Information Sciences 178 (2008) 3188–3202

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Induction of multiple fuzzy decision trees based on rough set technique

Xi-Zhao Wang *, Jun-Hai Zhai, Shu-Xia Lu

Key Lab. of Machine Learning and Computational Intelligence, College of Mathematics and Computer Science, Hebei University, Baoding 071002, China

ARTICLE INFO

Keywords:
Learning
Rough sets
Fuzzy attribute reduct
Fuzzy decision tree induction
Fusion
Integral

ABSTRACT

The integration of fuzzy sets and rough sets can lead to a hybrid soft-computing technique which has been applied successfully to many fields such as machine learning, pattern recognition and image processing. The key to this soft-computing technique is how to set up and make use of the fuzzy attribute reduct in fuzzy rough set theory. Given a fuzzy information system, we may find many fuzzy attribute reducts and each of them can have different contributions to decision-making. If only one of the fuzzy attribute reducts, which may be the most important one, is selected to induce decision rules, some useful information hidden in the other reducts for the decision-making will be losing unavoidably. To sufficiently make use of the information provided by every individual fuzzy attribute reduct in a fuzzy information system, this paper presents a novel induction of multiple fuzzy decision trees based on rough set technique. The induction consists of three stages. First several fuzzy attribute reducts are found by a similarity based approach, and then a fuzzy decision tree for each fuzzy attribute reduct is generated according to the fuzzy ID3 algorithm. The fuzzy integral is finally considered as a fusion tool to integrate the generated decision trees, which combines together all outputs of the multiple fuzzy decision trees and forms the final decision result. An illustration is given to show the proposed fusion scheme. A numerical experiment on real data indicates that the proposed multiple tree induction is superior to the single tree induction based on the individual reduct or on the entire feature set for learning problems with many attributes.

1. Introduction

Since the concept of rough set was originally proposed by Pawlak [1] in 1982, rough set theory (RST), as a new mathematical tool for processing incomplete information, has become a popular topic for many researchers and has been applied successfully to many fields [2–5]. An excellent introduction to rough sets can be found in [6], an excellent collection of some extensions of rough sets has been given in [7]. One particular use of rough set theory is attribute reduction in databases. It means, for a given a dataset with discrete attribute values, to find a subset of the original attributes with the most informative (named reduct) and then to remove all other attributes from the dataset so that the information loss is minimum.

Values of attributes in databases are possibly crisp or fuzzy but traditional RST can only handle the crisp case. Fuzzy values are usually transformed into crisp values to handle. One approach to transformation is to preprocess the data set by discretization in which continuous values of attributes are classified into several symbols. The discretization does not consider linguistic terms with membership functions. Noting that real numbers are well ordered but after discretization the symbols are totally non-ordered, the information loss (at least the order information loss) occurs in the process of discretization. To reduce the loss of this type of information, fuzzification has been proposed to stand for the discretization. Fuzzification is first used to fuzzify real attributes for obtaining a number of linguistic terms with semi-order, and then

* Corresponding author.
E-mail addresses: wangxz@hbu.cn (X.-Z. Wang), mczjh@hbu.cn (J.-H. Zhai), mclsx@hbu.cn (S.-X. Lu).

0020-0255/$ - see front matter
doi:10.1016/j.ins.2008.03.021

the attribute reduction technique in RST is used to generate rules for reasoning. Combining RST with fuzzy sets has led to a hybrid soft-computing technique in which fuzzy attribute reducts can be achieved by the use of fuzzy core [8–10].

In RST the attribute reduct is the crucial idea. Like RST fuzzy reducts are also particularly useful in fuzzy rough set approaches [11–13]. More and more researchers have been paying attention to investigating the approaches to find fuzzy attribute reducts in fuzzy information system [3,14,15]. The indiscernibility matrix was generalized to fuzzy rough set, by calculating the relative indiscernibility degree of fuzzy condition attribute set with respect to fuzzy decision attributes using fuzzy indiscernibility matrix, and several algorithms for finding fuzzy attribute reducts were given in [14]. Dependency degree was extended in [15] to fuzzy rough set, which is based on fuzzy positive region defined by the fuzzy equivalence class. By calculating fuzzy dependency degree of Q (fuzzy condition attributes) on P (fuzzy decision attribute), a fuzzy attribute reduct can be found, and an algorithm for finding all fuzzy attribute reducts was given in [15]. Using the algorithm in [15], the most important fuzzy attribute subsets (reducts) could be found to extract decision rules, but some more important subsets (reducts) could be discarded. Because every fuzzy attribute reduct may make significant and different contributions to decision-making, if only one of them is selected to induce decision rules, even if it is the most important one, partial useful information could be lost. In order to make full use of the information provided by every fuzzy attribute reduct, in this paper, we present a novel induction of multiple fuzzy decision trees based on the fuzzy rough set technique, the induction approach consist of three stages. First we find several fuzzy attribute reducts using an improved version based on the algorithm in [15], then a fuzzy decision tree for each fuzzy attribute reduct is generated by fuzzy ID3 algorithm and therefore a number of fuzzy decision trees are obtained. Finally we use the method of multiple classifiers fusion based on fuzzy integral to combine the outputs of the multiple fuzzy decision trees into a final result.

The rest of this paper is organized as follows. In Section 2, some notions and methods related to rough attribute reduction and fuzzy rough attribute reduction are introduced and an improved algorithm for finding reducts is given. In Section 3, the method of single fuzzy decision tree induction is introduced. In Section 4, the method of induction of multiple fuzzy decision trees based on fuzzy integral and fuzzy rough set technique is presented and an example is provided to illustrate our method. An experiment is conducted on a real dataset to verify the effectiveness of the proposed method in Section 5. Finally, Section 6 concludes this paper.

2. Rough set and fuzzy rough set attribute reduction

In this section, we recall some notions and methods related to rough attribute reduction and fuzzy rough attribute reduction, present the improved algorithm. Throughout this paper, we confine ourselves to the consideration only on the finite universe of discourse.

2.1. Rough set attribute reduction

An information system can be represented as a triple $\langle U, A, \{V_a\}_{a\in A}\rangle$, where U is set objects (the universe of discourse), A is set of attributes and V_a is a domain of the attribute a. In a decision system, $A = \{C \cup D\}$, where C is the set of conditional attributes and D is the set of decision attributes. With any $P \subseteq A$ there is an associated equivalence relation$IND(P)$ (called indiscernability relation)

$$IND(P) = \{(x, y) \in U \times U | a(x) = a(y) \text{ for all } a \in P\} \tag{1}$$

The partition of U, generated by $IND(P)$ is denoted by U/P and can be calculated as follows:

$$U/P = \otimes\{a \in P | U/IND\{a\}\} \tag{2}$$

where $A \otimes B = \{X \cap Y | \ \forall X \in A, \forall Y \in B, X \cap Y \neq \phi\}$, the equivalence classes of the P-indiscernibility relation are denoted $[x]_P$.

Let $X \subseteq U$, the P-lower approximation of a set can be defined as:

$$\underline{P}X = \cup\{Y \in U/P, Y \subseteq X\} = \{[x]_P | [x]_P \subseteq X\} \tag{3}$$

Let P and Q be equivalence relations over U, then the P-positive region of Q, denoted by $POS_P(Q)$ can be defined as follows:

$$POS_P(Q) = \cup_{X\in U/Q}\underline{P}X \tag{4}$$

The P-positive region of Q contains all objects of the universe U that can be properly classified into classes of U/Q using the knowledge expressed by the classification U/P.

An important concept in rough set techniques is dependence between attributes; dependency can be defined as follows: For P, $Q \subseteq A$, Q dependency on P in a degree k ($0 \leqslant k \leqslant 1$), denoted $P \Rightarrow_k Q$,

$$k = \gamma_P(Q) = \frac{|POS_P(Q)|}{|U|} \tag{5}$$

where $|S|$ denotes the cardinality of set S.

By calculating the change in dependency when an attribute is removed from the set of considered conditional attributes, a measure of the significance of the attribute can be obtained. Given P, Q and an attribute $x \in P$, the significance of attribute x upon Q is defined by

$$\rho_P(Q) = \gamma_P(Q) - \gamma_{P-\{x\}}(Q) \tag{6}$$

the higher the value $\rho_P(Q)$ is, the more significant the attribute is. If the significance is 0, then the attribute is dispensable.

The reduction of attributes is achieved by comparing equivalence relations generated by sets of attributes. In a decision systems, a reduct is formally defined as a subset R of the conditional attribute set C so that $\gamma_R(D) = \gamma_C(D)$. The intersection of all reduct sets is called the core. For a given data set, there exists only one attribute core, while there are many attribute reducts.

2.2. Fuzzy rough set attribute reduction

A typical fuzzy information system can be represented as Table 1, where $A_1, A_2, \ldots, A_n$ are n conditional attributes and A_i has k_i attribute values $A_{i1}, A_{i2}, \ldots, A_{ik_i}$ $(i = 1, 2, \ldots, n)$. Each attribute value A_{ij} $(i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, k_i)$ can be regarded as a fuzzy set defined on set of the objects $\{1, 2, \ldots, N\}$. That is

$$A_{ij} = \frac{a_{ij}^{(1)}}{1} + \frac{a_{ij}^{(2)}}{2} + \ldots \frac{a_{ij}^{(N)}}{N} \ (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, k_i) \tag{7}$$

which occupies a column in Table 1. The attribute C is the classification (decision) attribute with m attribute values $C_1, C_2, \ldots, C_m$, where

$$C_l = \frac{c_l^{(1)}}{1} + \frac{c_l^{(2)}}{2} + \ldots + \frac{c_l^{(N)}}{N} \ (l = 1, 2, \ldots, m) \tag{8}$$

There are totally N objects, and each objects corresponds to a row in Table 1. The value of the ith object with respect to jth attribute is $\{a_{j_1}^{(i)}, a_{j_2}^{(i)}, \ldots, a_{jk_j}^{(i)}\}$, which can be viewed as a fuzzy set defined on conditional attribute values $\{A_{j1}, A_{j2}, \ldots, A_{jk_j}\}$.

The rough set attribute reduct (RSAR) described in Section 2.1 can operate effectively on information system containing discrete-value attribute, but it can not effectively deal with the fuzzy information with fuzzy-value attribute such as Table 1. By using fuzzy rough set theory [3,11,16], fuzzy attribute reduct can be implemented in fuzzy information system.

Equivalence classes, lower approximation and dependency are central to rough set, in the same way fuzzy equivalence classes, fuzzy lower approximation and fuzzy dependency are central to fuzzy rough set. They are defined as follows respectively [3].

The fuzzy equivalence classes $[x]_S$ for object close to x can be defined as

$$\mu_{[x]_P}(y) = \mu_P(x, y) \tag{9}$$

where $\mu_P(x,y)$ is fuzzy similarity relation. For example, in Table 1, the fuzzy equivalence class $U/P(P=\{A_1\})$ can be represented as

$$U/P = U/\{A_1\} = \{A_{11}, A_{12}, \ldots, A_{1,k_1}\} \tag{10}$$

which can be regarded as fuzzy partition of U by a set of attributes $\{A_1\}$ using fuzzy similarity relations on U.

The fuzzy lower approximation is defined as follows:

$$\mu_{\underline{P}X}(x) = \sup_{F \in U/P} \min(\mu_F(x), \inf_{y \in U} \max\{1 - \mu_F(y), \mu_X(y)\}) \tag{11}$$

each set in U/P denotes an equivalence class.

The membership of an object $x \in U$ belonging to the fuzzy positive region defined by

$$\mu_{POS_P(Q)}(x) = \sup_{x \in U/Q} \mu_{\underline{P}X}(x) \tag{12}$$

Fuzzy dependency degree $\gamma_P(Q)$ $(0 \leqslant \gamma_P(Q) \leqslant 1)$ of Q on the set of attribute P is defined by

$$\gamma_P(Q) = \frac{\sum_{x \in U} \mu_{POS_P(Q)}(x)}{|U|} \tag{13}$$

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
1	$a_{11}^{(1)}$	$a_{12}^{(1)}$	…	$a_{1k_1}^{(1)}$	$a_{21}^{(1)}$	$a_{22}^{(1)}$	…	$a_{2k_2}^{(1)}$	…	$a_{n1}^{(1)}$	$a_{n2}^{(1)}$	…	$a_{nk_n}^{(1)}$	$c_1^{(1)}$	$c_2^{(1)}$	…	$c_m^{(1)}$
2	$a_{11}^{(2)}$	$a_{12}^{(2)}$	…	$a_{1k_1}^{(2)}$	$a_{21}^{(2)}$	$a_{22}^{(2)}$	…	$a_{2k_2}^{(2)}$	…	$a_{n1}^{(2)}$	$a_{n2}^{(2)}$	…	$a_{nk_n}^{(2)}$	$c_1^{(2)}$	$c_2^{(2)}$	…	$c_m^{(2)}$
⋮			…				…		…			…				…	
N	$a_{11}^{(N)}$	$a_{12}^{(N)}$	…	$a_{1k_1}^{(N)}$	$a_{21}^{(N)}$	$a_{22}^{(N)}$	…	$a_{2k_2}^{(N)}$	…	$a_{n1}^{(N)}$	$a_{n2}^{(N)}$	…	$a_{nk_n}^{(N)}$	$c_1^{(N)}$	$c_2^{(N)}$	…	$c_m^{(N)}$

Using the fuzzy dependency degree $\gamma_P(Q)$, an algorithm for generating fuzzy attribute reducts was presented in [15]. But using the algorithm, the most important fuzzy attribute subset (reduct) can be found, while some other more important subsets (reducts) could be discarded. In order to compensate for the inadequacy, in this paper, we propose an improved algorithm based on the algorithm in [15] as follows:

Suppose that $\{A_1, A_2, \ldots, A_n, D\}$ are a fuzzy attribute set in a fuzzy information system, where $A_1, A_2, \ldots, A_n$ are n fuzzy condition attributes, and D is a decision attribute. Let $B_i = \{A_1, A_2, \ldots, A_i\}$.

Step 1: Select one fuzzy condition attribute with the maximum $\gamma_{A_i}(D)$ $(i = 1, 2, \ldots, n)$ as candidate attribute of fuzzy attribute reduct. For the sake of convenience, suppose A_1 is selected.

Step 2: From $\{A_2, \ldots, A_n\}$, select another fuzzy condition attribute with maximum $\gamma_{A_1A_j}(D)$ $(j = 1, 2, \ldots, n)$ and $\gamma_{A_1A_j}(D) > \gamma_{A_1}(D)$ $(j = 2, 3, \ldots, n)$, suppose that A_2 is selected, hence A_1 and A_2 are two candidate attributes of fuzzy attribute reduct.

Step 3: If $B_k = \{A_1, A_2, \ldots, A_k\}$ is selected as candidate attributes of fuzzy attribute reduct and exist A_m $(k < m < n)$ satisfying $\gamma_{B_k \cup \{A_m\}}(D) > \gamma_{B_m}(D)$ $(m = k, k+1, \ldots, n)$, then several fuzzy attribute subsets $\{A_1, A_2, \ldots, A_{i-1}, A_p\}$ $(p = k, k+1, \ldots, n)$ are fuzzy attribute reducts, if $\gamma_{A_1,\ldots,A_{k-1}A_p}(D) > \gamma_{A_1,\ldots,A_{k-1}}(D)$, and different fuzzy attributes reduct are different in their degrees of importance.

3. Single fuzzy decision tree induction

One powerful heuristic for generating crisp decision trees is ID3 algorithm proposed by Quinlan [17,18], which makes a decision tree for classification from the symbolic data. As the increasing uncertainty is incorporated into the knowledge-based system, the fuzzy decision tree suggested by several authors [19–24] will be regarded as a generalization of the crisp case. Recently another decision tree learning algorithm with fuzzy labels was proposed by Qin and Lawry [25].

Fuzzy ID3 [22,24] are popular and efficient methods of making fuzzy decision trees from a group of training examples, which is very similar to ID3. The probability of fuzzy event is used to replace the probability of crisp value for numerical attribute values. The average fuzzy classification entropy is used to measure the disorder of the fuzzified data. According to the average fuzzy classification entropy of each attribute, the most suitable attribute is selected for branching. The average fuzzy classification entropy is defined as follows [22]:

Suppose there are N training examples and n attributes $A^{(1)}, A^{(2)}, \ldots, A^{(n)}$. For each k $(1 \leqslant k \leqslant n)$, the attribute $A^{(k)}$ takes m_k values of fuzzy subsets, $A_1^{(k)}, A_2^{(k)}, \ldots, A_{m_k}^{k}$, $A^{(n+1)}$ denotes the classification (decision) attribute, taking m values $C_1, C_2, \ldots, C_m$, which are also fuzzy sets. We use the symbol $M(A)$ to denote the sum of all membership degrees for a fuzzy set A.

The key of fuzzy ID3 is to select the expanded attribute, which can be performed in the following four steps:

(1) For each attribute value (fuzzy subset) $A_i^k, (i = 1, 2, \ldots, m_k)$, to calculate its relative frequencies with respect to class C_j, $(j = 1, 2, \ldots, m)$ as follows:

$$p_{ij}^{(k)} = M(A_i^{(k)} \cap C_j)/M(A_i^{(k)}) \tag{14}$$

the relative frequency is regarded as the degee of truthfulness of a fuzzy rule in [21].

(2) For each attribute value (fuzzy subset) A_i^k $(i = 1, 2, \ldots, m_k)$, to calculate its fuzzy classification entropy as follows:

$$Entr_i^{(k)} = -\sum_{j=1}^{m} p_{ij}^{(k)} \log_2^{p_{ij}^{(k)}} \tag{15}$$

(3) To calculate the average fuzzy classification entropy of the kth attribute as follows:

$$E_k = \sum_{i=1}^{m_k} \left(M(A_i^{(k)}) \Big/ \sum_{j=1}^{m_k} M(A_j^{(k)}) \right) Entr_i^{(k)} \tag{16}$$

(4) Select such an integer k_0, that

$$E_{k_0} = \min_{1 \leqslant k \leqslant n} \{E_k\} \tag{17}$$

In what follows, we briefly describe the induction based on fuzzy ID3 algorithm [21].

With given evidence significant level α and truth level threshold β, the induction process consists of the following steps:

1. Measure the average fuzzy classification entropy associated with each attribute and select the attribute with the smallest average fuzzy classification entropy as the root decision node.
2. Delete all empty branches of the decision node. For each nonempty branch of the decision node, calculate the truth level of classifying all objects within the branch into each class. If the truth level of classifying into one class is above a given threshold β, terminate the branch as a leaf. Otherwise, do further research if an additional attribute will further partition

the branch (i.e. generate more than one nonempty branch). If yes, select the attribute with the smallest average fuzzy classification entropy as a new decision node from the branch. If not, terminate this branch as a leaf. At the leaf, all objects will be labeled to one class with the highest truth level.

3. Repeat step 2 for all newly generated decision nodes until no further growth is possible, the decision tree then is complete.

By using fuzzy rough set method, several fuzzy attribute reducts of the fuzzy information system can be found, and then for every fuzzy attribute reduct, a fuzzy decision tree can be generated by fuzzy ID3 algorithm mentioned above.

4. Induction of multiple fuzzy decision tree based on fuzzy integral

In this section we first review the concepts of fuzzy measure, fuzzy integral, and method of multiple classifier fusion based on fuzzy integral, then introduce the induction of multiple fuzzy decision tree generated by several fuzzy attribute reducts with fuzzy rough set technique.

4.1. Fuzzy measure and fuzzy integral

Consider the finite space $X = (x_1, x_2, \ldots, x_n)$, we briefly recall the concepts of fuzzy measures and fuzzy integrals.

The concept of fuzzy measure was originally introduced by Sugeno [26] in the early 1970s in order to extend the classical measure through relaxation of the additive property. The fuzzy measure in finite space is defined as follows:

Definition 1 [26]. A set function g: $2^X \rightarrow [0,1]$ is a fuzzy measure if

(1) $g(\phi) = 0$, $g(X) = 1$.
(2) If $A \subseteq B$, then $g(A) \leqslant g(B)$.

A fuzzy measure g_λ is called a Sugeno measure [26], if it satisfies the following additional property for some $\lambda > -1$:
If $A \cap B = \phi$, then $g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) g_\lambda(B)$.
Let $g^i = g(\{x_i\})$, the value g^i is called the density of the measure. The value of λ can be determined by

$$\lambda + 1 = \prod_{i=1}^{n}(1 + \lambda g^i) \tag{18}$$

The value of $g(A_i)$ is computed recursively as follows:

$$g(A_1) = g(\{x_1\}) = g^1 \tag{19}$$

$$g(A_i) = g(\{x_1\}) = g^i + g(A_{i-1}) + \lambda g^i g(A_{i-1}) \ (1 \leqslant i \leqslant n) \tag{20}$$

Definition 2 [27,28]. Let g be a fuzzy measure on X. The discrete Choquet integral of function h: $X \rightarrow R^+$with respect to g is defined as:

$$(c)\int f\,\mathrm{d}\mu = \sum_{i=1}^{n}\{h(x_i) - h(x_{i-1})\}g(A_i) \tag{21}$$

where $0 \leqslant h(x_1) \leqslant \ldots \leqslant h(x_n) \leqslant 1$, and $h(x_0) = 0$.

4.2. Induction of multiple fuzzy decision tree based on fuzzy integral

In this section, multiple decision trees generated from fuzzy attribute reducts are regarded as multiple fuzzy classifiers. The final result is obtained by using fuzzy integral to combine the outputs of the multiple classifiers.

Multiple classifier fusion means that the outputs of all designed classifiers are used to derive a consensus decision. Since different designs of classifiers potentially offer complementary information about patterns to be classified; the fusion of various classifiers can make full use of information about subset of feature space. It has been demonstrated that in many situations combining the outputs on the several classifiers leads to an improved classification result [29–31]

There are many strategies for classifier fusion. The approaches used most often include the majority vote [32,33], average [34]; the weighted average [35,36], the Dempster–Shafer theory [37,38], and the fuzzy integral [39,40]. There is a brief list of the appropriate use of fusion schemes.

If labels are available, majority vote can be easily used.

If posterior probabilities are supplied, average can be used.

If the classifier outputs are interpreted as fuzzy membership values, fuzzy rules and Dempster–Shuffer techniques are suitable for the fusion.

If the fuzzy measure can be determined and the interaction among attributes exists, fuzzy integral is a good approach to multiple classifier fusion, which is selected in this paper as our fusion method.

The method of multiple classifier fusion based on fuzzy integral means that the outputs of all classifier are combined by fuzzy integral to form a final result.

The aggregating process is given in Fig. 1.

Suppose $D = \{D_1, D_2, \ldots, D_L\}$ be the set of Lindividual classifiers and $C = \{c_1, c_2, \ldots, c_n\}$ be the label set of classes, $X = (x_1, x_2, \ldots, x_m)$ be set of samples. For a given $x \in X$, the output of the classifier D_i is a n-dimension vector $(d_{i1}, d_{i2}, \ldots, d_{in})$, where d_{ij} denotes the output of the classifier D_i about x belonging to jth class c_j. For every $x \in X$, a decision profile based on L classifiers is obtained as follows:

$$D = \begin{bmatrix} d_{11}(x) & \ldots & d_{1n}(x) \\ \ldots & \ldots & \ldots \\ d_{L1}(x) & \ldots & d_{Ln}(x) \end{bmatrix}$$

When computing the degree that a sample x belongs to class c_i, the ith column of the decision profile D $(d_{i1}, d_{i2}, \ldots, d_{iL})^{\mathrm{T}}$ is first regarded as L functions on D, and then the fuzzy integral is computed as

$$e_i = (c)\int f_i \,\mathrm{d}\mu_i \;\; (i = 1, 2, \ldots, n) \tag{22}$$

where the integral is computed in terms of (21) and μ_i $(i = 1, 2, \ldots, n)$ are fuzzy measures which can be determined by training examples. Finally the value $\arg\max_{1 \leqslant i \leqslant n}\{e_i\}$, i.e., the degree that a sample x belongs to class c_i, is obtained.

The key to fusion based on fuzzy integral is how to determine the fuzzy measure. Many methods have been proposed by training or based on human designation. In our experiments, we use the method given in [41] to train our g-lamda fuzzy measure.

4.3. An Illustration

In this section, we give an example to illustrate the induction process using the small training data set shown in Table 2. The universe of discourse is $X = \{1, 2, \ldots, 16\}$. Temperature, outlook, humidity, wind and plan are five fuzzy attributes (fuzzy sets), Temperature, outlook, humidity, and wind are fuzzy condition attributes, while plan are fuzzy decision attribute.

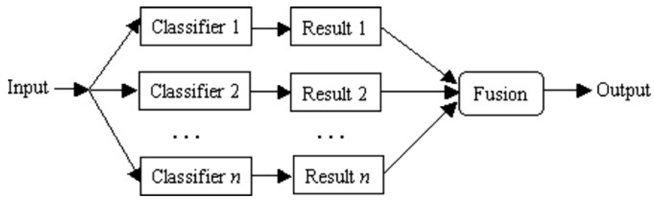

Fig. 1. Multiple classifiers fusion.

Table 2
Small training set with fuzzy representation

No.	Temperature			Outlook			Humidity		Wind		Plan		
	Hot	Mild	Cool	Sunny	Cloudy	Rain	Humid	Normal	Windy	Not-windy	*V*	*S*	*W*
1	0.7	0.2	0.1	1.0	0.0	0.0	0.7	0.3	0.4	0.6	0.0	0.6	0.4
2	0.6	0.2	0.2	0.6	0.4	0.0	0.6	0.4	0.9	0.1	0.7	0.6	0.0
3	0.0	0.7	0.3	0.8	0.2	0.0	0.2	0.8	0.2	0.8	0.3	0.6	0.1
4	0.2	0.7	0.1	0.3	0.7	0.0	0.8	0.2	0.3	0.7	0.9	0.1	0.0
5	0.0	0.1	0.9	0.7	0.3	0.0	0.5	0.5	0.5	0.5	1.0	0.0	0.0
6	0.0	0.7	0.3	0.0	0.3	0.7	0.3	0.7	0.4	0.6	0.2	0.2	0.6
7	0.0	0.3	0.7	0.0	0.0	1.0	0.8	0.2	0.1	0.9	0.0	0.0	1.0
8	0.0	1.0	0.0	0.0	0.9	0.1	0.1	0.9	0.0	1.0	0.3	0.0	0.7
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
10	0.7	0.2	0.1	0.0	0.3	0.7	0.8	0.2	0.9	0.1	0.0	0.3	0.7
11	0.6	0.3	0.1	1.0	0.0	0.0	0.7	0.3	0.2	0.8	0.4	0.7	0.0
12	0.2	0.6	0.2	0.0	1.0	0.0	0.7	0.3	0.7	0.3	0.7	0.2	0.1
13	0.7	0.3	0.0	0.0	0.9	0.1	0.1	0.9	0.0	1.0	0.0	0.4	0.6
14	0.1	0.6	0.3	0.0	0.9	0.1	0.7	0.3	0.7	0.3	1.0	0.0	0.0
15	0.0	0.0	1.0	0.0	0.3	0.7	0.2	0.8	0.8	0.2	0.4	0.0	0.6
16	1.0	0.0	0.0	0.5	0.5	0.0	1.0	0.0	1.0	0.0	0.7	0.6	0.0

Using the improved algorithm in Section 2, we can find several fuzzy attributes reducts of the fuzzy information system shown in Table 2. First of all, the lower approximation needs to be determined. Consider the first attribute in Table 2. Set P={Outlook} produces the fuzzy partition:

$$U/P = \{\{\text{Sunny}\}, \{\text{Cloudy}\}, \{\text{Rain}\}\}$$

Similarly, set Q = {Plan} produces the fuzzy partition:

$$U/Q = \{\{V\}, \{S\}, \{W\}\}$$

set B = {temperature} produces the fuzzy partition:

$$U/B = \{\{\text{Hot}\}, \{\text{Mild}\}, \{\text{Cool}\}\}$$

and set $P \cap B$ = {Outlook} ∩ {Temperature} produces the fuzzy partition:

$$U/(P \cap B) = \{\{\text{Sunny} \cap \text{Hot}\}, \{\text{Sunny} \cap \text{Mild}\}, \{\text{Sunny} \cap \text{Cool}\}, \{\text{Cloudy} \cap \text{Hot}\}, \{\text{Cloudy} \cap \text{Mild}\}, \{\text{Cloudy} \cap \text{Cool}\}, \{\text{Rain} \cap \text{Hot}\}, \{\text{Rain} \cap \text{Mild}\}, \{\text{Rain} \cap \text{Cool}\}\}$$

To determine the fuzzy P-lower approximation of Plan V ($\mu_{\underline{P}V}(x)$), each $F \in U/P$ must be taken into consideration. For F = Sunny,

$$\min(\mu_{\text{Sunny}}(x), \inf_{y \in U} \max\{1 - \mu_{\text{Sunny}}(y), \mu_V(y)\}) = \min(\mu_{\text{Sunny}}(x), 0.0)$$

Similarly, for F = Cloudy, $\min(\mu_{\text{Cloudy}}(x), 0.3)$ and F = Rain, $\min(\mu_{\text{Rain}}(x), 0.0)$. To calculate the extent to which an object $x$in the dataset belongs to the fuzzy P-lower approximation of V, the union of these values is calculated. For example, object 1 belongs to $\underline{P}X$ with a membership of $\mu_{\underline{P}V}$ (1), and,

$$\mu_{\underline{P}V}(1) = \sup\{\min(\mu_{\text{Sunny}}(1), 0.0), \min(\mu_{\text{Rain}}(1), 0.3), \min(\mu_{\text{Cool}}(1), 0.1)\} = 0.0$$

Likewise, for Wind:

$$\mu_{\underline{P}S}(1) = 0.3, \quad \mu_{\underline{P}W}(1) = 0$$

the degree to which object 1 belongs to the fuzzy positive region can be determined by considering the union of fuzzy P-lower approximations:

$$\mu_{POS_P(\text{Plan})}(1) = \sup_{D \in U/\text{Plan}} \mu_{\underline{P}D}(1) = 0.3$$

Similarly, for the remaining objects,

$$\begin{aligned}
&\mu_{POS_P(\text{Plan})}(2) = 0.3, \quad \mu_{POS_P(\text{Plan})}(3) = 0.3, \quad \mu_{POS_P(\text{Plan})}(4) = 0.3, \quad \mu_{POS_P(\text{Plan})}(5) = 0.3,\\
&\mu_{POS_P(\text{Plan})}(6) = 0.6, \quad \mu_{POS_P(\text{Plan})}(7) = 0.6, \quad \mu_{POS_P(\text{Plan})}(8) = 0.1, \quad \mu_{POS_P(\text{Plan})}(9) = 0.3,\\
&\mu_{POS_P(\text{Plan})}(10) = 0.6, \quad \mu_{POS_P(\text{Plan})}(11) = 0.3, \mu_{POS_P(\text{Plan})}(12) = 0.1, \quad \mu_{POS_P(\text{Plan})}(13) = 0.1,\\
&\mu_{POS_P(\text{Plan})}(14) = 0.1, \quad \mu_{POS_P(\text{Plan})}(15) = 0.6, \mu_{POS_P(\text{Plan})}(16) = 0.3
\end{aligned}$$

using these values, the degree of fuzzy dependency of Qon P = {Outlook} can be calculated

$$\gamma_P(\text{Plan}) = \frac{\sum_{x \in U} \mu_{POS_P(Plan)}(x)}{|U|} = \frac{5.2}{16}$$

Hence

$$\gamma_{\text{Outlook}}(\text{Plan}) = \frac{\sum_{x \in U} \mu_{POS_P(\text{Plan})(x)}}{|U|} = \frac{5.2}{16}$$

Similarly, we can calculate the following three values

$$\gamma_{\text{Temperature}}(\text{Plan}) = \frac{4.8}{16}, \gamma_{\text{Humidity}}(\text{Plan}) = \frac{3.2}{16}, \gamma_{\text{Wind}}(\text{Plan}) = \frac{3.2}{16}$$

As attribute "Outlook" causes the greatest increase in fuzzy dependency degree, it is added to the fuzzy attribute reduct candidate.

In order to calculate $\gamma_{\{\text{Outlook}\} \cap \{\text{Temperature}\}}$ (Plan), we must first obtain the membership of each linguistic term in $P \cap B = U/$ {{Outlook} ∩ {Temperature}}, repeat above calculating process, we can have

$$\gamma_{\{\text{Outlook,Temperature}\}}(\text{Plan}) = \frac{9.3}{16}, \gamma_{\{\text{Outlook,Humidity}\}}(\text{Plan}) = \frac{9.4}{16}, \gamma_{\{\text{Outlook,Wind}\}}(\text{Plan}) = \frac{8.4}{16}$$

then the attribute "Humidity" is also added to the fuzzy attribute reduct candidate, now two attributes are added to the fuzzy attribute reduct candidates. Next, we continue to add another member of attributes to the fuzzy attribute reduct candidate.

$$\gamma_{\{\text{Temperature,Outlook,Wind}\}}(\text{Plan}) = \frac{9}{16} \quad \gamma_{\{\text{Temperature,Outlook,Humidity}\}}(\text{Plan}) = \frac{9.8}{16} \quad \gamma_{\{\text{Outlook,Wind,Humidity}\}}(\text{Plan}) = \frac{9.2}{16}$$

while

$$\gamma_{\{\text{Temperature,Outlook,Humidity,Wind}\}}(\text{Plan}) = \frac{9.8}{16}$$

Hence, we obtain the following three conditional fuzzy attributes reducts: {Temperature, Outlook, Humidity}, {Temperature, Outlook Wind} and {Outlook Wind, Humidity}.

Correspondingly, the three fuzzy reduced decision tables, Tables 3–5, are achieved respectively.

In what follows, based on Tables 2–5 four fuzzy decision trees with $\alpha = 0.0$, $\beta = 0.80$ are generated by fuzzy ID3, see Figs. 2–5.

A fuzzy decision tree can be converted into several rules. Based on the rules, we can determine the degree to which class a sample belongs; an object may be classified into different classes with different degrees. The classification for a given object is obtained according to the method in [21].

Due to the three fuzzy attribute reducts have important and different contributions to the fuzzy system, every fuzzy attribute reduct provides us with different aspects of a given fuzzy information system, to make full use of the information

Table 3
Training set of fuzzy samples corresponding to fuzzy attribute reduct 1

No.	Temperature			Outlook			Humidity		Plan		
	Hot	Mild	Cool	Sunny	Cloudy	Rain	Humid	Normal	*V*	*S*	*W*
1	0.7	0.2	0.1	1.0	0.0	0.0	0.7	0.3	0.0	0.6	0.4
2	0.6	0.2	0.2	0.6	0.4	0.0	0.6	0.4	0.7	0.6	0.0
3	0.0	0.7	0.3	0.8	0.2	0.0	0.2	0.8	0.3	0.6	0.1
4	0.2	0.7	0.1	0.3	0.7	0.0	0.8	0.2	0.9	0.1	0.0
5	0.0	0.1	0.9	0.7	0.3	0.0	0.5	0.5	1.0	0.0	0.0
6	0.0	0.7	0.3	0.0	0.3	0.7	0.3	0.7	0.2	0.2	0.6
7	0.0	0.3	0.7	0.0	0.0	1.0	0.8	0.2	0.0	0.0	1.0
8	0.0	1.0	0.0	0.0	0.9	0.1	0.1	0.9	0.3	0.0	0.7
9	1.0	0.0	0.0	1.0	0.0	0.0	0.4	0.6	0.4	0.7	0.0
10	0.7	0.2	0.1	0.0	0.3	0.7	0.8	0.2	0.0	0.3	0.7
11	0.6	0.3	0.1	1.0	0.0	0.0	0.7	0.3	0.4	0.7	0.0
12	0.2	0.6	0.2	0.0	1.0	0.0	0.7	0.3	0.7	0.2	0.1
13	0.7	0.3	0.0	0.0	0.9	0.1	0.1	0.9	0.0	0.4	0.6
14	0.1	0.6	0.3	0.0	0.9	0.1	0.7	0.3	1.0	0.0	0.0
15	0.0	0.0	1.0	0.0	0.3	0.7	0.2	0.8	0.4	0.0	0.6
16	1.0	0.0	0.0	0.5	0.5	0.0	1.0	0.0	0.7	0.6	0.0

Table 4
Training set of fuzzy samples corresponding to fuzzy attribute reduct 2

No.	Temperature			Outlook			Wind		Plan		
	Hot	Mild	Cool	Sunny	Cloudy	Rain	Windy	Not-windy	*V*	*S*	*W*
1	0.7	0.2	0.1	1.0	0.0	0.0	0.4	0.6	0.0	0.6	0.4
2	0.6	0.2	0.2	0.6	0.4	0.0	0.9	0.1	0.7	0.6	0.0
3	0.0	0.7	0.3	0.8	0.2	0.0	0.2	0.8	0.3	0.6	0.1
4	0.2	0.7	0.1	0.3	0.7	0.0	0.3	0.7	0.9	0.1	0.0
5	0.0	0.1	0.9	0.7	0.3	0.0	0.5	0.5	1.0	0.0	0.0
6	0.0	0.7	0.3	0.0	0.3	0.7	0.4	0.6	0.2	0.2	0.6
7	0.0	0.3	0.7	0.0	0.0	1.0	0.1	0.9	0.0	0.0	1.0
8	0.0	1.0	0.0	0.0	0.9	0.1	0.0	1.0	0.3	0.0	0.7
9	1.0	0.0	0.0	1.0	0.0	0.0	0.4	0.6	0.4	0.7	0.0
10	0.7	0.2	0.1	0.0	0.3	0.7	0.9	0.1	0.0	0.3	0.7
11	0.6	0.3	0.1	1.0	0.0	0.0	0.2	0.8	0.4	0.7	0.0
12	0.2	0.6	0.2	0.0	1.0	0.0	0.7	0.3	0.7	0.2	0.1
13	0.7	0.3	0.0	0.0	0.9	0.1	0.0	1.0	0.0	0.4	0.6
14	0.1	0.6	0.3	0.0	0.9	0.1	0.7	0.3	1.0	0.0	0.0
15	0.0	0.0	1.0	0.0	0.3	0.7	0.8	0.2	0.4	0.0	0.6
16	1.0	0.0	0.0	0.5	0.5	0.0	1.0	0.0	0.7	0.6	0.0

Table 5
Training set of fuzzy samples corresponding to fuzzy attribute reduct 3

No.	Outlook			Humidity		Wind		Plan		
	Sunny	Cloudy	Rain	Humid	Normal	Windy	Not-windy	*V*	*S*	*W*
1	1.0	0.0	0.0	0.7	0.3	0.4	0.6	0.0	0.6	0.4
2	0.6	0.4	0.0	0.6	0.4	0.9	0.1	0.7	0.6	0.0
3	0.8	0.2	0.0	0.2	0.8	0.2	0.8	0.3	0.6	0.1
4	0.3	0.7	0.0	0.8	0.2	0.3	0.7	0.9	0.1	0.0
5	0.7	0.3	0.0	0.5	0.5	0.5	0.5	1.0	0.0	0.0
6	0.0	0.3	0.7	0.3	0.7	0.4	0.6	0.2	0.2	0.6
7	0.0	0.0	1.0	0.8	0.2	0.1	0.9	0.0	0.0	1.0
8	0.0	0.9	0.1	0.1	0.9	0.0	1.0	0.3	0.0	0.7
9	1.0	0.0	0.0	0.4	0.6	0.4	0.6	0.4	0.7	0.0
10	0.0	0.3	0.7	0.8	0.2	0.9	0.1	0.0	0.3	0.7
11	1.0	0.0	0.0	0.7	0.3	0.2	0.8	0.4	0.7	0.0
12	0.0	1.0	0.0	0.7	0.3	0.7	0.3	0.7	0.2	0.1
13	0.0	0.9	0.1	0.1	0.9	0.0	1.0	0.0	0.4	0.6
14	0.0	0.9	0.1	0.7	0.3	0.7	0.3	1.0	0.0	0.0
15	0.0	0.3	0.7	0.2	0.8	0.8	0.2	0.4	0.0	0.6
16	0.5	0.5	0.0	1.0	0.0	1.0	0.0	0.7	0.6	0.0

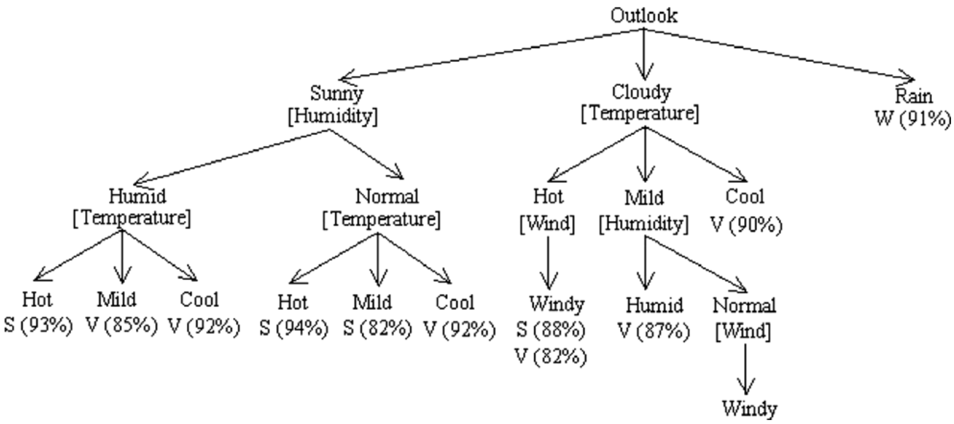

Fig. 2. The fuzzy decision tree 1 for Table 2.

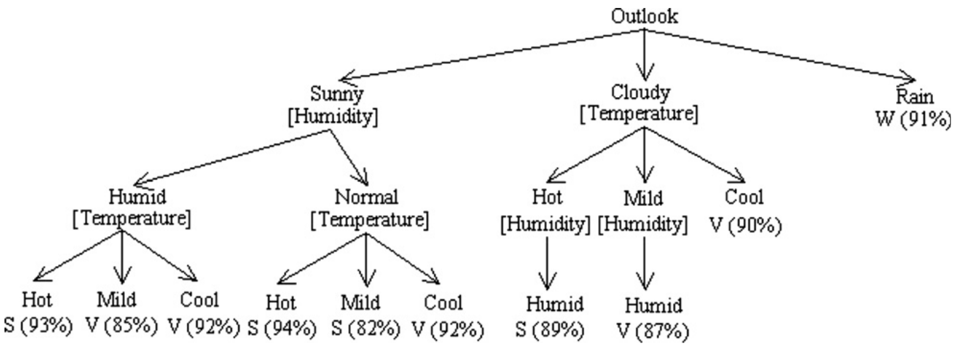

Fig. 3. The fuzzy decision tree 2 for Table 3.

provided by every fuzzy attribute reduct, we use the method of multiple classifiers fusion based on fuzzy integral to combines the outputs of the three fuzzy decision trees into a consensus result.

In the following, based on Choquet integral, we will aggregate the outputs of the three fuzzy decision trees. Here a fuzzy decision tree can be viewed as a fuzzy classifier.

First of all, we determine the fuzzy measures for every class. The important degree of the every fuzzy decision tree to every class can be determined using Tables 6–8; they can be viewed as fuzzy density. For example, the fuzzy density with

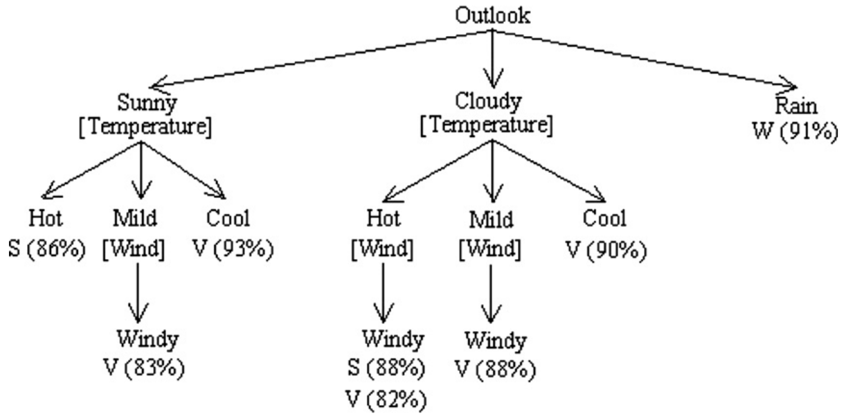

Fig. 4. The fuzzy decision tree 3 for Table 4.

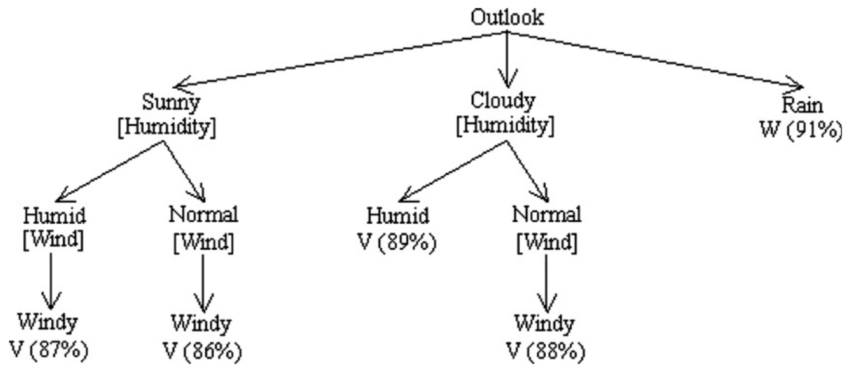

Fig. 5. The fuzzy decision tree 4 for Table 5.

respect to every class (V, S, W) of the fuzzy decision tree1 is respectively 0.406, 0.419 and 0.481. The fuzzy densities of the other two fuzzy decision trees can be similarly obtained. Based on the fuzzy densities, the value λ for every class can be determined by expression (18) in Section 4.1. According to expression (20), the fuzzy measures of the three fuzzy decision tree classifiers can be achieved as in Table 9, where D_i (i = 2, 3, 4) denotes the ith classifier (fuzzy decision tree i), μ_i (i = 1, 2, 3) denotes fuzzy measure of the ith class.

And then, for every $x \in X$, a decision profile can be determined based on the three fuzzy decision trees. Finally according to the Choquet integral (expression (21) in Section 4.1), the vector of the output of the three fuzzy decision trees can be obtained. For example, for sample 1, the decision profile is as follows:

$$\begin{bmatrix} & \mu_1 & \mu_2 & \mu_3 \\ D_1 & 0.2 & 0.7 & 0.0 \\ D_2 & 0.2 & 0.7 & 0.0 \\ D_3 & 0.4 & 0.0 & 0.0 \end{bmatrix}$$

Using Choquet integral, we have

$$e_V = 0.2(1 - 0.735) + 0.2(0.735 - 0.460) + 0.4 \times 0.460 = 0.28$$
$$e_S = 0.0(1 - 0.765) + 0.7(0.765 - 0.419) + 0.4 \times 0.419 = 0.54$$
$$e_W = 0$$

so the vector of the output is (0.28, 0.54, 0), selecting the class with the highest membership as the class label, so that sample 1 belongs to class S in Plan.

The classification results by fuzzy decision 1 and classification results by fusion of other three fuzzy decision trees (fuzzy decisions 2, 3 and 4) are shown in Table 10. Among 16 training cases, 14 cases are correctly classified (samples 2 and 16 are not correctly classified) by the method of the three fuzzy decision trees fusion. The classification accuracy is 87.5%, greater than the original classification 75% by fuzzy decision tree 1 (samples 2, 8, 13, and 16 are not correctly classified). The true degree of the fuzzy rules is more than 80%. Obviously, the classification accuracy is improved after aggregating the three

Table 6
Membership of {Outlook} ∩ {Temperature} ∩ {Humidity}

No.	Hot Sunny Humid	Hot Sunny Normal	Hot Cloudy Humid	Hot Cloudy Normal	Hot Rain Humid	Hot Rain Normal	Mild Sunny Humid	Mild Sunny Normal	Mild Cloudy Humid	Mild Cloudy Normal	Mild Rain Humid	Mild Rain Normal	Cool Sunny Humid	Cool Sunny Normal	Cool Cloudy Humid	Cool Cloudy Normal	Cool Rain Humid	Cool Rain Normal	*V*	*S*	M
1	0.7	0.3	0.0	0.0	0.0	0.0	0.2	0.2	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.6	0.4
2	0.6	0.4	0.4	0.4	0.0	0.0	0.2	0.2	0.2	0.2	0.0	0.0	0.2	0.2	0.2	0.2	0.0	0.0	0.7	0.6	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.7	0.2	0.2	0.0	0.0	0.2	0.3	0.2	0.2	0.0	0.0	0.3	0.6	0.1
4	0.2	0.2	0.2	0.2	0.0	0.0	0.3	0.2	0.7	0.2	0.0	0.0	0.1	0.1	0.1	0.1	0.0	0.0	0.9	0.1	0.0
5	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.0	0.0	0.5	0.5	0.3	0.3	0.0	0.0	1.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.3	0.3	0.7	0.0	0.0	0.3	0.3	0.3	0.3	0.2	0.2	0.6
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.2	0.0	0.0	0.0	0.0	0.7	0.2	0.0	0.0	1.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.8	0.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.7
9	0.4	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.7	0.0
10	0.0	0.0	0.3	0.2	0.7	0.2	0.0	0.0	0.2	0.2	0.2	0.2	0.0	0.0	0.1	0.1	0.1	0.1	0.0	0.3	0.7
11	0.6	0.3	0.0	0.0	0.0	0.0	0.3	0.3	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.4	0.7	0.0
12	0.0	0.0	0.2	0.2	0.0	0.0	0.0	0.0	0.6	0.3	0.0	0.0	0.0	0.0	0.2	0.2	0.0	0.0	0.7	0.2	0.1
13	0.1	0.2	0.1	0.7	0.0	0.0	0.1	0.2	0.1	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.6
14	0.0	0.0	0.1	0.1	0.1	0.1	0.0	0.0	0.6	0.3	0.1	0.1	0.0	0.0	0.3	0.3	0.1	0.1	1.0	0.0	0.0
15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.3	0.2	0.7	0.4	0.0	0.6
16	0.5	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.6	0.0

Table 7
Membership of {Outlook} ∩ {Temperature} ∩ {Wind}

No.	Hot Sunny Windy	Hot Sunny N-Windy	Hot Cloudy Windy	Hot Cloudy N-Windy	Hot Rain Windy	Hot Rain N-Windy	Mild Sunny Windy	Mild Sunny N-Windy	Mild Cloudy Windy	Mild Cloudy N-Windy	Mild Rain Windy	Mild Rain N-Windy	Cool Sunny Windy	Cool Sunny N-Windy	Cool Cloudy Windy	Cool Cloudy N-Windy	Cool Rain Windy	Cool Rain N-Windy	*V*	*S*	M
1	0.4	0.6	0.0	0.0	0.0	0.0	0.2	0.2	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.6	0.4
2	0.6	0.1	0.4	0.1	0.0	0.0	0.2	0.1	0.2	0.1	0.0	0.0	0.2	0.1	0.2	0.1	0.0	0.0	0.7	0.6	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.7	0.2	0.2	0.0	0.0	0.2	0.3	0.2	0.2	0.0	0.0	0.3	0.6	0.1
4	0.2	0.2	0.2	0.2	0.0	0.0	0.3	0.3	0.3	0.7	0.0	0.0	0.1	0.1	0.1	0.1	0.0	0.0	0.9	0.1	0.0
5	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.0	0.0	0.5	0.5	0.3	0.3	0.0	0.0	1.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.3	0.4	0.6	0.0	0.0	0.3	0.3	0.3	0.3	0.2	0.2	0.6
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.3	0.0	0.0	0.0	0.0	0.1	0.7	0.0	0.0	1.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.7
9	0.4	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.7	0.0
10	0.0	0.0	0.3	0.1	0.7	0.1	0.0	0.0	0.2	0.1	0.2	0.1	0.0	0.0	0.1	0.1	0.1	0.1	0.0	0.3	0.7
11	0.2	0.6	0.0	0.0	0.0	0.0	0.2	0.3	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.4	0.7	0.0
12	0.0	0.0	0.2	0.2	0.0	0.0	0.0	0.0	0.6	0.3	0.0	0.0	0.0	0.0	0.2	0.2	0.0	0.0	0.7	0.2	0.1
13	0.0	0.2	0.0	0.7	0.0	0.0	0.0	0.2	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.6
14	0.0	0.0	0.1	0.1	0.1	0.1	0.0	0.0	0.6	0.3	0.1	0.1	0.0	0.0	0.3	0.3	0.1	0.1	1.0	0.0	0.0
15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.2	0.7	0.2	0.4	0.0	0.6
16	0.5	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.6	0.0

Table 8
Membership of {Outlook} ∩ {Humidity} ∩ {Wind}

No.	Sunny Humid Windy	Sunny Humid N-Windy	Sunny Normal Windy	Sunny Normal N-Windy	Cloudy Humid Windy	Cloudy Humid N-Windy	Cloudy Normal Windy	Cloudy Normal N-Windy	Rain Humid Windy	Rain Humid N-Windy	Rain Normal Windy	Rain Normal N-Windy	*V*	*S*	M
1	0.4	0.6	0.3	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.4
2	0.0	0.6	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.6	0.0
3	0.2	0.8	0.1	0.1	0.2	0.2	0.1	0.1	0.0	0.0	0.0	0.0	0.3	0.6	0.1
4	0.3	0.3	0.2	0.2	0.3	0.7	0.2	0.2	0.0	0.0	0.0	0.0	0.9	0.1	0.0
5	0.5	0.5	0.5	0.5	0.3	0.3	0.3	0.3	0.0	0.0	0.0	0.0	1.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.3	0.3	0.3	0.3	0.4	0.6	0.3	0.3	0.2	0.2	0.6
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.8	0.1	0.2	0.0	0.0	1.0
8	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.2	0.0	0.8	0.0	0.8	0.3	0.0	0.7
9	0.4	0.4	0.6	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.7	0.0
10	0.0	0.0	0.0	0.0	0.3	0.1	0.1	0.1	0.7	0.1	0.1	0.1	0.0	0.3	0.7
11	0.0	0.7	0.2	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.7	0.0
12	0.0	0.0	0.0	0.0	0.3	0.7	0.3	0.3	0.0	0.0	0.0	0.0	0.7	0.2	0.1
13	0.1	0.0	0.2	0.0	0.1	0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.6
14	0.0	0.0	0.0	0.0	0.7	0.3	0.1	0.1	0.1	0.1	0.1	0.1	1.0	0.0	0.0
15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.2	0.4	0.0	0.6
16	0.0	0.5	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.6	0.0

Table 9
Fuzzy measures of three classifiers

	μ_1	μ_2	μ_3
D_1	0.406	0.419	0.481
D_2	0.413	0.438	0.463
D_3	0.406	0.375	0.463
D_1D_2	0.735	0.765	0.725
D_1D_3	0.730	0.715	0.725
D_2D_3	0.735	0.730	0.776
$D_1D_2D_3$	1.00	1.00	1.00

Table 10
Comparison classification results among known data, fuzzy decision 1, and multiple fuzzy decision tree fusion

No.	Classification known in training data			Classification results with fuzzy decision 1			Classification results with multiple fuzzy decision tree fusion		
	V	*S*	*W*	*V*	*S*	*W*	*V*	*S*	*W*
1	0.0	0.6	0.4	0.2	0.7	0.0	0.28	0.54	0.00
2	0.7	0.6	0.0	0.2	0.6	0.0	0.36	0.46	0.00
3	0.3	0.6	0.1	0.3	0.7	0.0	0.27	0.36	0.00
4	0.9	0.1	0.0	0.7	0.3	0.0	0.59	0.23	0.00
5	1.0	0.0	0.0	0.5	0.1	0.0	0.58	0.04	0.00
6	0.2	0.2	0.6	0.3	0.0	0.7	0.22	0.13	0.70
7	0.0	0.0	1.0	0.0	0.0	1.0	0.00	0.00	1.00
8	0.3	0.0	0.7	0.1	0.0	0.1	0.07	0.00	0.10
9	0.4	0.7	0.0	0.0	0.6	0.0	0.16	0.64	0.00
10	0.0	0.3	0.7	0.2	0.3	0.7	0.20	0.22	0.51
11	0.4	0.7	0.0	0.3	0.6	0.0	0.24	0.46	0.00
12	0.7	0.2	0.1	0.6	0.2	0.0	0.56	0.15	0.00
13	0.0	0.4	0.6	0.1	0.0	0.1	0.07	0.04	0.10
14	1.0	0.0	0.0	0.6	0.1	0.1	0.64	0.08	0.10
15	0.4	0.0	0.6	0.3	0.0	0.7	0.30	0.00	0.70
16	0.7	0.6	0.0	0.5	0.5	0.0	0.20	0.37	0.00

fuzzy decision trees. A fuzzy attribute reduct is an important set of attributes and every fuzzy attribute reduct provides us with different aspects of the system. By aggregating their outputs, the classification accuracy can be improved.

5. An experiment on real dataset

A dataset is 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 dataset has 170 normal cases and 42 abnormal cases. Totally 35 features are initially selected. They are 10 symmetric features, nine texture features and 16 statistical features including mean, variance, skewness, kurtosis, energy and entropy.

We use the cross-validation procedure. Each database is randomly partitioned into 10 disjoint subsets; each size is $m/10$ where m is the number of examples of the dataset. The procedure is then conducted 10 times, each time using a different one of these subsets as the validation (testing) set and combining the other nine subsets for the training set. The training and testing accuracies are then averaged. For each time, the experimental procedure is the following:

(A) Fuzzify the database according to a specified fuzzification procedure which is selected in this experiment as the triangular membership forms given in [21], then generate a single fuzzy decision tree.
(B) Calculate all fuzzy attribute reducts based on the algorithm listed at the end of Section 2.
(C) Generate a fuzzy decision tree for each reduct based on the fuzzy ID3 algorithm listed in Section 3.
(D) Convert each fuzzy decision tree into a set of fuzzy IF-THEN rules and calculate the corresponding training and testing accuracies.
(E) Fuse the individual outputs of each fuzzy decision tree based on the algorithm given in Section 4.2.

Our experimental results (averaged training and testing accuracies for the 10 times) together with the results given by the single fuzzy ID3 are summarized as follows:

	Training accuracy	Testing accuracy
Single fuzzy ID3	0.93	0.91
Our fusion scheme	0.98	0.97

Averagely the experimental results show that the fusion of multiple decision trees with respect to different reducts of attributes are superior to the single decision tree induction (single fuzzy ID3). The most difficult problem for the integral-based fusion is the determination of fuzzy measures. In this paper we select a training procedure given in [41] to obtain the g-lamda fuzzy measure.

6. Conclusion

Fuzzy attribute reduct is a set of important fuzzy attributes that plays a key role in a fuzzy information system. For a given fuzzy information system, we could find several fuzzy attribute reducts rather than only one, and each fuzzy attribute reduct could provide us with some important information from different aspects. If only one of them is selected to induce decision rules, even if it is the most important one, much useful information could be lost. In order to make full use of the information provided by every fuzzy attribute reduct, in this paper, we present a method of induction of multiple fuzzy decision trees based on fuzzy rough set, and use the fuzzy integral as the aggregating tool to combine the outputs of the multiple fuzzy decision trees into a consensus result. The experimental results confirm the effectiveness of the proposed method.

Acknowledgements

This research is supported by the National Natural Science Foundation of China (60473045), the Natural Science Foundation of Hebei Province (F2008000635), the Scientific Research Foundation of Hebei Province (06213548), Hebei Hi-tech project (04213533) and the plan of 100 excellent innovative scientists in Hebei Province.

References

[1] Z. Pawlak, Rough sets, International Journal of Information and Computer Sciences 11 (1982) 341–356.
[2] Rajen B. Bhatt, M. Gopal, On fuzzy-rough sets approach to feature selection, Pattern Recognition Letters 26 (2005) 965–975.
[3] Richard Jensen, Qiang Shen, Fuzzy-rough attribute reduction with application to web categorization, Fuzzy Sets and Systems 141 (2004) 469–485.
[4] AboulElla Hassanien, Fuzzy rough sets hybrid scheme for breast cancer detection, Image and Vision Computing 25 (2007) 172–183.
[5] Z. Pawlak, A. Skowron, Rough sets and Boolean reasoning, Information Sciences 177 (2006) 41–73.
[6] Z. Pawlak, A. Skowron, Rudiments of rough sets, Information Science 177 (2007) 3–27.
[7] Z. Pawlak, A. Skowron, Rough sets: some extension, Information Science 177 (2007) 28–40.
[8] D. Dubois, H. Prade, Putting rough sets and fuzzy sets together, intelligent decision support, in: R. Slowinski (Ed.), Handbook of Applications and Advances of the Rough Sets theory, Kluwer Academic Publishers, 1992.
[9] R.W. Swiniarski, A. Skowron, Rough set methods in feature selection and recognition, Pattern Recognition Letters 24 (2003) 833–849.
[10] Tingquan Deng, Yanmei Chen, Wenli Xu, Qionghai Dai, A novel approach to fuzzy rough sets based on a fuzzy covering, Information Sciences 177 (2007) 2308–2326.
[11] H. Thiele, Fuzzy rough sets versus rough fuzzy sets—an interpretation and a comparative study using concepts of modal logics, Tech. Report no. CI-30/98, University of Dortmund, 1998.
[12] Y.Y. Yao, A comparative study of fuzzy sets and rough sets, Information Science 109 (1988) 21–47.
[13] M. Beynon, Reducts within the variable precision rough sets model: a further investigation, European Journal of Operational Research 134 (2001) 592–605.
[14] Xi-Zhao Wang, Eric Tsang, Su-Yun Zhao, De-Gang Chen, Daniel Yeung, Learning fuzzy rules from fuzzy examples based on rough set techniques, Information Sciences 177 (2007) 4493–4514.
[15] Richard Jensen, Qiang Shen, Fuzzy-rough data reduction with ant colony optimization, Fuzzy Sets and Systems 149 (2005) 5–20.
[16] S.K. Pal, A. Skowron (Eds.), Rough-Fuzzy Hybridization: A New Trend in Decision Making, Springer, Singapore, 1999.

[17] J.R. Quinlan, Discovering Rule by Induction from large Collection of Examples, Expert Systems in the Micro Electronics Age, Edinburgh University Press, 1979.
[18] J.R. Quinlan, Induction of decision tree, Machine Learning 1 (1986) 81–106.
[19] H. Ichihashi, T. Shirai, K. Nagasaka, T. Miyoshi, Neuro-fuzzy ID3, Fuzzy Sets and Systems 81 (1996) 157–167.
[20] B. Jeng, Y.-M. Jeng, T.-P. Liang, FILM: a fuzzy inductive learning method for automated knowledge acquisition, Decision Support Systems 21 (1997) 61–73.
[21] Yufei Yuan, Michael J. Shaw, Induction of fuzzy decision trees, Fuzzy Sets and Systems 69 (1995) 125–139.
[22] Xizhao Wang, Bin Chen, Guoliang Qian, Feng Ye, On the optimization of fuzzy decision trees, Fuzzy Sets and Systems 112 (2000) 117–125.
[23] M. Umanol, H. Okamoto, I. Hatono, H. Tamura, F. Kawachi, S. Umedzu, J. Kinoshita, Fuzzy decision trees by fuzzy ID3 algorithm and its application to diagnosis systems, in: IEEE World Congress on Computational Intelligence, Proceedings of the Third IEEE Conference on Fuzzy Systems, 26–29 June 1994, vol. 3, 1994, pp. 2113–2118.
[24] Weber, Fuzzy-ID3: a class of methods for automatic knowledge acquisition, in: 2nd International Conference on Fuzzy Logic and Neural Networks, Iizuka, Japan, July 17–22, 1992, pp. 265–268.
[25] Zengchang Qin, Jonathan Lawry, Decision tree learning with fuzzy labels, Information Science 173 (2005) 255–275.
[26] M. Sugeno, Theory of fuzzy integrals and its applications, Ph.D. dissertation, Tokyo Institute of Technology, 1974.
[27] M. Sugeno, Fuzzy measures and fuzzy integrals: a survey, in: Gupta, Saridis, Gaines (Eds.), Fuzzy Automata and Decision Processes, 1977, pp. 89–102.
[28] G. Choquet, Theory of capacities, Annales de l Institut Fourier 5 (1954) 131–295.
[29] D. Opitz, R. Maclin, Popular ensemble methods: an empirical study, Journal of Artificial Intelligence Research 11 (1999) 169–198.
[30] G. Rogova, Combining the results of several neural network classifiers, Neural Network 7 (5) (1994) 777–781.
[31] D. Tax, R. Duin, M. Breukelen, Comparison between product and mean classifier combination rules, in: Workshop on statistical Techniques in Pattern Recognition. Prague, Czech Republic, 1997.
[32] R. Battiti, M. Colla, Democracy in neural nets: voting schemes for classification, Neural Networks 7 (4) (1994) 691–707.
[33] C. Ji, S. Ma, Combined weak classifiers, in: M.C. Ozer, M.I. Jordan, T. Petsche (Eds.), Advances in Neural Information Processing Systems, vol. 9, MIT Press, Cambridge, 1997, pp. 494–500.
[34] C.J. Merz, M.J. Pazzani, Combining neural network regression estimates with regularized linear weights, in: M.C. Mozer, M.I. Jordan, T. Petsche (Eds.), Advances in Neural Information Processing Systems, vol. 9, MIT Press, Cambridge, 1997, pp. 564–570.
[35] S. Hashem, Optimal linear combinations of neural networks, Neural Networks 10 (4) (1997) 599–614.
[36] L. Wang, K. Chen, H. Chi, Methods of linear combination based on different features, in: Proceedings of the International Conference on Neural Information Processing, Dunedin, New Zealand, vol. 2, 1997, pp. 1088–1091.
[37] L. Xu, A. Krzyzak, C.Y. Suen, Methods for combining multiple classifiers and their applications to handwriting recognition, IEEE Transactions on Systems, Man, and Cybernetics 22 (3) (1992) 418–435.
[38] T. Denoeux, A *k*-nearest neighbor classification rule based on Dempster–Shafer theory, IEEE Transactions on System, Man and Cybernetics – Part B 25 (5) (1995) 804–813.
[39] M. Grabisch, Fuzzy integral in multicriteria decision-making, Fuzzy Sets and Systems 69 (1995) 279–298.
[40] M. Grabisch, The representation of importance and interaction of features by fuzzy measures, Pattern Recognition Letters 17 (1996) 567–657.
[41] W. Wang, Z. Wang, G.J. Klir, Genetic algorithms for determining fuzzy measures from data, Journal of Intelligent and Fuzzy Systems: Applications in Engineering and Technology 6 (2) (1998) 171–183.