Fuzzy decision tree based on fuzzy-rough technique

Jun-hai Zhai

Abstract Using an efficient criterion in selection of fuzzy conditional attributes (i.e. expanded attributes) is important for generation of fuzzy decision trees. Given a fuzzy information system (FIS), fuzzy conditional attributes play a crucial role in fuzzy decision making. Besides, different fuzzy conditional attributes have different influences on decision making, and some of them may be more important than the others. Two well-known criteria employed to select expanded attributes are fuzzy classification entropy and classification ambiguity, both of which essentially use the ratio of uncertainty to measure the significance of fuzzy conditional attributes. Based on fuzzy-rough technique, this paper proposes a new criterion, in which expanded attributes are selected by using significance of fuzzy conditional attributes with respect to fuzzy decision attributes. An illustrative example as well as the experimental results demonstrates the effectiveness of our proposed method.

Keywords Fuzzy decision trees · Fuzzy rough sets · Fuzzy classification entropy · Fuzzy conditional attributes · Fuzzy decision attributes

1 Introduction

As a generalization of ID3 (Quinlan 1986), Fuzzy ID3 (Umano et al. 1994) uses the fuzzy classification entropy as the criterion to measure the uncertainty of the nodes. A similar version of fuzzy ID3 was proposed by Yuan and

J. Zhai (✉)
Key Laboratory of Machine Learning and Computational
Intelligence, College of Mathematics and Computer Science,
Hebei University, Baoding 071002, Hebei, China
e-mail: mczjh@hbu.edu.cn

Shaw (1995), in which the classification ambiguity is used as the criterion to measure the uncertainty of the nodes. Entropy and ambiguity are two types of uncertainties. Entropy describes the classification impurity of a partition with respect to the classes, while ambiguity describes the specificity of objects in selection of expanded attributes. One can find many types of uncertainties in the literature.

Rough sets developed by Pawlak (1982) are useful tools to deal with uncertainty (i.e. vagueness and ambiguity) of knowledge. Fuzzy rough sets (Dubois and Prade 1990; Kuncheva 1992; Nanda 1992) are generalizations of rough sets to deal with both fuzziness and vagueness in data, which integrate fuzzy sets and rough sets together. It is reasonable to believe that fuzzy rough sets are better than fuzzy sets or rough sets as tools to deal with uncertainty in complicated environments. In fuzzy rough sets, fuzzy similarity relations are employed to characterize the degree of similarity between two objects, which takes values on the unit interval. Most existing research on fuzzy rough sets focus on approximations of fuzzy sets (Chen et al. 2006; Hu et al. 2006a), fuzzy rough attribute reductions (Tsang et al. 2008; Jensen and Shen 2005, 2004; Hu et al. 2006b), and fuzzy rough attribute selections (Jensen and Shen 2007; Shen and Jensen 2004; Bhatt and Gopal 2005). Comparative studies can be found in Radzikowska and Kerre (2002), Yeung et al. (2005).

It is essential to study how to effectively make use of multiple fuzzy attribute reductions to extract fuzzy classification rules. Obviously, the quality of extracted rules heavily depends on the quality of fuzzy attribute reductions. Given a FIS, there are usually many fuzzy conditional attributes and multiple fuzzy attribute reductions [sometimes tens or hundreds of fuzzy attribute reductions can be found (Hu et al. 2007)]. Each fuzzy conditional attribute has a different significance, and therefore has

different contribution to classification. More specifically, some of these attributes may have great contribution to classification, and the others may have little or even no contribution. It implies that the important fuzzy conditional attributes will play a crucial role in a FIS. In this paper, based on fuzzy-rough technique, we propose a method with new criterion which uses significance instead of fuzzy entropy to select the most important fuzzy conditional attribute recursively as expanded node in generating a fuzzy decision tree.

The paper is organized as follows. Some related notions and theoretical background are given in Sect. 2. In Sect. 3, the method based on fuzzy-rough technique for generating fuzzy decision tree is presented. An example is provided to illustrate our proposed method in Sect. 4. The experimental comparison of our proposed method, fuzzy ID3, and the method by Yuan and Shaw (1995) are presented in Sect. 5. Section 6 concludes the paper.

2 Fuzzy information system and fuzzy rough sets

2.1 Fuzzy information system

An information system (IS) refers to a 4-tuple $\text{IS} = (U, A \cup C, V, f)$ where $U = \{x_1, x_2, \ldots, x_N\}$ is a non-empty finite set of objects and each x_i is usually represented as $x_i = (a_{i1}, a_{i2}, \ldots, a_{in})$; $A = \{a_1, a_2, \ldots, a_n\}$ is a set of conditional attributes; $C = \{c\}$ is the decision attribute (without loss of generality we suppose that $c \in \{1, 2, \ldots, p\}$); $V = \bigcup_{j=1}^{n} V_{a_j}$ where $V_{a_j}(1 \leq j \leq n)$ is the domain of value of attribute a_j; $f : U \times A \rightarrow V$ is called an information function, i.e. $f(x_i, a_j) = a_{ij}(1 \leq i \leq N; 1 \leq j \leq n)$.

A fuzzy information system (FIS) (Wang et al. 2001) is a fuzzy version of an IS, which is also a 4-tuple $\text{FIS} = (U, A \cup C, V, f)$ where U has the same meaning in IS and A is a finite set of fuzzy conditional attributes. In the FIS, we denote by $A = \{A_1, A_2, \ldots, A_n\}$ where $A_i(1 \leq i \leq n)$ represents a fuzzy conditional attribute which consists of a set of fuzzy linguistic terms $\text{FLT}_i = \{A_{i1}, A_{i2}, \ldots, A_{ik_i}\}(1 \leq i \leq n)$. C denotes a fuzzy decision attribute with a set of fuzzy linguistic terms $\text{FLT}_C = \{C_1, C_2, \ldots, C_m\}$. Each fuzzy linguistic term $A_{ij}(1 \leq i \leq n; 1 \leq j \leq k_i)$ or $C_l(1 \leq l \leq m)$ is considered as a fuzzy set on U, represented as

$$A_{ij} = \frac{a_{ij}^{(1)}}{x_1} + \frac{a_{ij}^{(2)}}{x_2} + \cdots + \frac{a_{ij}^{(N)}}{x_N}$$

and

$$C_l = \frac{c_l^{(1)}}{x_1} + \frac{c_l^{(2)}}{x_2} + \cdots + \frac{c_l^{(N)}}{x_N}$$

where $a_{ij}^{(p)}(1 \leq i \leq n; 1 \leq j \leq k_i; 1 \leq p \leq N)$ is the conditional membership degree, while $c_l^{(p)}(1 \leq l \leq m; 1 \leq p \leq N)$ is the class membership degree. V and f are defined same as that in the IS, respectively.

Generally a FIS can be described in Table 1. As an illustration, Table 2 gives a small FIS with four fuzzy conditional attributes and one fuzzy decision attribute. The four fuzzy conditional attributes and their linguistic terms are

$A_1 = \text{Outlook}, \text{FLT}_1 = \{A_{11}, A_{12}, A_{13}\}$
  $= \{Sunny, Cloudy, Rain\}$

$A_2 = \text{Temperature}, \text{FLT}_2 = \{A_{21}, A_{22}, A_{23}\}$
  $= \{Hot, Mild, Cool\}$

$A_3 = \text{Humidity}, \text{FLT}_3 = \{A_{31}, A_{32}\} = \{Humid, Normal\}$

$A_4 = \text{Wind}, \text{FLT}_4 = \{A_{41}, A_{42}\} = \{Windy, Not - Windy\}$

The fuzzy decision attribute and its linguistic terms are

$C = \{Plan\} \text{ and } \text{FLT}_C = \{C_1, C_2, C_3\} = \{V, S, W\}.$

Fuzzy classification rules can be extracted from a FIS by using fuzzy inductive learning methods, such as fuzzy decision trees (Yuan and Shaw 1995; Umano et al. 1994; Wang et al. 2001, 2000, 2008; Janikow 1998; Chiang and Hsu 2002; Pedrycz and Sosnowski 2005a, b, Wang et al. 2007), fuzzy rough sets (Shen and Chouchoulas 2002; Wang et al. 2007; Cao et al. 2003), fuzzy neural networks (Buckley and Hayashi 1994; Simpson 1992; Kasabov 1996), fuzzy support vector machines (Lin and Wang 2002; Chen and Wang 2003), and other techniques (Hung et al. 2008; Chen and Tsai 2008; Chen and Shie 2009; Tsai et al. 2006). This paper proposes a new methodology for rule extraction based on fuzzy-rough technique. To clearly describe our proposed methodology, we first give some definitions and notations.

Definition 2.1 Given a $\text{FIS} = (U, A \cup C, V, f)$ and a fuzzy set $F, \forall x \in U$, the significant level α of the membership degree is defined as follows (Wang et al. 2001).

$$\mu_{F_\alpha} = \begin{cases} \mu_F(x) & \text{if } \mu_F(x) \geq \alpha \\ 0 & \text{if } \mu_F(x) < \alpha \end{cases} \quad (1)$$

Definition 2.2 Let $A, B_i(1 \leq i \leq m)$ be fuzzy sets on U. A partition of A (Wang et al. 2001) is defined as $\{A \cap B_i | 1 \leq i \leq m\}$.

Definition 2.3 Let $F(U)$ be fuzzy power set on U, $F_i \subseteq F(U)(1 \leq i \leq n)$. T is called a fuzzy decision tree, if the following conditions are satisfied (Wang et al. 2000).

(1) Each node of the fuzzy decision tree belongs to $F(U)$;

Table 1 A fuzzy information system

No.	A_1				A_2				...	A_n				C			
	A_{11}	A_{12}	...	A_{1k_1}	A_{21}	A_{22}	...	A_{2k_2}	...	A_{n1}	A_{n2}	...	A_{nk_n}	C_1	C_2	...	C_m
X_1	$a_{11}^{(1)}$	$a_{12}^{(1)}$	...	$a_{1k_1}^{(1)}$	$a_{21}^{(1)}$	$a_{22}^{(1)}$	...	$a_{2k_2}^{(1)}$	...	$a_{n1}^{(1)}$	$a_{n2}^{(1)}$	...	$a_{nk_n}^{(1)}$	$c_1^{(1)}$	$c_2^{(1)}$	...	$c_m^{(1)}$
x_2	$a_{11}^{(2)}$	$a_{12}^{(2)}$	...	$a_{1k_1}^{(2)}$	$a_{21}^{(2)}$	$a_{22}^{(2)}$	...	$a_{2k_2}^{(2)}$	...	$a_{n1}^{(2)}$	$a_{n2}^{(2)}$	...	$a_{nk_n}^{(2)}$	$c_1^{(2)}$	$c_2^{(2)}$	...	$c_m^{(2)}$
$\vdots$			......				......		...			......				......	
x_N	$a_{11}^{(N)}$	$a_{12}^{(N)}$	...	$a_{1k_1}^{(N)}$	$a_{21}^{(N)}$	$a_{22}^{(N)}$	...	$a_{2k_2}^{(N)}$	...	$a_{n1}^{(N)}$	$a_{n2}^{(N)}$	...	$a_{nk_n}^{(N)}$	$c_1^{(N)}$	$c_2^{(N)}$	...	$c_m^{(N)}$

Table 2 A small fuzzy information system with 16 instances

No..	Outlook			Temperature			Humidity		Wind		Plan		
	Sunny	Cloudy	Rain	Hot	Mild	Cool	Humid	Normal	Windy	Not-windy	V	S	W
x_1	1.0	0.0	0.0	0.7	0.2	0.1	0.7	0.3	0.4	0.6	0.0	0.6	0.4
x_2	0.6	0.4	0.0	0.6	0.2	0.2	0.6	0.4	0.9	0.1	0.7	0.6	0.0
x_3	0.8	0.2	0.0	0.0	0.7	0.3	0.2	0.8	0.2	0.8	0.3	0.6	0.1
x_4	0.3	0.7	0.0	0.2	0.7	0.1	0.8	0.2	0.3	0.7	0.9	0.1	0.0
x_5	0.7	0.3	0.0	0.0	0.1	0.9	0.5	0.5	0.5	0.5	1.0	0.0	0.0
x_6	0.0	0.3	0.7	0.0	0.7	0.3	0.3	0.7	0.4	0.6	0.2	0.2	0.6
x_7	0.0	0.0	1.0	0.0	0.3	0.7	0.8	0.2	0.1	0.9	0.0	0.0	1.0
x_8	0.0	0.9	0.1	0.0	1.0	0.0	0.1	0.9	0.0	1.0	0.3	0.0	0.7
x_9	1.0	0.0	0.0	1.0	0.0	0.0	0.4	0.6	0.4	0.6	0.4	0.7	0.0
x_{10}	0.0	0.3	0.7	0.7	0.2	0.1	0.8	0.2	0.9	0.1	0.0	0.3	0.7
x_{11}	1.0	0.0	0.0	0.6	0.3	0.1	0.7	0.3	0.2	0.8	0.4	0.7	0.0
x_{12}	0.0	1.0	0.0	0.2	0.6	0.2	0.7	0.3	0.7	0.3	0.7	0.2	0.1
x_{13}	0.0	0.9	0.1	0.7	0.3	0.0	0.1	0.9	0.0	1.0	0.0	0.4	0.6
x_{14}	0.0	0.9	0.1	0.1	0.6	0.3	0.7	0.3	0.7	0.3	1.0	0.0	0.0
x_{15}	0.0	0.3	0.7	0.0	0.0	1.0	0.2	0.8	0.8	0.2	0.4	0.0	0.6
x_{16}	0.5	0.5	0.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	0.7	0.6	0.0

(2) For each non leaf node N, whose sub-nodes constitute a subset of $F(U)$ denoted by Γ (i.e. $\exists i (1 \leq i \leq n)$, $\Gamma = F_i \cap N$);

(3) Each leaf node corresponds to one or several values of classification decision.

2.2 Fuzzy rough sets

Fuzzy rough sets developed originally by D. Dubois integrate together the concepts of vagueness and indiscernibility. In the framework of Dubois of fuzzy rough sets, we review some preliminaries including several basic concepts and notations related to our proposed method.

Definition 2.4 Let U be a given universe. A partition $P = \{F_1, F_2, \ldots, F_n\}$ of U is a fuzzy partition if and only if the following two requirements hold,

(1) $\forall x_i \in U$, $\forall F_j \in P$, $\mu_{F_j}(x_i) \leq 1$;
(2) $\forall x_i \in U$, $\exists F_j \in P$, $\mu_{F_j}(x_i) \geq 0$.

where $\mu_{F_j}(x_i)$ denotes the membership degree to which x_i belongs F_j.

Definition 2.5 Let U be a given universe. R is a fuzzy equivalence relation over U if the following four requirements hold:

(1) R is a fuzzy relation on U;
(2) R is reflective, i.e. $R(x, x) = 1, \forall x \in U$;
(3) R is symmetric, i.e. $R(x, y) = R(y, x), \forall x, y \in U$;
(4) R is transitive, i.e. $R(x, z) \geq \min\{R(x, y), R(y, z)\}$, $\forall x, y, z \in U$.

An equivalence relation can generate a partition of universe U while a fuzzy equivalence relation can generate a fuzzy partition of universe U.

Definition 2.6 Let U be a given universe. R is a fuzzy equivalence relation over U. The fuzzy equivalence class $[x]_R$ is defined by

$$\mu_{[x]_R}(y) = \mu_R(x, y) \tag{2}$$

Definition 2.7 Let U be a given universe, X and P be two fuzzy sets on U. The fuzzy P-lower and P-upper approximations are defined as follows (Dubois and Prade 1990).

$$\mu_{\underline{P}X}(F_i) = \inf_{x \in U} \max\{1 - \mu_{F_i}(x), \mu_X(x)\}(i = 1, 2, \ldots, n) \tag{3}$$

$$\mu_{\overline{P}X}(F_i) = \sup_{x \in U} \min\{\mu_{F_i}(x), \mu_X(x)\}(i = 1, 2, \ldots, n) \tag{4}$$

where $F_i \in U/P (1 \leq i \leq n)$ denotes a fuzzy equivalence class, and U/P denotes the fuzzy partition of U with respect to P.

Definition 2.8 Let U be a given universe, X and P be two fuzzy sets on U, U/P be a fuzzy partition of U. For a given $x \in U$, the fuzzy P-lower approximation and the fuzzy P-upper approximation of X are defined as follows (Jensen and Shen 2005).

$$\mu_{\underline{P}X}(x) = \sup_{F \in U/P} \min\left(\mu_F(x), \inf_{y \in U} \max\{1 - \mu_F(y), \mu_X(y)\}\right) \tag{5}$$

$$\mu_{\overline{P}X}(x) = \sup_{F \in U/P} \min\left(\mu_F(x), \sup_{y \in U} \min\{\mu_F(y), \mu_X(y)\}\right) \tag{6}$$

The tuple $\left(\underline{P}X, \overline{P}X\right)$ is called a fuzzy rough set.

3 The fuzzy decision tree based on fuzzy-rough technique

3.1 New criterion for selection of expanded attribute

In this section, we will introduce a new criterion based on significance of fuzzy conditional attributes with respect to fuzzy decision attribute to select expanded attribute. Given a FIS, every fuzzy conditional attribute has different contribution to the decision attribute. In our proposed method, we will select the most important fuzzy conditional attribute as expanded attribute to generate a fuzzy decision tree. The significance of a fuzzy attribute with respect to another fuzzy attribute is introduced below.

Definition 3.1 Suppose P and Q are two fuzzy attributes in a given FIS. For a given $x \in U$, the significance of the fuzzy attribute P with respect to the fuzzy attribute Q is defined as follows,

$$\mu_{\text{POS}_P(Q)}(x) = \sup_{X \in U/Q} \mu_{\underline{P}X}(x) \tag{7}$$

In literature (Jensen and Shen 2005), equality (7) represents the membership degree of an object x belonging to the fuzzy positive region.

Definition 3.2 Suppose P and Q are two fuzzy attributes in a given FIS. The significance of P with respect to Q is defined by:

$$\tau_P(Q) = \frac{\sum_{x \in U} \mu_{\text{POS}_P(Q)}(x)}{|U|} \tag{8}$$

where $|U|$ is the cardinality of the universe U.

In literature (Jensen and Shen 2005), equality (8) represents the fuzzy dependency degree of Q on P. Obviously, $0 \leq \tau_P(Q) \leq 1$.

In order to illustrate the two concepts in Definitions 3.1 and 3.2, an example is provided as follows:

Example 1 Table 2 is a small FIS. The universe is $U = \{x_1, x_2, \ldots, x_{16}\}$. *Outlook*, *Temperature*, *Humidity*, and *Wind* are four fuzzy conditional attributes, while *Plan* is a fuzzy decision attribute. As an illustration, for $x_2 \in U$, we calculate the significance of fuzzy conditional attribute P with respect to fuzzy decision attribute Q.

Let $P = \{Outlook\}$ and $Q = \{Plan\}$. The fuzzy partitions induced by P and Q are $U/P = \{F_1, F_2, F_3\} = \{Sunny, Cloudy, Rain\}$ and $U/Q = \{V, S, W\}$ respectively.

Noting that $F_1 = Sunny$, $F_2 = Cloudy$ and $F_3 = Rain$, we have

$$\min\left(\mu_{Sunny}(x_2), \inf_{y \in U} \max\{1 - \mu_{Sunny}(y), \mu_V(y)\}\right)$$
$$= \min(\mu_{Sunny}(x_2), 0.0) = \min(0.6, 0.0) = 0.0,$$

$$\min\left(\mu_{Cloudy}(x_2), \inf_{y \in U} \max\{1 - \mu_{Cloudy}(y), \mu_V(y)\}\right)$$
$$= \min(\mu_{Cloudy}(x_2), 0.1) = \min(0.4, 0.1) = 0.1$$

$$\min\left(\mu_{Rain}(x_2), \inf_{y \in U} \max\{1 - \mu_{Rain}(y), \mu_V(y)\}\right)$$
$$= \min(\mu_{Rain}(x_2), 0.0) = \min(0.0, 0.0) = 0.0$$

which imply that the object x_2 belongs to $\underline{P}V$ with a membership degree of $\mu_{\underline{P}V}(x_2)$ where

$$\mu_{\underline{P}V}(x_2) = \sup\{\min(\mu_{Sunny}(x_2), 0.0),$$
$$\min(\mu_{Cloudy}(x_2), 0.1), \min(\mu_{Rain}(x_2), 0.0)\} = 0.1$$

Similarly, for S and W, we have

$$\mu_{\underline{P}S}(x_2) = 0.3, \quad \mu_{\underline{P}W}(x_2) = 0.1,$$

Therefore, for $x_2 \in U$, the significance of fuzzy conditional attribute P with respect to fuzzy decision attribute Q is

$$\mu_{\text{POS}_P(Q)}(x_2) = \sup_{X \in U/Q} \mu_{\underline{P}X}(x_2) = 0.3.$$

Similarly, the significance of fuzzy conditional attribute P with respect to fuzzy decision attribute Q for other objects in U can be obtained. As a result, we have that the significance of fuzzy conditional attribute P with respect to fuzzy decision attribute Q is

$$\tau_{Outlook}(Plan) = \frac{\sum_{x \in U} \mu_{\text{POS}_{Outlook}(Plan)}(x)}{|U|} = \frac{5.2}{16}$$

In this paper, we will use equality (8) to select the most important fuzzy conditional attribute as the expanded attribute to generate fuzzy decision tree.

3.2 The algorithm for generating fuzzy decision tree

This section introduces the algorithm for generating fuzzy decision trees based on the significance of fuzzy conditional attribute with respect to fuzzy decision attribute.

Input: FIS $= (U, A \cup C, V, f)$
Output: A group of fuzzy classification rules extracted from the generated fuzzy decision tree T

Step 1: Pre-preparing the FIS by using formula (1).
Step 2: Selecting the expanded attribute.

    Step 2.1: For each fuzzy conditional attribute A_i and its fuzzy linguistic terms $A_{ik_i}(1 \le i \le n)$, the significance of A_i with respect to the fuzzy decision attribute C is calculated by using formula (8).
    Step 2.2: Selecting i_0 according to $i_0 = Arg\max_{1 \le i \le n}\{\tau_{A_i}(C)\}$ (A_{i_0} is the expanded attribute).

Step 3: If the terminal conditions can not be satisfied, then partition U, and recursively select the expanded attribute until a fuzzy decision tree is generated (The terminal condition will be given in the following section).
Step 4: Extracting fuzzy classification rules from the fuzzy decision tree T.

3.3 Terminal conditions

In this section, we first give the definition of classification truth degree, and then provide terminal conditions in Sect. 3.2.

Definition 3.3 Suppose P and Q are two fuzzy sets on U. The truth degree of the fuzzy set P with respect to the fuzzy set Q is defined by:

$$\beta = T(P, Q) = \frac{\tau_P(Q)}{\sum_{x \in U} \mu_P(x)} \tag{9}$$

For example, in Table 2, the truth degree of the fuzzy set *Sunny* with respect to fuzzy set W is 0.75.

Terminal conditions used for generating a fuzzy decision tree are given below:

(1) If the classification truth degree of a branch with respect to one class exceeds a given threshold β, then the branch is terminated as a leaf node;
(2) At a branch, if no attribute can be expanded, the branch is terminated as a leaf node or as null node;
(3) If one of the conditions mentioned above is satisfied, the expanding of the branch in a fuzzy decision tree will be terminated.

During the generating of a fuzzy decision tree, if there is not a threshold to control the growth of the tree, the generated tree will be very complicated and usually the classification accuracy of the rules converted from the tree will be poor. The terminal condition (1), i.e. the threshold of classification truth degree β, plays the major role, which controls the growth of the tree. Different β will result in different fuzzy decision trees with different classification accuracies. In general, smaller β may lead to a smaller tree, but the classification accuracy will be lower. Larger β may lead to a larger tree, while the classification accuracy may be higher (on training set). The selection of β depends on the problem to be solved.

4 An illustrative example

In this section, we will demonstrate the process of generation of fuzzy decision tree by using the small FIS with 16 training instances given in Table 2. The universe is $U = \{x_1, x_2, \ldots, x_{16}\}$. The four fuzzy conditional attributes are *Outlook*, *Temperature*, *Humidity*, and *Wind*. The fuzzy decision attribute is *Plan*.

Let the significant level $\alpha = 0.0$, and the threshold of classification truth degree $\beta = 0.78$. By calculating the significance of each fuzzy conditional attribute $A_i(1 \le i \le 4)$ with respect to the fuzzy decision attribute $C = \{Plan\}$, we have

$$\tau_{A_1}(C) = \tau_{Outlook}(Plan) = \frac{\sum_{x \in U} \mu_{\text{POS}_{Outlook}(Plan)}(x)}{|U|} = \frac{5.2}{16}$$

$$\tau_{A_2}(C) = \tau_{Temperature}(Plan) = \frac{\sum_{x \in U} \mu_{\text{POS}_{Temperature}(Plan)}(x)}{|U|}$$
$$= \frac{4.8}{16}$$

$$\tau_{A_3}(C) = \tau_{Humidity}(Plan) = \frac{\sum_{x \in U} \mu_{POS_{Humidity}(Plan)}(x)}{|U|} = \frac{3.2}{16}$$

$$\tau_{A_4}(C) = \tau_{Wind}(Plan) = \frac{\sum_{x \in U} \mu_{POS_{Wind}(Plan)}(x)}{|U|} = \frac{3.2}{16}$$

Since the fuzzy conditional attribute *Outlook* has the biggest significance, it is selected as the root node. There are three branches (*Sunny*, *Cloudy* and *Rain*) from the root node *Outlook*. At the branch *Rain*, the classification truth degree for W is $0.79(0.79 > \beta)$ which meets the terminal condition (1) and becomes a leaf node with label W. At the branch *Cloudy*, $T(Cloudy, W) = 0.19 < \beta$, $T(Cloudy, S) = 0.19 < \beta$, $T(Cloudy, V) = 0.19 < \beta$. None of the classification truth degrees exceed the threshold $\beta = 0.78$, further partition based on additional attributes should be considered. We have three fuzzy partitions:

$Cloudy - Temperature = \{Cloudy \cap A_{2i}\}(i = 1, 2, 3)$

where $\{A_{21}, A_{22}, A_{23}\} = \{Hot, Mild, Cool\}$.

$Cloudy - Humidity = \{Cloudy \cap A_{3i}\}(i = 1, 2)$

where $\{A_{31}, A_{32}\} = \{Humid, Normal\}$.

$Cloudy - Wind = \{Cloudy \cap A_{4i}\}(i = 1, 2)$

where $\{A_{41}, A_{42}\} = \{Windy, Not - Windy\}$.

Similarly, we have

$$\tau_{Cloudy-Temperature}(Plan) = \frac{4.0}{16}, \tau_{Cloudy-Humidity}(Plan)$$
$$= \frac{5.6}{16}, \tau_{Cloudy-Wind}(Plan) = \frac{4.6}{16}$$

Because the value $\tau_{Cloudy-Humidity}(Plan)$ is the biggest one, the node *Humidity* is added to the branch *Cloudy*. From this node, the branch *Humid* terminates with label V (truth degree is $1 > \beta$) and branch *Normal* terminates with label W (truth degree is $0.89 > \beta$). At the branch *Sunny*, the node *Temperature* is selected as the expended attribute, and from this node, the branch *Hot* terminates with label S (truth degree is $0.83 > \beta$), the branch *Mild* terminates with label S (truth degree is $0.94 > \beta$) and the branch *Cool* terminates with label V (truth degree is $1 > \beta$). Finally we generate a fuzzy decision tree shown in Fig. 1.

Now we convert the fuzzy decision tree to a set of fuzzy classification rules. Each path of the fuzzy decision tree from root node to leaf node can be converted into a fuzzy classification rule with the highest classification truth degree. For the fuzzy decision tree shown in Fig. 1, based on the principle of maximal membership degree, six rules are obtained.

Rule 1: If *Outlook* is *Sunny* and *Temperature* is *Hot*, then *Plan* is *S* (0.83);
Rule 2: If *Outlook* is *Sunny* and *Temperature* is *Mild*, then *Plan* is *S* (0.94);
Rule 3: If *Outlook* is *Sunny* and *Temperature* is *Cool*, then *Plan* is *V* (1.00);
Rule 4: If *Outlook* is *Rain*, then *Plan* is *W* (0.79);
Rule 5: If *Outlook* is *Cloudy* and *Humidity* is *Humid*, then *Plan* is *V* (1.0);
Rule 6: If *Outlook* is *Cloudy* and *Humidity* is *Normal*, then *Plan* is *W* (0.89);

Matching to these fuzzy classification rules, each object in the FIS can acquire a result of classification. Since many rules can be applied to one object at the same time, the object can usually be classified into different classes with different degrees. We use the method in literature (Wang et al. 2001) to obtain the classification result for a given object.

On the same FIS given in Table 2, another fuzzy decision tree can be generated by using fuzzy ID3 (shown in Fig. 2). In Fig. 1, at *Sunny* branch the sub-node is *Temperature* and the *Cloudy* is *Humidity*, while in Fig. 2 at the branch *Sunny* the sub-node is *Humidity*, and the *Cloudy* is *Temperature*. The expanded attribute selected at the root node for both trees shown in Figs. 1 and 2 is the same, but at the sub-nodes, the expanded attributes selected for both trees are different. Fuzzy decision tree 1 has 6 leaf nodes and 9 inner nodes while fuzzy decision tree 2 has 10 leaf nodes and 18 inner nodes.

The classification results respectively by fuzzy ID3 and by our proposed method are listed in Table 3. Among 16 training instances, 15 instances are correctly classified by our proposed method (instance 2 is not correctly classified). The classification accuracy is 94%, which is higher than

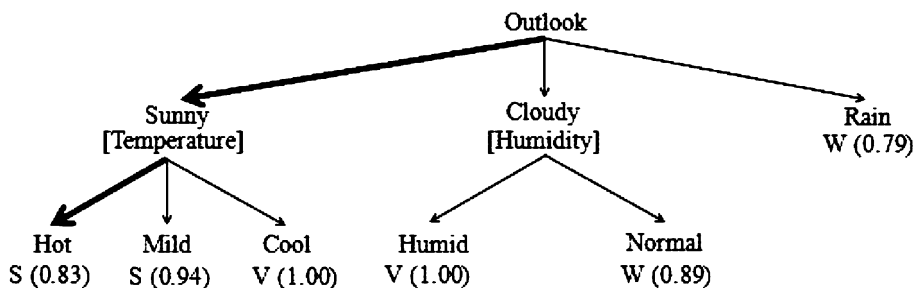**Fig. 1** Fuzzy decision tree 1 generated by our proposed method
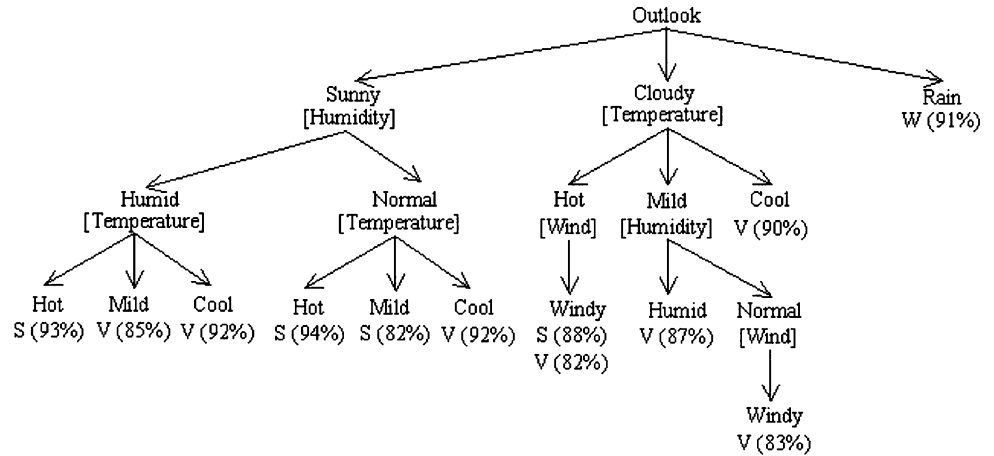
Fig. 2 Fuzzy decision tree 2 generated by fuzzy ID3

Table 3 Comparison between two methods

No..	Classification known in training data			Classification with our method			Classification with fuzzy ID3		
	V	S	W	V	S	W	V	S	W
x_1	0.0	0.6	0.4	0.1	0.7	0.0	0.2	0.7	0.0
x_2	0.7	0.6	0.0	0.4	0.6	0.4	0.2	0.6	0.0
x_3	0.3	0.6	0.1	0.3	0.7	0.2	0.2	0.7	0.0
x_4	0.9	0.1	0.0	0.7	0.3	0.2	0.7	0.2	0.0
x_5	1.0	0.0	0.0	0.7	0.1	0.3	0.5	0.1	0.0
x_6	0.2	0.2	0.6	0.3	0.0	0.7	0.3	0.0	0.7
x_7	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0
x_8	0.3	0.0	0.7	0.1	0.0	0.9	0.9	0.0	0.1
x_9	0.4	0.7	0.0	0.0	1.0	0.0	0.0	0.6	0.0
x_{10}	0.0	0.3	0.7	0.3	0.0	0.7	0.2	0.3	0.7
x_{11}	0.4	0.7	0.0	0.1	0.6	0.0	0.3	0.6	0.1
x_{12}	0.7	0.2	0.1	0.7	0.0	0.3	0.6	0.2	0.0
x_{13}	0.0	0.4	0.6	0.1	0.0	0.9	0.3	0.6	0.1
x_{14}	1.0	0.0	0.0	0.7	0.0	0.3	0.6	0.1	0.1
x_{15}	0.4	0.0	0.6	0.2	0.0	0.7	0.3	0.0	0.7
x_{16}	0.7	0.6	0.0	0.5	0.5	0.0	0.0	0.5	0.0

the classification accuracy 75% of the fuzzy decision tree 2 (instances 2, 8, 13, and 16 are not correctly classified). Obviously with respect to this example, the classification accuracy has been much improved.

5 Experiments

The effectiveness of our proposed method is demonstrated through numerical experiments in the environment of Matlab 7.0 on a Pentium 4 PC. Totally our experiments select 13 datasets among them 11 datasets are from *UCI* (Blake and Merz 1996) and 2 from real world. The 11 *UCI* datasets are Iris Dataset (DB1), Pima Dataset (DB2), Wine Dataset (DB3), Ionosphere Dataset (DB4), Breast Cancer

Dataset-WDBC (DB5), Breast Cancer Dataset-WPBC (DB6), Glass Dataset (DB7), Blood Transfusion Service Center Dataset (DB8), Image Segmentation Dataset (DB9), Parkinsons Dataset (DB10), and Wine (red) Quality Dataset (DB11). The two real world datasets are RenRu Dataset (DB12) and CT Dataset (DB13). The CT Dataset is obtained by collecting 212 medical CT images from Baoding local hospital. All CT images are classified into two classes (i.e., normal class and abnormal class). The CT Dataset 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 Dataset is created by the key laboratory of machine learning and computational

Table 4 Summary of the 13 datasets used in our experiments

DB	Number of instances	Number of attributes	Number of classes	Number of instances per class
DB1	150	4	3	50, 50, 50
DB2	768	8	2	500, 268
DB3	170	13	3	59, 71, 48
DB4	337	34	2	119, 218
DB5	555	30	2	357, 198
DB6	191	33	2	146, 45
DB7	160	9	6	16, 76, 17, 13, 9, 29
DB8	748	4	2	570, 178
DB9	194	19	7	14, 30, 30, 30, 30, 30, 30
DB10	195	22	2	48, 147
DB11	1,599	11	6	10, 53, 681, 638, 199, 18
DB12	148	26	2	92, 56
DB13	212	35	2	170, 42

intelligence of Hebei Province, China. The RenRu Dataset is 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 13 datasets are summarized in Table 4.

The conducted experiments consist of the following two steps.

5.1 Fuzzification

An *IS* with real-valued conditional attributes used in our experiments is first transformed into a FIS, which includes two steps. Firstly, we use the following algorithm to determine the fuzzy membership degree of instances.

Input: A *IS* with real-valued conditional attributes
Output: A FIS with fuzzy decision attribute (The conditional attributes remain unchanged).

Step 1: For each class $k \in \{1, 2, \ldots, p\}$, calculating the center c_k of class k;
Step 2: For each instance $x_i \in U$, calculating the distance d_{ik} between x_i and c_k;
Step 3: For each instance $x_i \in U$, calculating the fuzzy membership degree of class as follows,

$$\mu_k(x_i) = \frac{(d_{ik}^2)^{-1}}{\sum_{k=1}^{p}(d_{ik}^2)^{-1}}(1 \leq i \leq N; 1 \leq k \leq p)$$

Secondly, we use the algorithm in literature (Chen and Shie 2009) to fuzzify real-valued conditional attributes. The algorithm is listed below.

Input: A *IS* with real-valued conditional attributes.
Output: A FIS with fuzzy conditional attributes (The decision attribute remains unchanged).

Step 1: For $i = 1 \rightarrow n$;
Step 2: Let $k = 2$ and $T = T_0$;
Step 3: Constructing the membership function of real-valued conditional attributes a_i as follows:

Step 3.1: Clustering the values $\{a_{1i}, a_{2i}, \ldots, a_{Ni}\}$ of a_i into k clusters, $m_1, m_2, \ldots, m_k$ are the centers of the k clusters respectively;
Step 3.2: For $t = 1 \rightarrow k$
If $(t = 1)$ then

$$\mu_{V_{i1}}(a_i) = \begin{cases} 1 - \frac{a_i - m_1}{a_{\min} - m_1} \times 0.5 & a_{\min} \leq a_i \leq m_1 \\ 1 - \frac{a_i - m_1}{m_2 - m_1} & a_1 \leq a_i \leq m_2 \\ 0 & \text{others} \end{cases}$$

Else if $(t = k)$ then

$$\mu_{V_{ik}}(a_i) = \begin{cases} 1 - \frac{a_i - m_k}{m_{k-1} - m_k} & m_{k-1} \leq a_i \leq m_k \\ 1 - \frac{a_i - m_k}{a_{\max} - m_k} \times 0.5 & a_k \leq a_i \leq a_{\max} \\ 0 & \text{others} \end{cases}$$

Else

$$\mu_{V_{ij}}(a_i) = \begin{cases} 1 - \frac{a_i - m_j}{m_{j-1} - m_j} & m_{j-1} < a_i < m_j \\ 1 - \frac{a_i - m_j}{m_{j+1} - m_j} & m_j < a_i < m_{j+1} \\ 0 & \text{others} \end{cases}$$

Step 4: Calculating the increment $\Delta Gain(a_i)$ of information gain of attribute a_i;
Step 5: If ($k = 2$ or $\Delta Gain(a_i) > T_0$) then $k \leftarrow k + 1$, and go to Step 3;
Else exit and $k \leftarrow k - 1$;
Step 6: If $(i < n)$ then $i \leftarrow i + 1$, and go to Step 2;
Else exit.

5.2 Tree generation

We implement our algorithm on a Pentium 4 PC in the environment of Matlab 7.0. The algorithm is tested on the 13 selected datasets by tenfold cross-validation. For each dataset, we run tenfold cross-validation ten times, the experimental results are the average of the ten outputs. A comparison among our proposed method, Yuan and Shaw 1995 and fuzzy ID3 (Umano et al. 1994) is conducted. In our experiments, we select different threshold T to verify the effectiveness of our proposed approach. We find out that the threshold is sensitive to classification accuracy for a few datasets. For instance, when $T = 0.1$, the average accuracy of our proposed method on DB7 is 0.6028, and when $T = 0.07$, the average accuracy increase to 0.6918,

Table 5 Experimental results with $T_0 = 0.1$

DB	Yuan's method (Yuan and Shaw 1995) Average accuracy	Fuzzy ID3 (Umano et al. 1994) Average accuracy	Our proposed method Average accuracy
DB1	**0.9533**	0.9531	0.9333
DB2	0.7718	0.7636	**0.7839**
DB3	0.5433	0.6017	**0.6372**
DB4	0.7261	0.6506	**0.9011**
DB5	0.9530	0.9784	**0.9804**
DB6	0.9585	**0.9595**	0.8575
DB7	0.5917	**0.6460**	0.6028
DB8	0.8511	0.8421	**0.8905**
DB9	0.5524	**0.7757**	0.6033
DB10	0.8555	0.8595	**0.8645**
DB11	0.5882	0.6849	**0.7065**
DB12	**0.8727**	0.8393	0.8347
DB13	0.8818	0.9157	**0.9328**

but the latter costs much more time than the former because of much more intervals generated in the process of fuzzification. It is not really true that more intervals will lead to higher accuracy. The experimental results with $T = 0.1$ are listed in Table 5. Our proposed method outperforms fuzzy ID3 (Umano et al. 1994) and Yuan's method (Yuan and Shaw 1995) on eight datasets (see the results with boldface in Table 5). Fuzzy ID3 outperforms our proposed method and Yuan's method (Yuan and Shaw 1995) on three datasets. Yuan's method outperforms our proposed method and the fuzzy ID3 (Yuan and Shaw 1995) on two datasets. The performances of the three methods are almost same on some datasets (e.g. DB1, DB2, DB5, and DB7). In order to further verify the effectiveness of our proposed method, we have analyzed statistically the experimental results by using paired t test (Dietterich 1998). Firstly, for each dataset and for each method, we run tenfold cross-validation ten times and then obtain three statistics denoted by X_1, X_2, and X_3 (where X_i is a 100-dimensional vector, $1 \leq i \leq 3$) corresponding to Yuan's method, fuzzy ID3, and our proposed method respectively. Next we apply paired t test to the experimental results by computing the values of MATLAB function t test2(X_1, X_3) and t test2(X_2, X_3). The p values of paired t test for the 10×10-fold cross-validation are listed in Table 6. Secondly, for each dataset and for each method, we average the results of 10×10-fold cross-validation and then obtain three statistics also denoted by X_1, X_2, and X_3 (where X_i is a ten-dimensional vector, $1 \leq i \leq 3$). The p values of paired t test for the average of 10×10-fold cross-validation are

Table 6 P values of paired t test for 10×10-fold cross-validation

DB	p value 1	p value 2
DB1	1.2100e-002	1.2100e−002
DB2	6.5043e−005	3.2000e−003
DB3	1.9385e−004	1.2860e−e001
DB4	1.1906e−024	3.3042e−035
DB5	1.9507e−018	3.6970e−001
DB6	3.1932e−022	1.7467e−022
DB7	3.8100e−001	4.0853e−028
DB8	3.7137e−008	2.2946e−011
DB9	1.4526e−036	1.2637e−030
DB10	5.6400e−001	7.4610e−001
DB11	1.4615e−046	1.6947e−032
DB12	3.6929e−027	2.7404e−016
DB13	5.5199e−007	3.9000e−003

Table 7 P values of paired t test for the average of 10×10-fold cross-validation

DB	p value 1	p value 2
DB1	4.4040e−008	4.4040e−008
DB2	4.3396e−019	5.6477e−009
DB3	3.3000e−003	6.0000e−003
DB4	1.4023e−013	2.4153e−018
DB5	1.5557e−021	3.7163e−005
DB6	7.5177e−023	1.6372e−023
DB7	5.0215e−004	5.9798e−021
DB8	4.6340e−015	1.6828e−016
DB9	7.0729e−020	1.6123e−013
DB10	2.2630e−001	4.1590e−001
DB11	4.0137e−027	1.7855e−030
DB12	3.0726e−021	1.9112e−015
DB13	4.2795e−010	1.6244e−006

listed in Table 7. The small *p* values cast doubt on the validity of the null hypothesis. Statistically the *p* values of paired *t* test (shown in Tables 6, 7) verified the effectiveness of our proposed method.

6 Conclusions

In this paper, based on fuzzy-rough technique, a new method for generating fuzzy decision tree is presented, which uses significance of fuzzy conditional attributes with respect to fuzzy decision attribute as a criterion. Our proposed method takes both fuzziness and roughness existing in an information system into consideration. In comparison with other similar techniques, the fuzzy classification rules converted from fuzzy decision trees generated by using our proposed method demonstrate stronger generalization ability. In addition, numerically experimental results statistically verify the effectiveness of our proposed method.

References

Bhatt RB, Gopal M (2005) On fuzzy-rough sets approach to feature selection. Pattern Recognit Lett 26(7):965–975

Blake CL, Merz CJ (1996) UCI Repository of machine learning databases. http://www.ics.uci.edu/~mlearn/MLRepository.html

Buckley JJ, Hayashi Y (1994) Fuzzy neural networks: a survey. Fuzzy Sets Syst 66(1):1–13

Cao G, Shiu SCK, Wang X (2003) A fuzzy-rough approach for the maintenance of distributed case-based reasoning systems. Soft Comput 7(8):491–499

Chen SM, Shie JD (2009) Fuzzy classification systems based on fuzzy information gain measures. Expert Syst Appl 36(3):4517–4522

Chen SM, Tsai FM (2008) Generating fuzzy rules from training instances for fuzzy classification systems. Expert Syst Appl 35(3):611–621

Chen Y, Wang J (2003) Support vector learning for fuzzy rule-based classification systems. IEEE Trans Fuzzy Syst 11(6):716–728

Chen DG, Zhang WX, Yeung DS, Tsang ECC (2006) Rough approximations on a complete completely distributive lattice with applications to generalized rough sets. Inf Sci 176(13):1829–1848

Chiang IJ, Hsu JYJ (2002) Fuzzy classification trees for data analysis. Fuzzy Sets Syst 130(1):87–99

Dietterich TG (1998) Approximate statistical test for comparing supervised classification learning algorithms. Neural Comput 7(10):1895–1923

Dubois D, Prade H (1990) Rough fuzzy sets and fuzzy rough sets. Intern J Gen Syst 17:191–208

Hu Q, Yu D, Xie Z, Liu J (2006a) Fuzzy probabilistic approximation spaces and their information measures. IEEE Trans Fuzzy Syst 14(2):191–201

Hu Q, Yu D, Xie Z (2006b) Information-preserving hybrid data reduction based on fuzzy-rough techniques. Pattern Recognit Lett 27(5):414–423

Hu Q, Yu D, Xie Z (2007) EROS: ensemble rough subspace. Pattern Recognit 40(12):3728–3739

Hung WL, Chang YC, Chuang SC (2008) Fuzzy classification maximum likelihood algorithms for mixed-weibull distributions. Soft Comput 12(10):1013–1018

Janikow CZ (1998) Fuzzy decision tree: issues and methods. IEEE Trans Syst Man Cybern B 28(1):1–14

Jensen R, Shen Q (2004) Fuzzy-rough attributes reduction with application to web categorization. Fuzzy Sets Syst 141(3):469–485

Jensen R, Shen Q (2005) Fuzzy-rough data reduction with ant colony optimization. Fuzzy Sets Syst 149(1):5–20

Jensen R, Shen Q (2007) Fuzzy-rough sets assisted attribute selection. IEEE Trans Fuzzy Syst 15(1):73–89

Kasabov NK (1996) Learning fuzzy rules and approximate reasoning in fuzzy neural networks and hybrid systems. Fuzzy Sets Syst 82(2):135–149

Kuncheva LI (1992) Fuzzy rough sets: application to feature selection. Fuzzy Sets Syst 51:147–153

Lin CF, Wang SD (2002) Fuzzy support vector machines. IEEE Trans Neural Netw 13(2):464–471

Nanda S (1992) Fuzzy rough sets. Fuzzy Sets Syst 45:157–160

Pawlak Z (1982) Rough sets. Intern J Inform Comput Sci 11:341–356

Pedrycz W, Sosnowski ZA (2005a) Genetically optimized fuzzy decision trees. IEEE Trans Syst Man Cybern B 35(3):633–641

Pedrycz W, Sosnowski A (2005b) C-fuzzy decision trees. IEEE Trans Syst Man Cybern 34(4):498–511

Quinlan JR (1986) Induction of decision trees. Mach Learn 1:81–106

Radzikowska AM, Kerre EE (2002) A comparative study of fuzzy rough sets. Fuzzy Sets Syst 126(22):137–155

Shen Q, Chouchoulas A (2002) A rough-fuzzy approach for generating classification rules. Pattern Recognit 35(11):341–354

Shen Q, Jensen R (2004) Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring. Pattern Recognit 7(7):1351–1363

Simpson PK (1992) Fuzzy min–max neural networks-part 1: classification. IEEE Trans Neural Netw 3(5):776–786

Tsai YC, Cheng CH, Chang JR (2006) Entropy-based fuzzy rough classification approach for extracting classification rules. Expert Syst Appl 31(2):436–443

Tsang ECC, Chen DG, Yeung DS, Wang XZ, Lee JWT (2008) Attributes reduction using fuzzy rough sets. IEEE Trans fuzzy Syst 16(5):1130–1141

Umano M, Okamolo H, Hatono I, Tamura H (1994) Fuzzy decision trees by fuzzy ID3 algorithm and its application to Diagnosis System. Proc Third IEEE Intern Conf Fuzzy Syst 3:2113–2118

Wang XZ, Chen B, Qian G, Ye F (2000) On the optimization of fuzzy decision trees. Fuzzy Sets Syst 112(1):117–125

Wang XZ, Yeung DS, Tsang ECC (2001) A comparative study on heuristic algorithms for generating fuzzy decision trees. IEEE Trans Syst Man Cybern B 31(2):215–226

Wang XM, Nauck DD, Spott M (2007a) Intelligent data analysis with fuzzy decision trees. Soft Comput 11(5):439–457

Wang XZ, Tsang ECC, Zhao SY, Chen DG, Yeung DS (2007b) Learning fuzzy rules from fuzzy examples based on rough set techniques. Inf Sci 177(20):4493–4514

Wang XZ, Zhai JH, Lu SX (2008) Induction of multiple fuzzy decision trees based on rough set technique. Inf Sci 178(16):3188–3202

Yeung DS, Chen DG, Tsang ECC, Lee JWT, Wang XZ (2005) On the generalization of fuzzy rough sets. IEEE Trans Fuzzy Syst 13(3):343–361

Yuan Y, Shaw MJ (1995) Induction of fuzzy decision trees. Fuzzy Sets Syst 69(2):125–139