Information Sciences
Volume 465, October 2018, Pages 425-438

Tolerance rough fuzzy decision tree

Junhai Zhai [a], Xizhao Wang [b], Sufang Zhang [c], Shaoxing Hou [d]

Show more

https://doi.org/10.1016/j.ins.2018.07.006

Get rights and content

Abstract

Fuzzy decision tree (FDT) is an extension of decision tree. Fuzzy classification rules can be extracted by FDT from fuzzy decision tables with fuzzy conditional attributes and fuzzy decision attributes. However, it is very time consuming for fuzzifying conditional attributes, and fuzzification of conditional attributes will inevitably lead to information loss. In order to deal with this problem, based on tolerance rough fuzzy set, this paper proposed an algorithm named TRFDT (Tolerance Rough Fuzzy Decision Tree) and theoretically proved that the proposed algorithm is convergent with a very large probability. TRFDT can directly handle fuzzy decision tables with continuous-valued conditional attributes and fuzzy decision attributes. Accordingly, TRFDT has fast learning speed and good generalization ability, which have been experimentally proved by comparing TRFDT with two state-of-the-art approaches fuzzy ID3 and FDT-YS.

Previous article

Next article

Keywords

Fuzzy decision tree; Rough set; Tolerance rough set; Rough fuzzy set; Tolerance rough fuzzy set

1. Introduction

Fuzzy decision tree (FDT) [29] is an extension of decision tree [24]. Fuzzy classification rules can be extracted by FDT from fuzzy decision tables with fuzzy conditional attributes and fuzzy decision attributes. Because fuzzy classification systems based on fuzzy decision tree are generally robust, FDT have been widely and successfully applied to many fields [25], such as decision-making [9], [50], [51], classification and prediction [38], [45], [49], biological information processing [15], [16], etc. Fuzzy ID3 [29] and the fuzzy decision tree algorithm proposed by Yuan and Shaw [39] (denoted by FDT-YS for convenience in this paper) are two famous fuzzy decision tree algorithms. Fuzzy ID3 was directly extended from ID3 [24] by replacing information entropy with average fuzzy classification entropy to select expanded attributes. Different from fuzzy ID3, FDT-YS [39] employed ambiguity as heuristic to select expanded attributes, the termination condition for leaf node is same as fuzzy ID3. Along with this technology route, many fuzzy decision tree algorithms were proposed by different researchers. Based on fuzzy rough set, Zhai [43] proposed a fuzzy rough decision tree algorithm which cleverly combined the roughness of knowledge and fuzziness of data. Compared with fuzzy ID3, the testing accuracy can be improved by this algorithm. Olaru and Wehenkel [21] proposed a novel fuzzy decision tree named soft decision tree which combined tree-growing and pruning, to determine the structure of the soft decision tree, refitting and backfitting techniques were used to improve the generalization capability of soft decision tree. Jang [12] extended CART algorithm to fuzzy environment, and proposed fuzzy CART. The fuzzy CART can produce more comprehensible fuzzy rules. Lertworaprachaya et al. [17] extended the fuzzy ID3 algorithm to interval-valued conditional attributes. They represented fuzzy membership values as intervals to model uncertainty and employed the look-ahead based fuzzy decision tree induction method to construct decision tree. They also investigated the significance of different neighbourhood values and defined a new parameter insensitive to specific data sets using fuzzy sets. Zaitseva and Levashenko [41] proposed a FDT based method for construction of structure function, which was used to represent the correlation of the system performance level and the states of its components. Based on Hadoop MapReduce, Segatori et al. [26] proposed a distributed learning scheme for generating fuzzy decision trees, the proposed scheme is suitable for managing big data sets even with a modest commodity hardware support.

Except these algorithms, there are also some other algorithms proposed from different technology routes. For instance, Based on axiomatic fuzzy sets, Liu et al. [19] proposed a fuzzy decision tree algorithm for extracting fuzzy classification rules, which can be applied to data sets with mixed data type attributes. Zeinalkhani and Eftekhari [42] proposed a two steps algorithm for constructing fuzzy decision tree. In the first step, discretization divides domain of continuous attributes to several partitions, and then, in the second step, an fuzzy membership degree is defined on each partition. Finally a fuzzy decision tree is generated. Wang et al. [34] proposed fuzzy rule based decision trees. In contrast with traditional axis-parallel decision trees in which only a single feature is taken into account at each node, each node of the proposed decision trees involves multiple features. Pedrycz and Sosnowski [23] proposed a clustering fuzzy decision tree algorithm, which introduced the idea of granular computing into the process of induction of fuzzy decision tree. Based on case-based reasoning, Chang et al. [2] proposed a fuzzy decision tree algorithm and they applied the proposed algorithm to data classification. Wang et al. [32] proposed an algorithm for generating fuzzy decision trees with carefully selected samples, the proposed algorithm can significantly improve the generalization ability of fuzzy decision trees. Wang et al. [31] conducted a further investigation on the relationship between the uncertainty and generalization ability of fuzzy learning algorithms, and obtained very valuable conclusion: the classifier with higher uncertainty outputs has better performance for complex boundary problems. Recently, Wang [35] edited a special issue on learning with uncertainty. In addition, Wang et al. also studied the applications of uncertainty in machine learning. For instance, fuzziness based semi-supervised learning and its application in intrusion detection was studied in [1]. Fuzziness based nonlinear regression was studied in [10]. OWA operator based link prediction ensemble and its application in social network data mining was investigated in [8]. Fuzzy rough set based feature selection methods were investigated in [33], [36], [48]. A comparative study on heuristic algorithms for generating fuzzy decision trees was given in [30], an excellent survey of fuzzy decision tree algorithms can be found in [13].

These algorithms mentioned above can only induce fuzzy decision trees from fuzzy decision tables with fuzzy-valued conditional attributes and fuzzy-valued decision attribute. When these algorithms are applied to fuzzy decision tables with continuous-valued conditional attributes and fuzzy-valued decision attribute, it is inevitable for these algorithms to fuzzify the continuous-valued conditional attributes, but it is difficult to determine the fuzzy membership degree. Based on our previous work [46], in this paper, an induction algorithm of fuzzy decision tree named TRFDT is proposed, TRFDT can directly induce fuzzy decision tree from fuzzy decision tables with continuous-valued conditional attributes and fuzzy-valued decision attribute. The degree of tolerance rough fuzzy dependency [46], [47] is employed to select expanded attributes, the Luca-Termini fuzzy entropy [20] is employed to select the optimal cut, and the Kosko fuzzy entropy [14] is used as termination condition for leaf nodes. We theoretically proved that the proposed algorithm is convergent with a very large probability. The experimental results and analysis including statistical analysis verified that the proposed algorithm TRFDT is effective and efficient.

The remainder of this paper is structured as follows. Section 2 presents the preliminaries, including rough set, tolerance rough set, rough fuzzy set. Tolerance rough fuzzy set and tolerance rough fuzzy decision tree are presented in Section 3. The theoretical proof of the convergence of the proposed algorithm is also presented in this section. An example is provided in Section 4 to illustrate the generation of tolerance rough fuzzy decision trees by the proposed algorithm. Experimental results and statistical analysis are given in Sections 5 and6 concludes this paper.

2. Preliminaries

In this section, we briefly review preliminaries including rough set [22], tolerance rough set [27], rough fuzzy set [4]. Some new extended models of rough set can be found in [5], [6], [7], [28],

[37].

2.1. Rough set

In this paper, we discuss problems in the framework of classification. Let $DT = (U, A \cup C)$ be a decision table with symbolic-valued conditional attributes. $U = \{x_1, x_2, \ldots, x_n\}$, $A = \{a_1, a_2, \ldots, a_d\}$, the instances in U are categorized into k classes: $C_1, C_2, \ldots, C_k$, i.e., $U/C = \{C_1, C_2, \ldots C_k\}$. Let $x \in U$ and R is an equivalence relation induced by a subset of A, the equivalence class containing x is given by:

$$[x]_R = \{y|xRy\}. \tag{1}$$

Given a decision table, for arbitrary target concept $C_i \in U/C (1 \le i \le k)$, the lower approximation and the upper approximation of C_i with respect to R are defined by

$$\underline{R}(C_i) = \{[x]_R | [x]_R \subseteq C_i\}. \tag{2}$$

and

$$\overline{R}(C_i) = \{[x]_R | [x]_R \cap C_i \neq \phi\}. \tag{3}$$

Given an equivalence relation $R \subseteq A$, the positive region of C with respect to R is defined by (4).

$$POS_R(C) = \bigcup_{C_i \in U/C} \underline{R}(C_i). \tag{4}$$

The degree of dependency of C with respect to R is given by (5).

$$\gamma_R(C) = \frac{|POS_R(C)|}{|U|} = \frac{\left|\bigcup_{C_i \in U/C} \underline{R}(C_i)\right|}{|U|}. \tag{5}$$

2.2. Tolerance rough set

Tolerance rough set (TRS) [27] is extended from rough set by replacing equivalence relation with similarity relation. The target concept is same as in rough set.

Given a decision table $DT = (U, A \cup C)$,R is a similarity relation defined on U, if and only if R satisfies the following conditions:

(1) Reflexivity, i.e., for each $x \in U$, xRx;

(2) Symmetry, i.e., for each $x, y \in U$, xRy, and yRx.

We can define many similarity relations on U, such as the definitions given by (6)–(8).

$$R_a(x, y) = 1 - \frac{|a(x) - a(y)|}{|a_{max} - a_{min}|}. \tag{6}$$

$$R_a(x, y) = exp\left(-\frac{(a(x) - a(y))^2}{2{\sigma_a}^2}\right). \tag{7}$$

$$R_a(x,y) = max\left\{min\left\{\frac{a(y)-(a(x)-\sigma_a)}{a(x)-(a(x)-\sigma_a)}, \frac{(a(x)+\sigma_a)-a(y)}{(a(x)+\sigma_a)-a(x)}\right\}, 0\right\}. \tag{8}$$

In (6)–(8), $a \in A$, $x \in U$, and a_{max} and a_{min} denote the maximum and minimum of a respectively, σ_a is variance of attribute a.

Given a similarity threshold τ, for $\forall x \in U$ and $\forall R \subseteq A$, we can define the similarity class of x as follows.

$$[x]_{R_\tau} = \left\{y \middle| (y \in U) \wedge \left(\frac{\sum_{a \in R} R_a(x,y)}{|R|} \geq \tau\right)\right\}. \tag{9}$$

Given a decision table and a similarity threshold τ, for arbitrary target concept C_i, the tolerance lower approximation and the tolerance upper approximation of C_i with respect to R are defined by

$$\underline{R_\tau}(C_i) = \left\{x \middle| (x \in U) \wedge \left([x]_{R_\tau} \subseteq C_i\right)\right\}, \tag{10}$$

and

$$\overline{R_\tau}(C_i) = \left\{x \middle| (x \in U) \wedge \left([x]_{R_\tau} \cap C_i \neq \phi\right)\right\}. \tag{11}$$

2.3. Rough fuzzy set

Rough fuzzy set (RFS) [4] is an extension of rough set. In RFS, the used knowledge is equivalence relation, while the target concept is a fuzzy set. Without loss of generality, we also use $C_i(1 \leq i \leq k)$ to denote the k fuzzy target concepts, which are k fuzzy sets. The corresponding decision table is called fuzzy decision table.

Given a fuzzy decision table, for arbitrary fuzzy target concept $C_i(1 \leq i \leq k)$. The rough fuzzy lower approximation and the rough fuzzy upper approximation of C_i with respect to equivalence relation R are defined by

$$\underline{R}(C_i) = \mu_{\underline{R}(C_i)}([x]_R) = inf\{\mu_{C_i}(y) | y \in [x]_R\} \tag{12}$$

and

$$\overline{R}(C_i) = \mu_{\overline{R}(C_i)}([x]_R) = sup\{\mu_{C_i}(y) | y \in [x]_R\} \tag{13}$$

According to the fuzzy extension principle [40], (12) and (13) can be equivalently written as follows.

$$\underline{R}(C_i) = \mu_{\underline{R}(C_i)}(x) = inf\left\{\max_{y \in U}\{\mu_{C_i}(y), 1-\mu_R(x,y)\}\right\} \tag{14}$$

and

$$\overline{R}(C_i) = \mu_{\overline{R}(C_i)}(x) = sup\left\{\min_{y \in U}\{\mu_{C_i}(y), \mu_R(x,y)\}\right\} \tag{15}$$

3. Tolerance rough fuzzy set and tolerance rough fuzzy decision tree

In this section, we first introduce the tolerance rough fuzzy set (TRFS) which is proposed in our previous work [47], and then based on TRFS, we present the proposed algorithm TRFDT, finally we give a theorem and its proof. The theorem ensures that TRFDT is convergent with a very large probability.

3.1. Tolerance rough fuzzy set

TRFS is extended from RFS by replacing equivalence relation with similarity relation. TRFS can deal with the fuzzy decision table with continuous-valued conditional attributes and fuzzy-valued decision attribute. Given a fuzzy decision table $DT = (U, A \cup C)$,R is a similarity relation defined on U, τ is a similarity threshold. For $\forall C_i \in U/C (1 \leq i \leq k)$, the tolerance rough fuzzy lower approximation and tolerance rough fuzzy upper approximation of C_i with respect to R are defined by

$$\underline{R_\tau}\left(C_i\right) = \mu_{\underline{R_\tau}(C_i)}\left([x]_{R_\tau}\right) = inf\left\{\mu_{C_i}\left(y\right)\middle|y \in [x]_{R_\tau}\right\}, \tag{16}$$

and

$$\overline{R_\tau}\left(C_i\right) = \mu_{\overline{R_\tau}(C_i)}\left([x]_{R_\tau}\right) = sup\left\{\mu_{C_i}\left(y\right)\middle|y \in [x]_{R_\tau}\right\}. \tag{17}$$

Similarly to (14) and (15), we have the following equivalent definitions:

$$\begin{aligned}\underline{R_\tau}\left(C_i\right) &= \mu_{\underline{R_\tau}(C_i)}\left(x\right)\\ &= inf\left\{\max_{y\in U}\left\{\left(\mu_{C_i}\left(y\right), 1-\mu_{R_\tau}\left(x,y\right)\right\}\right\}\right.,\end{aligned} \tag{18}$$

and

$$\begin{aligned}\overline{R_\tau}\left(C_i\right) &= \mu_{\overline{R_\tau}(C_i)}\left(x\right)\\ &= sup\left\{\min_{y\in U}\left\{\mu_{C_i}\left(y\right), \mu_{R_\tau}\left(x,y\right)\right\}\middle|y \in U\right\}.\end{aligned} \tag{19}$$

3.2. Tolerance rough fuzzy decision tree

In this section, we first introduce the heuristics for selecting the expanded attribute and the optimal cut, and then discuss terminal condition of the proposed algorithm.

Given a fuzzy decision table $DT = (U, A \cup C)$ and a similarity threshold τ. For $\forall a_i \in A$ and $\forall C_j \in U/C (1 \leq j \leq k)$, the degree of dependency of C_j with respect to a_i is given by (20).

$$\gamma_{(a_i)_\tau}\left(C_j\right) = \frac{\sum_{x\in U}\mu_{(a_i)_\tau(C_j)}(x)}{|U|}. \tag{20}$$

The degree of dependency of C with respect to a_i is given by (21).

$$\gamma_{(a_i)_\tau}\left(C\right) = \frac{1}{k}\sum_{j=1}^{k}\gamma_{(a_i)_\tau}\left(C_j\right). \tag{21}$$

The heuristic for selecting the expanded attributes is defined by (22).

$$a^* = argmax_{a_i \in A} \left\{ \gamma_{(a_i)_\tau} (C) \right\}. \tag{22}$$

For $\forall C_j \in U/C (1 \leq j \leq k)$, the Luca-Termini fuzzy entropy [20] of C_j is defined by (23).

$$H_{LucaT} (C_j) = -\frac{1}{|U|} \sum_{x \in U} \left[\mu_{C_j} (x) \, log_2 \mu_{C_j} (x) + \left(1 - \mu_{C_j} (x)\right) log_2 \left(1 - \mu_{C_j} (x)\right) \right]. \tag{23}$$

The heuristic for selecting the optimal cut of the expanded attribute is defined by (24).

$$H_{LucaT} (C) = \frac{1}{k} \sum_{j=1}^{k} H_{LucaT} (C_j). \tag{24}$$

The proposed algorithm TRFDT is given in Algorithm 1.

Algorithm 1. TRFDT: Tolerance Rough Fuzzy Decision Tree.

Input: Fuzzy decision table $DT = (U, A \cup C)$, $U = \{x_1, x_2, \cdots, x_n\}$, $A = \{a_1, a_2, \cdots, a_d\}$, $C = \{C_1, C_2, \ldots, C_k\}$, $C_j (1 \leq j \leq k)$ are k fuzzy sets, a similarity threshold τ and a parameter λ.
Output: A fuzzy decision tree.

1 **for** $(i = 1; i \leq n; i = i + 1)$ **do**
2   **for** $(j = 1; j \leq k; j = j + 1)$ **do**
3   Calculate the degree of dependency of C_j with respect to a_i by (20);
4   **end**
5 **end**
6 **for** $(i = 1; i \leq n; i = i + 1)$ **do**
7   Calculate the degree of dependency of C with respect to a_i by (21);
8 **end**
9 Select expanded attribute a^* by (22);
10 Select the optimal cut of the expanded attribute a^* by (24);
11 Partition data set U into two subsets U_1 and U_2;
12 **for** $(j = 1; j \leq k; j = j + 1)$ **do**
13   Calculate the Kosko fuzzy entropy $H_{Kosko}(C_j)$ on U_1 and U_2;
14 **end**
15 **if** $(H_{Kosko}(C_j) \leq \lambda)$ **then**
16   Generate a leaf node marked with C_j;
17 **else**
18   Repeat 1 to 18;
19 **end**
20 Output a fuzzy decision tree.

It is important for setting or selecting an appropriate terminal condition in the process of generating fuzzy decision tree. An unsuitable terminal condition may result in large fuzzy decision trees with over-fitting or small fuzzy decision trees with under-fitting. In this paper, we employs Kosko fuzzy entropy [14] of a fuzzy decision class $C_j (1 \leq j \leq k)$ given by (25) as the terminal condition for leaf nodes. For a node, if the Kosko fuzzy entropy of fuzzy decision class C_j is less than a predefined threshold λ, then the node becomes a leaf node.

$$H_{Kosko} (C_j) = \frac{\sum_{x \in U} min\left\{\mu_{C_j}(x), \mu_{C_j^c}(x)\right\}}{\sum_{x \in U} max\left\{\mu_{C_j}(x), \mu_{C_j^c}(x)\right\}}. \tag{25}$$

Remarks:

(1) Other fuzzy entropy can also used as the terminal condition.

(2) Because a node of a fuzzy decision tree corresponds a subset of instances. Consequently, a node and its corresponding subset of instances can be view as equivalent.

Given a fuzzy decision class C_j, without loss of generality, we denote the fuzzy entropies of C_j on a node U_1 and on its parent node U by $H(U_1)$ and $H(U)$ respectively. Regarding $H(U)$ and $H(U_1)$, we have the following theorem which ensures that the proposed algorithm is convergent with a very large probability.

Theorem 1

Let $U_1 \subseteq U$, $|U| = n, |U_1| = n_1$. *If* $n \geq 2n_1$, *then* $E(U) \leq \frac{n_1}{n} E(U_1)$ *holds with a probability greater than* $1 - \left(\frac{1}{2}\right)^{n-n_1} + \left(\frac{1}{4}\right)^{n-n_1}$.

For the sake of simplicity, we denote the fuzzy membership degree of x_i *belong to a fuzzy decision class by* μ_i, *let* $y_i = min\{\mu_i, 1-\mu_i\}$. *without loss of generality, we suppose that* y_i *follows uniform distribution on interval* (0.1, 0.5). *Hence, we have,*

$$H(U) = \frac{y_1+y_2+\cdots+y_n}{n-(y_1+y_2+\cdots+y_n)}. \tag{26}$$

and

$$\frac{n_1}{n} H(U_1) = \frac{n_1}{n} \times \frac{y_1+y_2+\cdots+y_{n_1}}{n_1-\left(y_1+y_2+\cdots+y_{n_1}\right)}. \tag{27}$$

Under these assumptions, theorem 1 is equivalent to the following Theorem 2

Theorem 2

Suppose that y_i *follows uniform distribution on interval* (0.1, 0.5) *and* $n \geq 2n_1$. *Then the following inequality* (28) *holds with a probability greater than* $1 - \left(\frac{1}{2}\right)^{n-n_1} + \left(\frac{1}{4}\right)^{n-n_1}$.

$$\frac{n(y_1+y_2+\cdots+y_n)}{n-(y_1+y_2+\cdots+y_n)} \geq \frac{n_1\left(y_1+y_2+\cdots+y_{n_1}\right)}{n_1-\left(y_1+y_2+\cdots+y_{n_1}\right)} \tag{28}$$

Proof

Let,

$$\begin{aligned} S(y_1, y_2, \cdots, y_n) &= nn_1\left(y_1+y_2+\cdots+y_n\right) - n\left(y_1+y_2+\cdots+y_{n_1}\right)\left(y_1+y_2+\cdots\right. \\ &\quad - nn_1\left(y_1+y_2+\cdots+y_{n_1}\right) + n_1\left(y_1+y_2+\cdots+y_{n_1}\right)\left(y_1+y_2+\right. \\ &= nn_1\left(y_{n_1}+y_{n_2}\cdots+y_n\right) - (n-n_1)\left(y_1+y_2+\cdots+y_{n_1}\right)\left(y_1+\cdots\right. \end{aligned}$$

Compute derivatives of $S(y_1, y_2, \ldots, y_n)$ with respect to $y_1, y_2, \ldots, y_n$, we have,

$$\frac{\partial S}{\partial y_1} = -\left(n - n_1\right)\left(y_1 + y_2 + \cdots + y_n\right) - \left(n - n_1\right)\left(y_1 + y_2 + \cdots + y_{n_1}\right) < 0.$$

Similarly, we have,

$$\frac{\partial S}{\partial y_2} < 0, \cdots, \frac{\partial S}{\partial y_{n_1}} < 0.$$

While,

$$\begin{aligned}\frac{\partial S}{\partial y_{n_1+1}} &= nn_1 - \left(n - n_1\right)\left(y_1 + y_2 + \cdots + y_{n_1}\right) > nn_1 - \left(n - n_1\right) \times 0.5n_1 \\ &= n_1\left(n - 0.5n + 0.5n_1\right) = n_1\left(0.5n + 0.5n_1\right) = 0.5n_1\left(n + n_1\right) > 0.\end{aligned}$$

Similarly, we have,

$$\frac{\partial S}{\partial y_{n_1+2}} > 0, \frac{\partial S}{\partial y_{n_1+3}} > 0, \cdots, \frac{\partial S}{\partial y_n} > 0.$$

For $n > 2n_1$, let

$$\begin{aligned}&y_1 = y_2 = \cdots = y_{n_1} = 0.5; \\ &y_{n_1+1} + y_{n_1+2} + \cdots + y_n = 0.2\left(n - n_1\right).\end{aligned}$$

Then, we have,

$$S\left(y_1, y_2, \cdots, y_n\right) = 0.1nn_1 - 0.2n_1^2 > 0.2n_1^2 - 0.2n_1^2 = 0.$$

Hence, in order to prove Theorem 2, only to prove that $y_{n_1+1} + y_{n_1+2} + \cdots + y_n > 0.2\left(n - n_1\right)$ holds with probability greater than $1 - \left(\frac{1}{2}\right)^{n-n_1} + \left(\frac{1}{4}\right)^{n-n_1}$.

While

$$\begin{aligned}&P(y_{n_1+1} + y_{n_1+2} + \cdots + y_n \geq 0.2\left(n - n_1\right)) \\ &= 1 - P(y_{n_1+1} + y_{n_1+2} + \cdots + y_n < 0.2\left(n - n_1\right)).\end{aligned}$$

where $0.1 \le y_i \le 0.5$.

$$\begin{aligned}&P(y_{n_1+1} + \cdots + y_n < 0.2\left(n - n_1\right)) \\ &= \int\int\cdots\int_{y_{n_1+1}+\cdots+y_n<0.2(n-n_1)} f\left(y_{n_1+1} + \cdots + y_n\right) dy_{n_1+1}\cdots dy_n.\end{aligned}$$

where $f(y_{n_1+1} + \cdots + y_n)$ is joint probability density function.

Let A_i and B_i are two random events $y_i \in (0.1, 0.2)$ and $y_i \in (0.2, 0.3)$, $i = n_1 + 1, n_1 + 2, \ldots, n$. We have

$$
\begin{aligned}
&P(y_{n_1+1}+y_{n_1+2}+\cdots+y_n<0.2\,(n-n_1))\\
&<P\left(A_{n_1+1}A_{n_1+2}\cdots A_n\right)+P\left(A_{n_1+1}\cdots A_{n_1+j-1}B_{n_1+j}A_{n_1+j+1}\cdots A_n\right)\\
&+P(A_{n_1+1}\cdots A_{n_1+i-1}B_{n_1+i}A_{n_1+i+1}\cdots A_{n_1+j-1}B_{n_1+j}A_{n_1+j+1}\cdots A_n)+\cdots\\
&+P(B_{n_1+1}\cdots B_{n_1+j-1}A_{n_1+j}B_{n_1+j+1}\cdots B_n)\\
&\quad=\frac{1}{4^{n-n_1}}+C_{n-n_1}^{1}\frac{1}{4^{n-n_1}}+\cdots+C_{n-n_1}^{n-n_1-1}\frac{1}{4^{n-n_1}}\\
&\quad=\left(\frac{1}{4}+\frac{1}{4}\right)^{n-n_1}-\frac{1}{4^{n-n_1}}=\left(\frac{1}{2}\right)^{n-n_1}-\frac{1}{4^{n-n_1}}.
\end{aligned}
$$

Consequently,

$$P\left(y_{n_1+1}+y_{n_1+2}+\cdots+y_n\geq 0.2\,(n-n_1)\right)>1-\left(\tfrac{1}{2}\right)^{n-n_1}+\tfrac{1}{4^{n-n_1}}.$$

Hence, Theorem 2 holds. □

Remarks:

(1) Regarding the partition of set of instances, generally inequality $n \geq 2n_1$ holds.

(2) Inequality (28) is not always true, it holds with a probability. For example, let the fuzzy membership degrees of x_1 and x_2 are (0.15, 0.85) and (0.45, 0.55), respectively, we have

$$\tfrac{2(0.45+0.15)}{2-(0.45+0.15)}=\tfrac{2}{3}<\tfrac{0.45}{1-0.45}=\tfrac{9}{11}.$$

In other words, inequality (28) doesn't hold.

In the following, the computations of probabilities for two cases $n-n_1=2$ and $n-n_1=3$ are given.

(1) For the case $n-n_1=2,$ we have,

$$P\left(0.2\leq y_{n_1+1}+y_{n_1+2}<0.2*2\right)=\tfrac{\frac{1}{2}\times0.2\times0.2}{0.4^2}=\tfrac{1}{8}.$$

Hence,

$$P\left(y_{n_1+1}+y_{n_1+2}\geq 0.2*2\right)=1-\tfrac{1}{8}=\tfrac{7}{8}.$$

while,

$$1-\left(\tfrac{1}{2}\right)^2+\left(\tfrac{1}{4}\right)^2=\tfrac{13}{16}<\tfrac{14}{16}.$$

The error rate is $\left(\tfrac{14}{16}-\tfrac{13}{16}\right)/\tfrac{14}{16}=\tfrac{1}{14}.$

(2) For the case $n-n_1=3,$ we have,

$$\begin{aligned}
&P(0.3 \leq y_{n_1+1} + y_{n_1+2} + y_{n_1+3} < 0.2 \times 3) \\
&= P(0.3 \leq y_{n_1+1} + y_{n_1+2} + y_{n_1+3} < 0.6) \\
&= \frac{\int_{0.1}^{0.5} dy_{n_1+1} \int_{0.1}^{0.6-y_{n_1+1}} dy_{n_1+2} \int_{0.1}^{0.6-y_{n_1+1}-y_{n_1+2}} dy_{n_1+3}}{0.5^3 - 3 \times 0.1^2 \times 0.5 + 2 \times 0.1^3} \\
&= \frac{\int_{0.1}^{0.5} dy_{n_1+1} \int_{0.1}^{0.6-y_{n_1+1}} dy_{n_1+2} \int_{0.1}^{0.6-y_{n_1+1}-y_{n_1+2}} dy_{n_1+3}}{0.5^3 - 3 \times 0.1^2 \times 0.5 + 2 \times 0.1^3} \\
&\quad - \frac{\int_{0.1}^{0.3} dy_{n_1+1} \int_{0.1}^{0.3-y_{n_1+1}} dy_{n_1+2} \int_{0.1}^{0.3-y_{n_1+1}-y_{n_1+2}} dy_{n_1+3}}{0.5^3 - 3 \times 0.1^2 \times 0.5 + 2 \times 0.1^3} \\
&= \frac{\frac{1}{3} \times \frac{1}{2} \times 0.4 \times 0.4 \times 0.4 - \frac{1}{3} \times \frac{1}{2} \times 0.2 \times 0.2 \times 0.2}{0.5^3 - 3 \times 0.1^2 \times 0.5 + 2 \times 0.1^3} = \frac{7}{84}.
\end{aligned}$$

Hence,

$$P\left(y_{n_1+1} + y_{n_1+2} + y_{n_1+3} \geq 0.6\right) = 1 - \frac{7}{84} = \frac{77}{84}.$$

The error rate is $\left(\frac{77}{84} - \frac{57}{64}\right) / \frac{77}{84} = \frac{5}{176}$.

Similarly, for the case $n - n_1 = 4$, we can obtain that the inequality (28) holds with a probability greater than $1 - \frac{1}{2^4} + \frac{1}{4^4} = \frac{241}{256} = 0.94$.

Based on the above computations, we can find that the bigger the value of $n - n_1$, the bigger the corresponding probability is. Generally, in the process of generation of fuzzy decision tree, the values of $n - n_1$ are much greater than 4. In other words, the inequality (28) holds with a very large probability. This conclusion provides a theoretical support for the proposed algorithm TRFDT, which ensures that TRFDT is convergent with a very large probability (i.e., $1 - \left(\frac{1}{2}\right)^{n-n_1} + \left(\frac{1}{4}\right)^{n-n_1}$).

4. An example

In this section, we provide an example to illustrate the generation of tolerance fuzzy rough decision trees by the proposed algorithm.

Example: Given a fuzzy decision table with 6 instances (see Table 1). In Table 1, there are three continuous-valued conditional attributes a_1, a_2, a_3 and two fuzzy decision classes, i.e., two fuzzy target concept C_1, C_2. In the following, the process of constructing fuzzy decision tree by TRFDT are given as follows.

(1) Select the expanded attribute

We first compute the degree of dependency of C_1 with respect to a_1 by (20). According to (18), we have,

$$\underline{a_1}\left(C_1\right) = \mu_{\underline{a_1}(C_1)}\left(x\right) = inf\{\max_{y \in U}\{(\mu_{C_1}\left(y\right), 1 - \mu_{a_1}\left(x, y\right)\}\},$$

where $\mu_{a_1}(x,y)$ is computed by (6).

For $x = x_1$, we have,

$$
\begin{aligned}
\mu_{\underline{a_1}(C_1)}(x_1) &= inf\left\{\underset{y\in U}{max}\left\{\mu_{C_1}(y), 1-\mu_{a_1}(x_1,y)\right\}\right\} \\
&= inf\{max\{\mu_{C_1}(x_1), 1-\mu_{a_1}(x_1,x_1)\}, \\
&\quad max\{\mu_{C_1}(x_2), 1-\mu_{a_1}(x_1,x_2)\}, \\
&\quad max\{\mu_{C_1}(x_3), 1-\mu_{a_1}(x_1,x_3)\}, \\
&\quad max\{\mu_{C_1}(x_4), 1-\mu_{a_1}(x_1,x_4)\}, \\
&\quad max\{\mu_{C_1}(x_5), 1-\mu_{a_1}(x_1,x_5)\}, \\
&\quad max\{\mu_{C_1}(x_6), 1-\mu_{a_1}(x_1,x_6)\}\} \\
&= inf\{max\{0.14, 0.00\}, max\{0.92, 0.12\}, max\{0.19, 0.12\}, \\
&\quad max\{0.83, 0.87\}, max\{0.80, 0.75\}, max\{0.12, 0.62\}\} \\
&\quad inf\{0.14, 0.92, 0.19, 0.87, 0.80, 0.62\} = 0.14.
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
\mu_{\underline{a_1}(C_1)}(x_2) &= 0.14, \ldots \\
\mu_{\underline{a_1}(C_1)}(x_4) &= 0.25, \mu_{\underline{a_1}(C_1)}(x_5) = 0.12, \mu_{\underline{a_1}(C_1)}(x_6) = 0.12.
\end{aligned}
$$

According to (20), the degree of dependency of C_1 with respect to a_1 can be obtained.

$$
\begin{aligned}
\gamma_{a_1}(C_1) &= \frac{1}{|U|}\sum_{x\in U}\mu_{a_1(C_1)}(x) \\
&= \frac{0.14+0.14+0.14+0.25+0.12+0.12}{6} = 0.15.
\end{aligned}
$$

Similarly, the degree of dependency of C_2 with respect to a_1 can be obtained.

$$
\begin{aligned}
\gamma_{a_1}(C_2) &= \frac{1}{|U|}\sum_{x\in U}\mu_{a_1(C_2)}(x) \\
&= \frac{0.12+0.08+0.25+0.17+0.17+0.20}{6} = 0.17.
\end{aligned}
$$

Hence, we can obtain the degree of dependency of C with respect to a_1 by (21) as follows.

$$\gamma_{a_1}(C) = \frac{1}{2}\sum_{j=1}^{2}\gamma_{a_1}(C_j) = \frac{0.15+0.17}{2} = 0.16.$$

Similarly, we have,

$$\gamma_{a_2}(C) = 0.15, \gamma_{a_3}(C) = 0.18.$$

Obviously, attribute a_3 has the maximum, it is selected as expanded attribute, i.e., the root node of the fuzzy decision tree.

(2) Select the optimal cut

Next we select the optimal cut of the expanded attribute a_3. We sort the instances in Table 1 by a_3, it is easy to find that there are 5 cuts which are $t_1 = \frac{0.4+0.6}{2} = 0.50$,$t_2 = \frac{0.6+0.7}{2} = 0.65$, $t_3 = \frac{0.7+0.8}{2} = 0.75$,$t_4 = \frac{0.8+0.9}{2} = 0.85$ and $t_5 = \frac{0.9+1.2}{2} = 1.05$.

The cut t_1 partition U into U_1 and U_2, where $U_1 = \{x_1\}$,$U_1 = U - U_1$. According to (23), the Luca-Termini fuzzy entropy of U_1 is computed as follow,

$$\begin{aligned} H_{LucaT}^{U_1}(C) &= -\frac{1}{|U_1|}\sum_{x\in U_1}\left[\mu_{C_1}(x)\, log_2\mu_{C_1}(x) + (1-\mu_{C_1}(x))\, log_2(1-\mu_{C_1}(x))\right] \\ &= -\left[\mu_{C_1}(x_1)\, log_2\mu_{C_1}(x_1) + (1-\mu_{C_1}(x_1))\, log_2(1-\mu_{C_1}(x_1))\right] \\ &= -(0.14\times log_2 0.14 + 0.86\times log_2 0.86) = 0.58. \end{aligned}$$

The Luca-Termini fuzzy entropy of U_2 is computed as follow,

$$\begin{aligned} H_{LucaT}^{U_2}(C) &= -\frac{1}{|U_2'}\sum\left[\mu_{C_1}(x)\, log_2\mu_{C_1}(x) + (1-\mu_{C_1}(x))\, log_2(1-\mu_{C_1}(x))\right] \\ &= -\left[\mu_{C_1}\right. \end{aligned}$$

$$\begin{aligned} &= -\left[\mu_{C_1}(x_3)\, log_2\mu_{C_1}(x_3) + (1-\mu_{C_1}(x_3))\, log_2(1-\mu_{C_1}(x_3))\right] \\ &= -\left[\mu_{C_1}(x_4)\, log_2\mu_{C_1}(x_4) + (1-\mu_{C_1}(x_4))\, log_2(1-\mu_{C_1}(x_4))\right] \\ &= -\left[\mu_{C_1}(x_5)\, log_2\mu_{C_1}(x_5) + (1-\mu_{C_1}(x_5))\, log_2(1-\mu_{C_1}(x_5))\right] \\ &= -\left[\mu_{C_1}(x_6)\, log_2\mu_{C_1}(x_6) + (1-\mu_{C_1}(x_6))\, log_2(1-\mu_{C_1}(x_6))\right] \\ &= -[(0.80\times log_2 0.80 + 0.20\times log_2 0.20) \\ &\quad + (0.19\times log_2 0.19 + 0.81\times log_2 0.81) \\ &\quad + (0.12\times log_2 0.12 + 0.88\times log_2 0.88) \\ &\quad + (0.92\times log_2 0.92 + 0.08\times log_2 0.08) \\ &\quad + (0.83\times log_2 0.83 + 0.17\times log_2 0.17)] = 0.60. \end{aligned}$$

Hence, according to (24), we can obtain the Luca-Termini fuzzy entropy of cut t_1,

$$H_{LucaT}(t_1) = \frac{H_{LucaT}^{U_1}(C) + H_{LucaT}^{U_2}(C)}{2} = 0.59.$$

Similarly, we can obtain the Luca-Termini fuzzy entropy of other cuts t_2, t_2, t_4, t_5,

$$\begin{aligned} H_{LucaT}(t_2) &= 0.61, H_{LucaT}(t_3) = 0.60, \\ H_{LucaT}(t_4) &= 0.58, H_{LucaT}(t_5) = 0.62. \end{aligned}$$

Because that the Luca-Termini fuzzy entropy of t_4 is the maximal, it is selected the optimal cut. Repeat process (1) and (2), we finally obtain the fuzzy decision tree shown in Fig. 1.

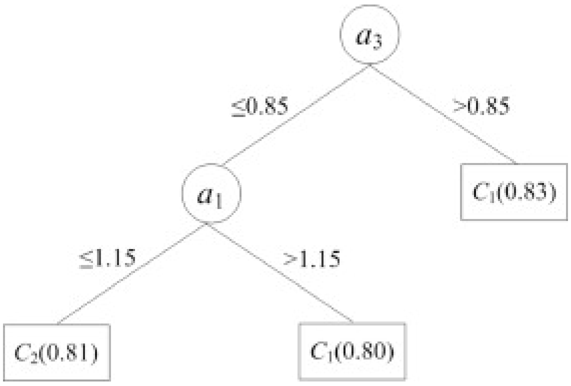


Fig. 1. The fuzzy decision tree induced from Table 1 by the proposed algorithm.

Table 1. A small fuzzy decision table with 6 instances.

Instances	Continuous-valued conditional attributes			Fuzzy-valued decision attributes	
	a_1				
x_1	0.60	0.70	0.40	0.14	0.86
x_2	0.50	1.20	0.90	0.92	0.08
x_3	0.70	0.60	0.70	0.19	0.81
x_4	1.30	0.50	1.20	0.83	0.17
x_5	1.20	0.80	0.60	0.80	0.20
x_6	1.10	1.10	0.80	0.12	0.88

5. Experimental results and statistical analysis

In order to verify the effectiveness of the proposed algorithm TRFDT, we experimentally compared TRFDT with two state-of-the-art approaches which are fuzzy ID3 [29] and FDT-YS [39] on two aspects: testing accuracy and CPU time. The experiments are conducted on 11 data sets with 10-fold cross-validation in the environment Matlab 2013b, the experimental results are the average of the 10 outputs. In Algorithm 1, the λ is a predefined parameter by user. Theoretically, the smaller value of the λ, the higher testing accuracy of the TRFDT is. But if set λ=0.0, the induced fuzzy decision trees may be very large, whereas very large decision trees are prone to overfitting. In our experiments, we determined the suitable value of λ by experimental trials, and we found that when $\lambda = 0.2$, the testing accuracy of the TRFDT does not increase on almost all data sets used in our experiments. Hence, in the subsequent experiments, we set $\lambda = 0.2$. The selected data sets include 9 UCI data sets [18] and 2 real world data sets [43]. The 2 real world data sets are RenRu data set and CT data set. The CT data set was obtained by collecting 212 medical CT images from Baoding local hospital. All

CT images are classified into 2 classes (i.e., normal class and abnormal class). The CT data set has 170 normal instances and 42 abnormal instances. Totally 35 features are initially selected. They are 10 symmetric features, 9 texture features and 16 statistical features including mean, variance, skewness, kurtosis, energy and entropy. The RenRu data set was created by the key laboratory of machine learning and computational intelligence of Hebei Province, China. The RenRu data set was obtained by collecting 148 Chinese characters REN and RU with different typeface, font and size, in which there are 92 Chinese characters REN and 56 Chinese characters RU. For each Chinese character, it is described by 26 numerical features. The basic information of the 11 data sets are summarized in Table 2.

Table 2. The basic information of the 11 UCI data sets.

Data sets	Number of instances	Number of attributes	Number of classes
CT	212	35	2
RenRu	148	26	2
Iris	50	[illegible]	[illegible]
Glass	[illegible]	[illegible]	[illegible]
Pima	768	8	2
Parkinsons	195	22	2
Wine	130	13	3
Breast-WDBC	555	30	2
Image Segmentation	194	19	7
Ionosphere	337	34	2
Blood Transfusion	748	4	2

In the experiments, for the selected data sets with continuous-valued conditional attributes and crisp decision attribute. we first fuzzify the crisp decision attribute to determine the fuzzy membership degrees of instances. Class center based fuzzification algorithm [43] (see the following Algorithm 2) is used in this paper. Next, the proposed algorithm is compared with fuzzy ID3 and FDT-YS on two aspects: the testing accuracy and CPU time. For each data set, we run 10-fold cross-validation 10 times, the experimental results are the average of the 10 outputs. The experimental results on testing accuracy and CPU time are listed in Table 2 and Table 3 respectively.

Algorithm 2. Fuzzification.

Input: $DT = (U, A \cup C)$, $U = \{x_1, x_2, \cdots, x_n\}$, $A = \{a_1, a_2, \cdots, a_d\}$, $C = \{C_1, C_2, \ldots, C_k\}$.
Output: $\mu_{C_j}(x_i)(1 \le i \le n; 1 \le j \le k)$.

for $(j = 1; j \le k; j = j + 1)$ **do**
  Calculate the center c_j of class j;
end
for $(i = 1; i \le n; i = i + 1)$ **do**
  for $(j = 1; j \le k; j = j + 1)$ **do**
    Calculate the distance d_{ij} between x_i and center c_j;
  end
end
for $(i = 1; i \le n; i = i + 1)$ **do**
  for $(j = 1; j \le k; j = j + 1)$ **do**
    Calculate fuzzy membership degree of x_i with respect to class j with the following formula;
    $\mu_{C_j}(x_i) = \frac{(d_{ij}^2)^{-1}}{\sum_{j=1}^{k}(d_{ij}^2)^{-1}}$.
    // $d_{ij} = \| x_i - c_j \|_2$, c_j is the center of class C_j.
  end
end
Output $\mu_{C_j}(x_i)$.

Table 3. The comparison between TRFDT, fuzzy ID3 and FDT-YS on testing accuracy.

Data sets			
CT	0.9157	0.8818	**0.9238**
RenRu	0.8393	0.8727	**0.8806**
Iris	0.9531	**0.9533**	0.9405
Glass	0.5778	**0.5838**	0.5698
Pima	**0.6460**	0.5917	0.6320
Parkinsons	0.8595	0.8555	**0.8753**
Wine	0.6017	0.5433	**0.6410**
Breast-WDBC	0.9784	0.9530	**0.9819**
Image segmentation	**0.7757**	0.5524	0.7643
Ionosphere	0.6506	0.7261	**0.8107**
Blood transfusion	0.8421	0.8522	**0.8911**

From Tables 2 and 3, we can find that (a) regarding the testing accuracy, the proposed algorithm TRFDT outperforms fuzzy ID3 on 9 data sets, and outperforms FDT-YS also on 9 data sets, (b) regarding the CPU time, the proposed algorithm TRFDT outperforms fuzzy ID3 and FDT-YS on all data sets. In order to further confirm the observation (b), we have analyzed statistically the experimental results by paired *t*-test with confidence level 0.05 [3], [11], [44]. Firstly, for each data set and for each algorithm, we run 10-fold cross-validation 10 times and then obtain 3 statistics denoted by X_1, X_2 and X_3 corresponding to fuzzy ID3, FDT-YS and our proposed algorithm TRFDT respectively. Next we apply paired *t*-test to the experimental results by computing the values of MATLAB functions t−test2(X_1, X_3) and t−test2(X_2, X_3). The p-values of paired t-test are listed in Table 4. The p-values listed in Table 4 confirm statistically that the observation is correct with probability 0.95.

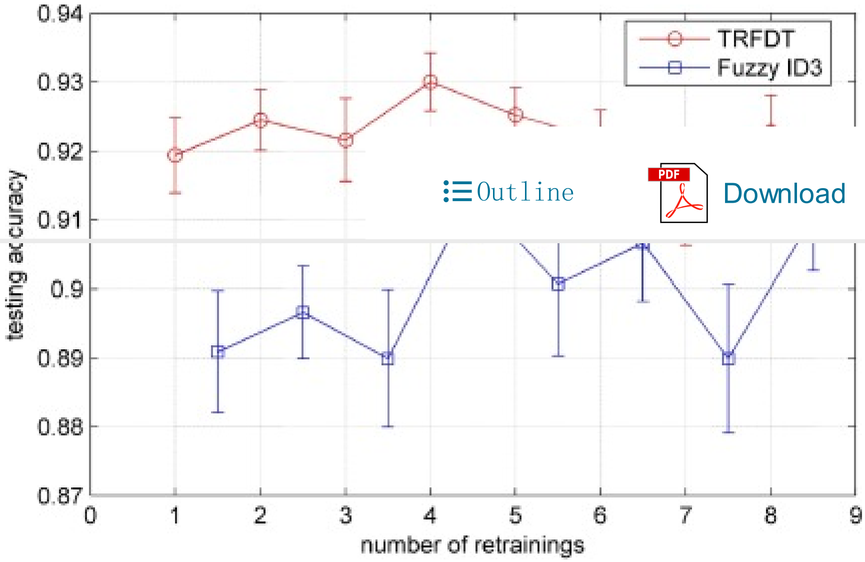
Table 4. The comparison between TRFDT, fuzzy ID3 and FDT-YS on CPU time(s).

Data sets	**Fuzzy ID3**	**FDT-YS**	**TRFDT**
CT	1.28	1.52	**1.13**
RenRu	1.19	1.21	**1.08**
Iris	0.18	0.28	**0.17**
Glass	0.28	0.25	**0.23**
Pima	1.04	1.22	**0.97**
Parkinsons	0.82	0.92	**0.72**
Wine	0.76	0.84	**0.68**
Breast-WDBC	[illegible]	[illegible]	[illegible]
Image Segmentation	0.72	0.71	**0.69**
Ionosphere	1.51	1.68	**1.47**
Blood Transfusion	0.78	0.92	**0.76**

Table 5. The p-values of statistical analysis on CPU times.

Data sets	***p*-value1**	***p*-value2**
CT	1.738e−04	3.184e−04
RenRu	3.401e−05	1.652e−03
Iris	6.000e−04	3.124e−05
Glass	1.760e−06	2.617e−05
Pima	4.121e−04	4.126e−04
Parkinsons	3.019e−05	1.738e−06
Wine	1.344e−07	2.886e−05
Breast-WDBC	2.618e−03	3.009e−04
Image Segmentation	5.162e−04	5.005e−05
Ionosphere	3.333e−05	2.965e−06
Blood Transfusion	2.735e−04	4.410e−04

In addition, we also investigated the sensitivity of the proposed algorithm to random perturbation. For each data set, we run the proposed algorithm 8 times with 10-fold cross-validation, we added random perturbation to the input at every time, and record every testing accuracy, the interval between the minimum and the maximum of the testing accuracy is the corresponding fluctuation amplitude. The experimental results are presented in Figs. 2 to5. Due to the limitation of pages here we only present the experimental results on 2 data sets which are CT and Parkinsons, the experimental results on other data sets are similar. The experimental results compared TRFDT with Fuzzy ID3 and FDT-YS on data set CT are given in Figs. 2 and3, while the experimental results compared TRFDT with Fuzzy ID3 and FDT-YS on data set Parkinsons are given in Fig. 4 and5 respectively. From the experimental results given in Figs. 2 to5, we can find that the proposed algorithm TRFDT outperforms the two other algorithms on sensitivity.



Download high-res image (233KB) Download full-size image

Fig. 2. The experimental results of sensitivity compared TRFDT with fuzzy ID3 on data set CT.

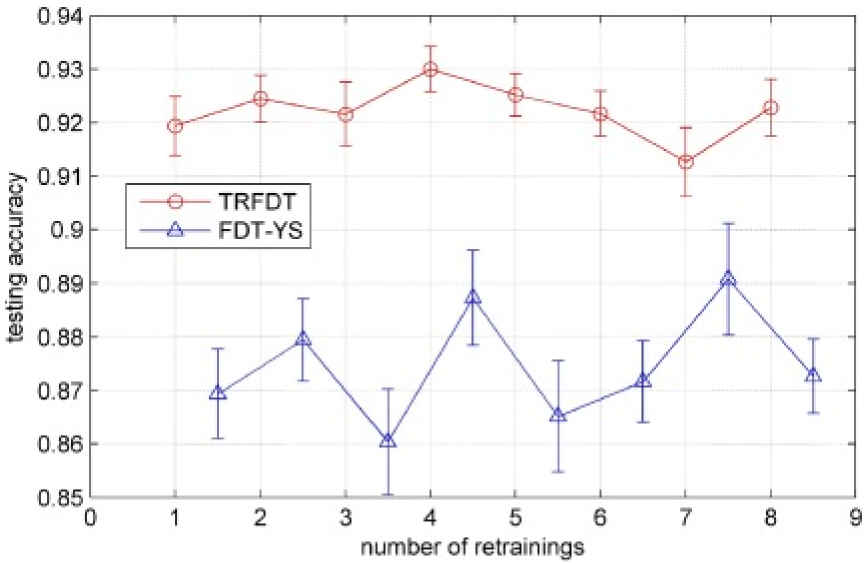


Fig. 3. The experimental results of sensitivity compared TRFDT with FDT-YS on data set CT.

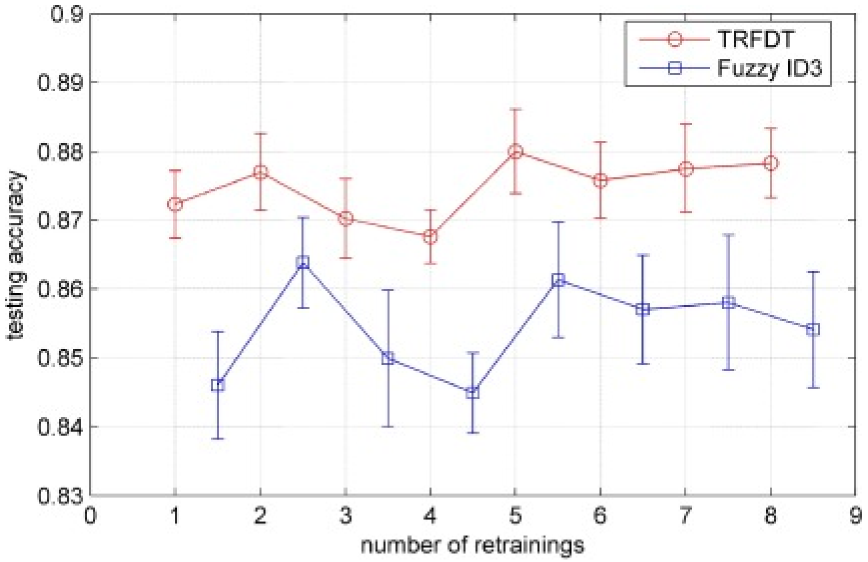


Fig. 4. The experimental results of sensitivity compared TRFDT with fuzzy ID3 on data set Parkinsons.

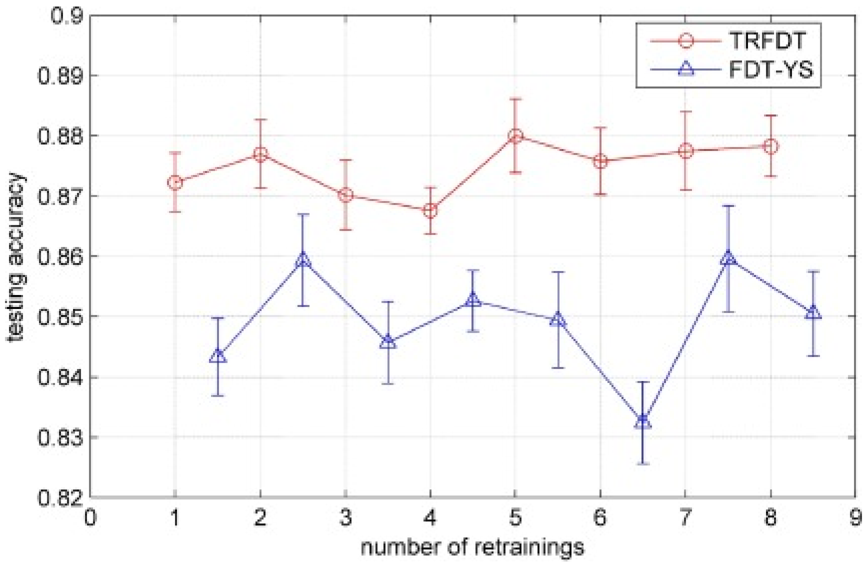


Fig. 5. The experimental results of sensitivity compared TRFDT with FDT-YS on data set Parkinsons.

6. Conclusion

Based on tolerance rough fuzzy set, an algorithm of induction of fuzzy decision tree is proposed in this paper, and we theoretically proved that the proposed algorithm is convergent with a very large probability. The advantage of the proposed algorithm is that it can directly induce fuzzy decision tree from fuzzy decision table with continuous-valued conditional attributes and fuzzy-valued decision attributes, the fuzzification of the continuous-valued conditional attributes can be avoid. Thereby, the proposed algorithm TRFDT has fast learning speed and good generalization ability, which have been verified by experimental comparison with two state-of-the-art approaches fuzzy ID3 and FDT-YS. Furthermore, the experimental results also demonstrated that the proposed algorithm TRFDT outperforms the Fuzzy ID3 and FDT-YS on sensitivity. The promising future works of this study include (1) the investigation of applications of the proposed algorithm TRFDT in practice, (2) the study of scalability of the proposed algorithm TRFDT in the large scale data circumstance, and corresponding implementation by MapReduce in Hadoop environment.

Acknowledgement

This research is supported by Basic Research Project of Knowledge Innovation Program in Shenzhen (JCYJ20150324140036825), by National Natural Science Foundations of China (71371063), and by the Natural Science Foundation of Hebei Province (F2017201026).

References

[1] R.A.R. Ashfaq, Wang X.Z., Huang J.Z.X., *et al.*
Fuzziness based semi-supervised learning approach for intrusion detection system
Inf. Sci. (Ny), 378 (2017), pp. 484-497
Article Download PDF View Record in Scopus

[2] Chang P., Fan C., Dzan W.
A CBR-based fuzzy decision tree approach for database classification
Expert Syst. Appl., 37 (1) (2010), pp. 214-225
Article Download PDF View Record in Scopus

[3] T.G. Dietterich
Approximate statistical test for comparing supervised classification learning algorithms
Neural Comput., 7 (10) (1998), pp. 1895-1923
CrossRef View Record in Scopus

[4] D. Dubois, H. Prade
Rough fuzzy sets and fuzzy rough sets
Int. J. Gen. Syst., 17 (2) (1990), pp. 191-208
CrossRef View Record in Scopus

[5] M.E.A. El-Monsef, M.A. El-Gayar, R.M. Aqeel
A comparison of three types of rough fuzzy sets based on two universal sets
Int. J. Mach. Learn. Cybern., 8 (1) (2017), pp. 343-353

[6] Fan B.J., Tsang E.C.C., Xu W.H., *et al.*
Double-quantitative rough fuzzy set based decisions: a logical operations method

Inf. Sci. (Ny), 378 (2017), pp. 264-281
Article Download PDF View Record in Scopus

[7] Gong Z.T., Zhang X.X.
The further investigation of variable precision intuitionistic fuzzy rough set model
Int. J. Mach. Learn. Cybern., 8 (5) (2017), pp. 1565-1584
CrossRef View Record in Scopus

[8] He Y.L., Liu J.N.K., Hu Y.X., *et al.*
OWA operator based link prediction ensemble for social network
Expert Syst. Appl., 42 (1) (2015), pp. 21-50
Article Download PDF View Record in Scopus

[9] He Y.L., Wang X.Z., Huang J.Z.X.
Recent advances in multiple criteria decision making techniques
Int. J. Mach. Learn. Cybern. (2016), 10.1007/s13042-015-0490-y
In press

[10] He Y.L., Wang X.Z., Huang J.Z.X.
Fuzzy nonlinear regr
Inf. Sci. (Ny), 364–365

Article Download PDF View Record in Scopus

[11] Janez D.
Statistical comparisons of classifiers over multiple data sets
J. Mach. Learn. Res., 7 (1) (2006), pp. 1-30
View Record in Scopus

[12] J. Jang, Structure determination in fuzzy modeling: a fuzzy CART approach, Proceedings of the Third IEEE Conference on Fuzzy Systems 1 (1994) 480–485.

[13] C.Z. Janikow
Fuzzy decision tree: issues and methods
IEEE Trans. Syst. Man Cybern. Part B, 28 (1) (1998), pp. 1-14
CrossRef View Record in Scopus

[14] B. Kosko
Fuzzy entropy and conditioning
Inf. Sci. (Ny), 40 (2) (1986), pp. 165-174
Article Download PDF View Record in Scopus

[15] A. Kumar, M. Hanmandlu, H.M. Gupta
Ant colony optimization based fuzzy binary decision tree for bimodal hand knuckle verification system
Expert Syst. Appl., 40 (2) (2013), pp. 439-449
Article Download PDF View Record in Scopus

[16] A. Kumar, M. Hanmandlu, H.M. Gupta
Fuzzy binary decision tree for biometric based personal authentication

Neurocomputing, 99 (1) (2013), pp. 87-97
Article Download PDF CrossRef View Record in Scopus

[17] Y. Lertworaprachaya, Yang Y., John R.
Interval-valued fuzzy decision trees with optimal neighbourhood perimeter
Appl. Soft Comput., 24 (2014), pp. 851-866
Article Download PDF View Record in Scopus

[18] M. Lichman
UCI Machine Learning Repository, University of California, School of Information and Computer Science, Irvine, CA (2013)

http://archive.ics.uci.edu/ml

[19] Liu X.D., X. Feng X., Pedrycz W.
Extraction of fuzzy rules from fuzzy decision trees: an axiomatic fuzzy sets (AFS) approach
Data Knowl. Eng., 84 (3) (2013), pp. 1-25
Article Download PDF CrossRef View Record in Scopus

[20] A.D. Luca, S. Termini
A definition of a non-

Inf. Control, 20 (4) (1972), pp. 301-312

[21] C. Olaru, L. Wehenkel
A complete fuzzy decision tree technique
Fuzzy Sets Syst., 138 (2) (2003), pp. 221-254
Article Download PDF View Record in Scopus

[22] Pawlak Z.
Rough sets
Int. J. Inf. Comput. Sci., 11 (5) (1982), pp. 341-356
CrossRef View Record in Scopus

[23] Pedrycz W., A. Sosnowski
C-Fuzzy decision trees
IEEE Trans. Syst. Man Cybern., 34 (4) (2005), pp. 498-511
CrossRef View Record in Scopus

[24] J.R. Quinlan
Induction of decision trees
Mach. Learn., 1 (1986), pp. 81-106
CrossRef View Record in Scopus

[25] L. Rokach, O. Maimon
Data Mining with Decision Trees: Theory and Applications
World Scientific ((2nd ed.)) (2015)

[26] A. Segatori, F. Marcelloni, Pedrycz W.
On distributed fuzzy decision trees for big data
IEEE Trans. Fuzzy Syst., 26 (1) (2018), pp. 174-192

View Record in Scopus

[27] A. Skowron
Tolerance approximation spaces
Fundam. Inform., 27 (2–3) (1996), pp. 245-253
View Record in Scopus

[28] Tsang E.C.C., Sun B.Z., Ma W.M.
General relation-based variable precision rough fuzzy set
Int. J. Mach. Learn. Cybern., 8 (3) (2017), pp. 891-901
CrossRef View Record in Scopus

[29] M. Umano, H. Okamolo, I. I. Hatono, et al., Fuzzy decision trees by fuzzy ID3 algorithm and its application to diagnosis system, Proceedings of the IEEE International Conference on Fuzzy Systems 3 (1994) 2113–2118.

[30] Wang X.Z., Yeung D.S., Tsang E.C.C.
A comparative study on heuristic algorithms for generating fuzzy decision trees
IEEE Trans. Syst. Man Cybern Part B, 31 (2) (2001), pp. 215-226
CrossRef View Rec

[31] Wang X.Z., Xing H.J., L
A study on relationship between generalization abilities and fuzziness of base classifiers in ensemble learning
IEEE Trans. Fuzzy Syst., 23 (5) (2015), pp. 1638-1654
CrossRef View Record in Scopus

[32] Wang X.Z., Dong L.C., Yan J.H.
Maximum ambiguity-based sample selection in fuzzy decision tree induction
IEEE Trans. Knowl. Data Eng., 24 (8) (2012), pp. 1491-1505
CrossRef View Record in Scopus

[33] Wang C.Z., Hu Q.H., Wang X.Z., *et al.*
Feature selection based on neighborhood discrimination index
IEEE Trans. Neural Netw. Learn. Syst. (2018), 10.1109/TNNLS.2017.2710422

[34] Wang X.C., Liu X.D., Pedrycz W., *et al.*
Fuzzy rule based decision trees
Pattern Recognit., 48 (1) (2015), pp. 50-59
Article Download PDF View Record in Scopus

[35] Wang X.Z.
Learning from big data with uncertainty-editorial
J. Intell. Fuzzy Syst., 28 (5) (2015), pp. 2329-2330
CrossRef View Record in Scopus

[36] Wang C.Z., Qi Y.L., Shao M.W., *et al.*
A fitting model for feature selection with fuzzy rough sets
IEEE Trans. Fuzzy Syst., 25 (4) (2017), pp. 741-753
CrossRef View Record in Scopus

[37] Yang X., Li T.R., Liu D., *et al.*
A unified framework of dynamic three-way probabilistic rough sets
Inf. Sci. (Ny), 420 (2017), pp. 126-147
Article Download PDF CrossRef View Record in Scopus

[38] Yang R., Wang Z.Y.
Cross-oriented Choquet integrals and their applications on data classification
J. Intell. Fuzzy Syst., 28 (1) (2015), pp. 205-216
View Record in Scopus

[39] Yuan Y., Shaw M.J.
Induction of fuzzy decision trees
Fuzzy Sets Syst., 69 (2) (1995), pp. 125-139
Article Download PDF View Record in Scopus

[40] Zadeh L.A.
Fuzzy sets
Inf. Control, 8 (3) (1965), pp. 338-353
Article Download PDF View Record in Scopus

[41] Zaitseva E., V. Levashe
Construction of a reliability structure function based on uncertain data
IEEE Trans. Reliab., 65 (4) (2016), pp. 1710-1723
CrossRef View Record in Scopus

[42] M. Zeinalkhani, M. Eftekhari
Fuzzy partitioning of continuous attributes through discretization methods to construct fuzzy decision tree classifiers
Inf. Sci. (Ny), 278 (26) (2014), pp. 715-735
Article Download PDF View Record in Scopus

[43] Zhai J.H.
Fuzzy decision tree based on fuzzy-rough technique
Soft Comput., 15 (6) (2011), pp. 1087-1096
CrossRef View Record in Scopus

[44] Zhai J.H., Xu H.Y., Wang X.Z.
Dynamic ensemble extreme learning machine based on sample entropy
Soft Comput., 16 (9) (2012), pp. 1493-1502
CrossRef View Record in Scopus

[45] Zhai J.H., Zhang S.F., Wang C.X.
The classification of imbalanced large data sets based on MapReduce and ensemble of ELM classifiers
Int. J. Mach. Learn. Cybern., 8 (3) (2017), pp. 1009-1017
CrossRef View Record in Scopus

[46] Zhai J.H., Hou S.X., Zhang S.F.
Induction of tolerance rough fuzzy decision tree

Proceeding of the International Conference on Machine Learning and Cybernetics, Guangzhou, China (2015), pp. 843-848
CrossRef View Record in Scopus

[47] Zhai J.H., Zhang Y., Zhu H.Y.
Three-way decisions model based on tolerance rough fuzzy set
Int. J. Mach. Learn. Cybern., 8 (1) (2017), pp. 35-43
CrossRef View Record in Scopus

[48] Zhang X., Mei C.M., Chen D.G., *et al.*
Feature selection in mixed data: a method using a novel fuzzy rough set-based information entropy
Pattern Recognit., 56 (2016), pp. 1-15
Article Download PDF CrossRef View Record in Scopus

[49] Zhao S.Y., Chen H., Li C.P., *et al.*
A novel approach to building a robust fuzzy rough classifier
IEEE Trans. Fuzzy Syst., 23 (4) (2015), pp. 769-786
CrossRef View Record in Scopus

[50] Zhao H.Y., Ma W.M., S
A novel decision making approach based on intuitionistic fuzzy soft sets
Int. J. Mach. Learn. Cybern., 8 (3) (2017), pp. 1107-1117
CrossRef View Record in Scopus

[51] Zhou W., Zhang S.L.
The decision delay in finite-length MMSE-DFE systems
Wirel. Personal Commun., 83 (1) (2015), pp. 175-189
CrossRef View Record in Scopus