 7 shows the result of Nemenyis procedure as a post-hoc test for $\alpha = 0.05$. Nemenyis procedure rejects those hypotheses that have a $p - value \leq 0.003333$. As it can be seen from Table 7, both tree depth and fuzzy information gain stopping criteria, perform better than fuzzy entropy, accuracy and number of instances in terms of mean depth of generated trees.

Stopping Criteria	Ranking
Tree Depth	1.65
Fuzzy Information Gain	2.425
NMGNI	3.15
Number of Instances	4.25
Accuracy	4.75
Fuzzy Entropy	4.775

Table 6. Average Ranking of Stopping Criteria (Friedman Test) for Mean Depth of Rules

Stopping Criteria	p-value
Tree Depth vs. Fuzzy Entropy	0
Tree Depth vs. Accuracy	0
Tree Depth vs. Number of Instances	0.000011
Fuzzy Information Gain vs. Fuzzy Entropy	0.000071
Fuzzy Information Gain vs. Accuracy	0.000085
Fuzzy Information Gain vs. Number of Instances	0.002037
NMGNI vs. Fuzzy Entropy	0.006019
NMGNI vs. Accuracy	0.006841
Tree Depth vs. NMGNI	0.011230
NMGNI vs. Number of Instances	0.062979
Tree Depth vs. Fuzzy Information Gain	0.190200
Fuzzy Information Gain vs. NMGNI	0.220397
Number of Instances vs. Fuzzy Entropy	0.374857
Number of Instances vs. Accuracy	0.398025
Accuracy vs. Fuzzy Entropy	0.966293

Table 7. Post-hoc Comparison for a=0.05 for Mean Depth of Rules

6. Conclusions

Accuracy and size of induced FDT can be used to measure the power of FDT induction algorithms. Occam's razor principle of machine learning advises the induction of simplest and most accurate FDTs. Accordingly, a novel iterative approach (i.e. IDFID3), which employs one stopping criterion, for FDT induction has been proposed. In each iteration, IDFID3 determines a specific threshold value for a stopping criterion from the FDT constructed in previous iteration. IDFID3 controls the growth of FDT by dynamically setting the threshold value of stopping criterion. Despite the common FID3, which causes the tree's growth in a greedy manner like depth first search, the proposed IDFID3 tries to construct the FDT along the depth and width of it. In addition, a new stopping criterion (i.e. NMGNI) has been proposed that approximates the amount of usefulness of a node for expansion. Moreover, based on the proposed IDFID3 algorithm, a new method for comparing different stopping criteria independent of their threshold values has been proposed. Proposed stopping criterion and the previously developed ones are employed to construct FDTs using IDFID3 algorithm. The results obtained from experiments show that the proposed stopping criterion outperforms the other stopping criteria except the number of nodes and also the number of instances stopping criterion performs better than the fuzzy information gain stopping criterion in terms of classification accuracy and the number of nodes of induced FDT. Moreover, tree depth and fuzzy information gain stopping criteria, outperform the fuzzy entropy, accuracy and the number of instances in terms of mean depth of generated trees. Therefore, it is concluded that tree depth and fuzzy information gain result in producing more interpretable FDTs comparing to the other criteria.

Acknowledgements. The authors would like to thank M. Nikoupour, F. Ghafarizadeh and Z. Zeinalkhani for their valuable comments and suggestions.

References

[1] J. Alcal-Fdez, A. Fernandez, J. Luengo, J. Derrac, S. Garca, L. Snchez 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, **17** (2011), 255-287.

[2] J. Alcal-Fdez, L. Snchez, S. Garca, M. del Jesus, S. Ventura, J. Garrell, J. Otero, C. Romero, J. Bacardit, V. Rivas, J. Fernndez and F. Herrera, *KEEL: a software tool to assess evolutionary algorithms for data mining problems*, Soft Computing, **13(3)** (2009), 307-318.

[3] L. Bartczuk and D. Rutkowska, *Type-2 fuzzy decision trees*, In: Proceedings of the 9th International Conference on Artificial Intelligence and Soft Computing, Springer-Verlag, Zakopane, Poland, (2008), 197-206.

[4] R. B. Bhatt and M. Gopal, *Neuro-fuzzy decision trees*, International Journal of Neural Systems, **16(1)** (2006), 63-78.

[5] B. Chandra, P. Paul Varghese, *Fuzzy sliq decision tree algorithm*, IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, **38** (2008), 1294-1301.

[6] B. Chandra and P. Paul Varghese, *Fuzzifying gini index based decision trees*, Expert Systems with Applications, **36(4)** (2009), 8549-8559.

[7] P. C. Chang, C. Y. Fan and W. Y. Dzan, *A CBR-based fuzzy decision tree approach for database classification*, Expert Systems with Applications, **37(1)** (2010), 214-225.

[8] Y. L. Chen, T. Wang, B. S. Wang and Z. J. Li, *A survey of fuzzy decision tree classifier*, Fuzzy Information and Engineering, **1(2)** (2009), 149-159.

[9] M. E. Cintra, M. C. Monard and H. A. Camargo, *Evaluation of the pruning impact on fuzzy C4.5*, In: Anais Congresso Brasileiro de Sistemas Fuzzy, (2010), 257-264.

[10] J. Demar, *Statistical comparisons of classifiers over multiple data sets*, Journal of Machine Learning Research, **7** (2006), 1-30.

[11] A. Frank and A. Asuncion, *UCI machine learning repository*, http://archive.ics.uci.edu/ml, Irvine, 2010.

[12] S. Garcia and F. Herrera, *An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons*, Journal of Machine Learning Research, **9** (2008), 2677-2694.

[13] J. S. R. Jang, C. T. Sun and E. Mizutani, *Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence*, Prentice Hall, London, 1997.

[14] C. Z. Janikow, *Fuzzy decision trees: issues and methods*, IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics, **28(1)** (1998), 1-14.

[15] R. Jensen and Q. Shen, *Fuzzy-rough feature significance for fuzzy decision trees*, In: Proceedings of the 2005 UK Workshop on Computational Intelligence, (2005), 89-96.

[16] U. Khan, H. Shin, J. Choi and M. Kim, *wFDT - weighted fuzzy decision trees for prognosis of breast cancer survivability*, In: The Australasian Data Mining Conference, (2008), 141-152.

[17] R. E. Korf, *Depth-first iterative-deepening: an optimal admissible tree search*, In: G. R. Peter, ed., Expert Systems, IEEE Computer Society Press, (1990), 380-389.
[18] D. McNeill and P. Freiberger, *Fuzzy logic*, Simon & Schuster, New York, 1994.
[19] T. M. Mitchell, *Machine learning*, McGraw-Hill, New York, 1997.
[20] W. Pedrycz and Z. A. Sosnowski, *Designing decision trees with the use of fuzzy granulation*, IEEE Transactions on Systems, Man and Cybernetics-Part A: Systems and Humans, **30(2)** (2000), 151-159.
[21] W. Pedrycz and Z. A. Sosnowski, *Genetically optimized fuzzy decision trees*, IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics, **35(3)** (2005), 633-641.
[22] P. Pulkkinen and H. Koivisto, *Fuzzy classifier identification using decision tree and multi-objective evolutionary algorithms*, International Journal of Approximate Reasoning, **48(2)** (2008), 526-543.
[23] J. R. Quinlan, *C4.5: programs for machine learning*, Morgan Kaufmann Publishers, San Francisco, 1993.
[24] L. Rokach and O. Maimon, *Top-down induction of decision trees classifiers - a survey*, IEEE Transactions on Systems, Man and Cybernetics-Part C: Applications and Reviews, **35(4)** (2005), 476-487.
[25] L. Rokach and O. Maimon, *Data mining with decision trees: theroy and applications*, World Scientific, Singapore, 2008.
[26] J. Sanz, A. Fernandez, H. Bustince and F. Herrera, *IIVFDT: ignorance functions based interval-valued fuzzy decision tree with genetic tuning*, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, in press, 2012.
[27] M. Umano, H. Okamoto, I. Hatono, H. Tamura, F. Kawachi, S. Umedzu and J. Kinoshita, *Fuzzy decision trees by fuzzy ID3 algorithm and its application to diagnosis systems*, In: Proceedings of Third IEEE Conference on Fuzzy Systems, (1994), 2113-2118.
[28] T. Wang, Z. Li, Y. Yan and H. Chen, *A survey of fuzzy decision tree classifier methodology*, In: B. Y. Cao, ed., Fuzzy Information and Engineering, Springer, Berlin / Heidelberg, (2007), 959-968.
[29] X. Wang and C. Borgelt, *Information measures in fuzzy decision trees*, In: Proceedings of 13th IEEE International Conference on Fuzzy Systems, (2004), 85-90.
[30] X. Wang, B. Chen, G. Qian and F. Ye, *On the optimization of fuzzy decision trees*, Fuzzy Sets and Systems, **112(1)** (2000), 117-125.
[31] X. Wang and H. Jiarong, *On the handling of fuzziness for continuous-valued attributes in decision tree generation*, Fuzzy Sets and Systems, **99(3)** (1998), 283-290.
[32] Y. Yuan and M. J. Shaw, *Induction of fuzzy decision trees*, Fuzzy Sets and Systems, **69(2)** (1995), 125-139.
[33] M. Zeinalkhani and M. Eftekhari, *A new measure for comparing stopping criteria of fuzzy decision tree*, In: International Conference on Computer and Knowledge Engineering, Mashhad, (2011), 120-123.
[34] H. Zhang and B. H. Singer, *Recursive partitioning and applications*, 2nd ed., Springer, London, 2010.

Mohsen Zeinalkhani*, Department of Computer Engineering, Shahid Bahonar University of Kerman, Kerman, Iran
E-mail address: zeinalkhani@gmail.com

Mahdi Eftekhari, Department of Computer Engineering, Shahid Bahonar University of Kerman, Kerman, Iran
E-mail address: m.eftekhari@mail.uk.ac.ir

*Corresponding author