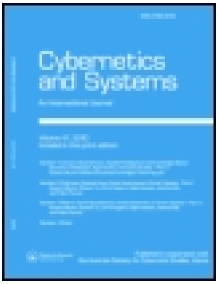

Cybernetics and Systems: An International Journal
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ucbs20

A NEW METHOD FOR CONSTRUCTING FUZZY DECISION TREES AND GENERATING FUZZY CLASSIFICATION RULES FROM TRAINING EXAMPLES

Shyi-Ming Chen, Shih-Yirng Lin
[a] Department of Electronic Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, R.O.C.
[b] Department of Computer and Information Science, National Chiao Tung University, Hsinchu, Taiwan, R. O. C.
Published online: 29 Oct 2010.

Cybernetics and Systems: An International Journal, 31:763–785, 2000

0196-9722/00 $12.00 + .00

A NEW METHOD FOR CONSTRUCTING FUZZY DECISION TREES AND GENERATING FUZZY CLASSIFICATION RULES FROM TRAINING EXAMPLES

SHYI-MING CHEN

Department of Electronic Engineering,
National Taiwan University of Science and Technology,
Taipei, Taiwan, R.O.C.

SHIH-YIRNG LIN

Department of Computer and Information Science,
National Chiao Tung University,
Hsinchu, Taiwan, R. O. C.

This paper presents a new method for constructing fuzzy decision trees and generating fuzzy classification rules from training instances using compound analysis techniques. The proposed method can generate simpler fuzzy classification rules and has a better classification accuracy rate than the existing method. Furthermore, the proposed method generated less fuzzy classification rules.

In recent years, there were many researchers focusing on the research of the inductive learning for rules generation, e.g., Hart (1995), Hunt et al. (1966), Minger (1989), Quinlan (1979, 1986), and Yasdi (1991).

This work was supported in part by the National Science Council, Republic of China, under Grant NSC 89-2213-E-011-060.

Address correspondence to Professor Shyi-Ming Chen, Department of Electronic Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, R.O.C.

The extended models using fuzzy set theory (Zadeh, 1965) for fuzzy rules generation can be found in Chang and Pavlidis (1977), Boyen and Wehenkel (1999), Chen and Yeh (1997), Lin and Chen (1996), Sun (1995), Wu and Chen (1999), Yasdi (1991), and Yuan and Shaw (1995). Chang and Pavlidis (1977) presented fuzzy decision tree algorithms. Boyen and Wehenkel (1999) presented a method to fuzzy tree induction from examples for power system security assessment. In Chen and Yeh (1997), the authors have presented a method for generating fuzzy rules from relational database systems for estimating null values by constructing fuzzy decision trees. In Lin and Chen (1996), the authors have presented a method for generating fuzzy rules from fuzzy decision trees. In Sun (1995) a fuzzy approach to decision trees was presented. In Wu and Chen (1999) the authors have presented a method for constructing membership functions and fuzzy rules from training examples. Yasdi (1991) presented a method for learning classification rules from a database in the context of knowledge acquisition and representation. Yuan and Shaw (1995) presented a method for generating fuzzy rules from fuzzy decision trees.

There are some problems with the traditional ID3 learning method (Quinlan, 1986). For example, it cannot deal with cognitive uncertainties (Yuan & Shaw, 1995) such as vagueness and ambiguity associated with human thinking and perception. Furthermore, it is sensitive to "noise" (Sun, 1995). Yuan and Shaw (1995) proposed an algorithm combined with fuzzy logic to solve these problems. However, there are some drawbacks in the method presented in Yuan and Shaw (1995) described as follows:

(1) It takes much computation time to find the "entropy" of attributes by using a more complex test function.
(2) Yuan and Shaw's method generated more decision nodes and induced more fuzzy rules.

In this article, a new method for constructing fuzzy decision tree and generating fuzzy rules from training instances was presented, using compound analysis techniques. It is a simpler, more efficient, and more effective method, where a simple test function is used to find the entropy of attributes; instead of using fuzzy subsethood (Sun, 1995), the correctness of classification is used as the criteria (Jeng & Liang, 1993) to stop the expansion of the fuzzy decision trees. In the process of constructing fuzzy decision trees, the proposed method does not directly find the entropy of attributes, but tries to find some factors that are negative to each class. The proposed method can overcome the drawbacks of the one presented in Yuan and Shaw (1995).

BASIC CONCEPTS OF CONSTRUCTING FUZZY DECISION TREES AND GENERATING FUZZY CLASSIFICATION RULES

Quinlan (1986) proposed a learning algorithm called ID3 to construct decision trees from a set of training instances based on information theory. A decision tree is a tree supporting an inference for classifications of all possible instances, where every path from the root to each terminal node forms a rule.

A fuzzy decision tree is an extension of Quinlan's decision tree. It can avoid unexpected results caused by "noise," which might take place with a nonfuzzy approach, and it can deal with cognitive uncertainty such as vagueness and ambiguity associated with human thinking and perception.

Before constructing a fuzzy decision tree, one must fuzzify the training data set by applying the concepts of fuzzy sets (Chen, 1986; Zadeh, 1965, 1975) and fuzzification. One can construct a fuzzy decision tree from the training data set and then generate fuzzy classification rules from the constructed fuzzy decision tree. The objective is to generate a set of fuzzy classification rules having the following form:

$$\text{If A is } A_i \text{ AND B is } B_j \text{ THEN C is } C_k \tag{1}$$

where A and B are attributes of an instance; A_i and B_j are linguistic terms of A and B, respectively; C_k is a class term of the classification attribute C. Then, one can forecast the value of the class term C_k of the classification attribute C. Let a and b be membership values of the instance in A_i and B_j, respectively and let c be the forecasted value of C_k, where $c = \min(a, b)$ and "min" is the minimum operator. When two or more rules have the same conclusion (i.e., they all conclude that "IF conditions THEN C is C_k"), then one generates several forecasted values with respect to C_k, and one takes the largest one.

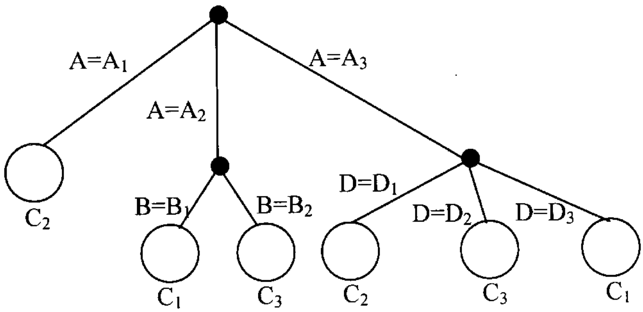

Figure 1. A fuzzy decision tree.

Figure 1 shows an example of a fuzzy decision tree. From Figure 1, one can obtain the fuzzy rules shown as follows:

IF A is A_1 THEN C is C_2
IF A is A_2 AND B is B_1 THEN C is C_1
IF A is A_2 AND B is B_2 THEN C is C_3
IF A is A_3 AND D is D_1 THEN C is C_2
IF A is A_3 AND D is D_2 THEN C is C_3
IF A is A_3 AND D is D_3 THEN C is C_1

where

(1) C_1, C_2, and C_3 are class terms of the classification attribute C.
(2) A, B, and D are attributes; A_1, A_2, and A_3 are linguistic terms of the attribute A; B_1 and B_2 are linguistic terms of the attribute B; and D_1, D_2, and D_3 are linguistic terms of the attribute D, respectively.

A NEW METHOD FOR CONSTRUCTING FUZZY DECISION TREES AND GENERATING FUZZY RULES

In the following, a new method to generate fuzzy classification rules is proposed based on constructing fuzzy decision trees using compound analysis techniques. One generalizes the traditional ID3 algorithm (Quinlan, 1986) to deal with cognitive uncertainties such as vagueness and ambiguity associated with human thinking and perception. In addition, combining with fuzzy logic, one can reduce the sensitivity to "noise." Most important of all one tries to find influential factors that can directly produce classifications. Furthermore, removing instances that are dominated by these influential factors, one might find a weaker factor that can produce a classification beyond the effects of the influential factors. After taking these into consideration, one can construct fuzzy decision trees with less edges and decision nodes.

First, the definitions of significance level are briefly reviewed (Yuan & Shaw, 1995) the correctness of the classification (Jeng & Liang, 1993), and the formula for calculating the entropy of attributes (Sun, 1995), respectively.

Definition 3.1: Given a fuzzy set A of the universe of discourse U with the membership function μ_A, μ_A: U→[0,1]. The α-significance level A_α of the fuzzy set A is defined as follows:

$$\mu_{A_\alpha}(u) = \begin{cases} \mu_A(u), & \text{if } \mu_A(u) \geq \alpha \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

where $0 \leq \alpha \leq 1$.

Example 3.1: Under the significance level $\alpha = 0.5$, one can translate Table 1 into Table 2 as shown.

The translated fuzzy relation shown in Table 2 can be further reduced into a reduced fuzzy relation such that each attribute's value of the cases contains the set of linguistic terms whose membership values in Table 2 are larger than the significance level $\alpha = 0.5$. In this case, Table 2 can be translated into a reduced fuzzy relation as shown in Table 3.

From Table 1, one can see that "outlook," "temperature," "humidity," and "wind" are the attributes of the fuzzy relation, whose values are the sets of linguistic terms {sunny, cloudy, rain}, {hot, mild, cool}, {humid, normal}, {windy, not windy}, respectively; "plan" is called the classification attribute whose values are "volleyball," "swimming," and "weightlifting," where "volleyball," "swimming," and "weightlifting" are called class terms.

Definition 3.2: Let A be an attribute, and C be the classification attribute of an instance, and let A_i be a linguistic term of A and C_k be a class term of C. The degree of correctness of the classification (Jeng & Liang, 1993), denoted by cc(A_i, C_k), is the ratio of the number of instances in the decision nodes of the decision trees which have "A is A_i and C is C_k" to the number of instances which have "A is A_k" in

Table 1. A fuzzy relation

Case	Outlook			Temperature			Humidity		Wind		Plan		
	sunny	cloudy	rain	hot	mild	cool	humid	normal	windy	not windy	volleyball	swimming	weightlifting
1	0.9	0.1	0.0	1.0	0.0	0.0	0.8	0.2	0.4	0.6	0.0	0.8	0.2
2	0.8	0.2	0.0	0.6	0.4	0.0	0.0	1.0	0.0	1.0	1.0	0.7	0.0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
15	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.8	0.2	0.0	0.0	1.0
16	1.0	0.0	0.0	0.5	0.5	0.0	0.0	1.0	0.0	1.0	0.8	0.6	0.0

Table 2. A translated fuzzy relation under the significance level $\alpha = 0.5$

Case	Outlook			Temperature			Humidity		Wind		Plan		
	sunny	cloudy	rain	hot	mild	cool	humid	normal	windy	not windy	volleyball	swimming	weightlifting
1	0.9	0.0	0.0	1.0	0.0	0.0	0.8	0.0	0.0	0.6	0.0	0.8	0.0
2	0.8	0.0	0.0	0.6	0.0	0.0	0.0	1.0	0.0	1.0	1.0	0.7	0.0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
15	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.8	0.0	0.0	0.0	1.0
16	1.0	0.0	0.0	0.5	0.5	0.0	0.0	1.0	0.0	1.0	0.8	0.6	0.0

Table 3. A reduced fuzzy relation under the significance level $\alpha = 0.5$

Case	Outlook	Temperature	Humidity	Windy	Plan
1	{sunny}	{hot}	{humid}	{not windy}	{swimming}
2	{sunny}	{hot}	{normal}	{not windy}	{volleyball, swimming}
⋮	⋮	⋮	⋮	⋮	⋮
15	{rain}	{cool}	{humid}	{windy}	{weightlifting}
16	{sunny}	{hot, mild}	{normal}	{not windy}	{volleyball, swimming}

Table 4. A reduced fuzzy relation

Case	Outlook	Temperature	Humidity	Windy	Plan
1	{sunny}	{hot}	{normal}	{not windy}	{volleyball}
2	{cloudy}	{mild}	{normal}	{not windy}	{volleyball}
3	{cloudy}	{cool}	{humid}	{not windy}	{weightlifting}
4	{cloudy}	{cool}	{normal}	{not windy}	{volleyball}
5	{cloudy}	{mild}	{normal}	{not windy}	{volleyball}
6	{sunny}	{hot, mild}	{normal}	{not windy}	{volleyball}

the decision node of the decision trees. The degree of correctness of the classification is a real value between zero and one.

Example 3.2: Based on Definition 3.2 one can calculate the degree of correctness of the classification for each linguistic term in Table 4 shown as follows:

cc(sunny; volleyball) = 2/2 = 1,
cc(sunny; weightlifting) = 0/2 = 0,
cc(cloudy; volleyball) = 3/4 = 0.75,
cc(cloudy; weightlifting) = 1/4 = 0.25,
cc(hot; volleyball) = 2/2 = 1,
cc(hot; weightlifting) = 0,
cc(cool; volleyball) = 1/2 = 0.5,
cc(cool; weightlifting) = 0.5,
cc(mild; volleyball) = 3/3 = 1,
cc(mild; weightlifting) = 0/3 = 0,
cc(humid; volleyball) = 0/1 = 0,
cc(humid; weightlifting) = 1/1 = 1,

cc(normal; volleyball) = 5/5 = 1,
cc(normal; weightlifting) = 0/5 = 0,
cc(not windy; volleyball) = 5/6 = 0.833,
cc(not windy; weightlifting) = 1/6 = 0.167.

Let C be a classification attribute with class term $C_1, C_2, \ldots, C_k$ and let A be an attribute having m linguistic terms $A_1, A_2, \ldots, A_m$. The entropy of A is calculated as follows (Sun, 1995)

$$(1/m)\left\{\sum_{i=1}^{m}\sum_{j=1}^{k}\left(-p_{ij}\ \log\ p_{ij} - n_{ij}\ \log\ n_{ij}\right)\right\}, \tag{3}$$

where

(1) $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, k$,
(2) $p_{ij} = cc(A_i; C_j)$,
(3) $n_{ij} = 1 - p_{ij}$.

Example 3.3: According to formula (3) and Table 4, one can calculate the entropy of the attributes "outlook," "temperature," "humidity," and "wind," respectively, shown as follows:

Entropy of "outlook" = 0.122,
Entropy of "temperature" = 0.100,
Entropy of "humidity" = 0,
Entropy of "wind" = 0.196. (4)

In the following, the concept of "disadvantageous linguistic terms" is presented, which will be used in the proposed method for fuzzy classification rules generation.

Definition 3.3: Let β be a disadvantageous threshold value determined by the user, where $\beta \in [0,1]$, A_i be a linguistic term of attribute A, and let C be the classification attribute and C_j be a class term of C. If $cc(A_i; C_j)$ is less than $1-\beta$, we say that A_i is a disadvantageous linguistic term to C_j.

Example 3.4: Assume that the disadvantageous threshold value given by the user is 0.8. Then, based on Example 3.2, we can see that

$$\begin{aligned} &\mathrm{cc(sunny;\ weightlifting)} < (1-0.8) = 0.2,\\ &\mathrm{cc(hot;\ weightlifting)} < (1-0.8) = 0.2,\\ &\mathrm{cc(mild;\ weightlifting)} < (1-0.8) = 0.2,\\ &\mathrm{cc(humid;\ volleyball)} < (1-0.8) = 0.2. \end{aligned} \tag{5}$$

Thus, one can see that {sunny, hot, mild} and {humid} are disadvantageous linguistic terms with respect to "weightlifting" and "volleyball," respectively.

To generate fuzzy classification rules for forecasting unknown values, one only considers linguistic terms and class terms of an attribute and the classification attribute, respectively, whose membership values are not less than the significance level α, $\alpha \in [0,1]$, where the value of α is determined by the user. The linguistic terms and the class terms whose membership values are not less than the significance level α will be kept in the instance of the training data set. For example, if the significance level α is 0.5, then we can reduce the training data set shown in Table 1 into Table 3.

In the following, a method for generating fuzzy classification rules is presented from constructed fuzzy decision trees using compound analysis techniques. Let α be a significance level, β be a disadvantageous threshold value, γ be a candidate threshold value, and θ be a criteria threshold value. The values of α, β, γ, and θ are given by the user, where $\alpha \in [0,1]$, $\beta \in [0,1]$, $\gamma \in [0,1]$, $\theta \in [0,1]$, and $\gamma \leq \theta$. The algorithm for constructing fuzzy decision trees is presented as follows.

Step 1: Let T contain the training instances which have been reduced by significance level α, where the linguistic terms and the class terms whose membership value are not less than significance level α will be kept in the instances of the training data set. Let T be the root node.

Step 2: In T, based on the disadvantageous threshold value β, find all possible disadvantageous linguistic terms with respect to each class term.

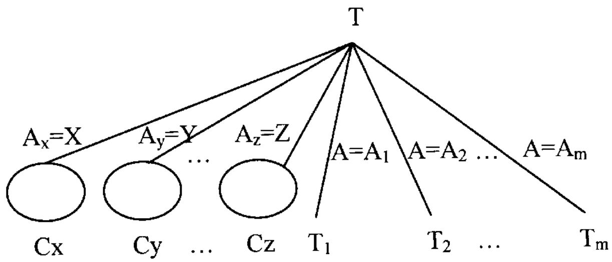

Figure 2. A sprouting tree.

Let D be an empty set. D will record any linguistic term having the degree of correctness of the classification not less than θ

Step 3: Let X, Y, ... , Z be linguistic terms, S be a set, and C_x, C_y, ... , C_z be class terms of the classification attribute C. Assume there exists X, Y, ... , Z in T such that cc(X; Cx)$\geq$cc(Y; Cy)$\geq$ ... $\geq$cc(Z; Cz)$\geq\gamma$ and S = {X, Y, ... , Z}.

Step 4: If S is empty, then go to Step 6. Otherwise, go to Step 5.

Step 5: Let A_K denote the attribute of linguistic term K. Select a linguistic term K from the set S in the order of X, Y, ... , Z and remove K from the set S.

If cc(K; C_k)$\geq\theta$, then create an edge labeled "A_k = K" from the root node T to a node labeled C_k as shown in Figure 2 and record any linguistic term d in C_k if d is a disadvantageous linguistic term to class term C. Put K into the set D. Remove class term C_k of the classification attribute C from the instances in T whose attribute A_k contains linguistic term K. Delete the instances in T whose classification attribute is empty.

If S is empty, then go to Step 2
else go to Step 5
else if S is empty go to Step 2.
else go to Step 5.

Step 6: If T is empty, then stop. Otherwise, let A be an attribute in node T having linguistic terms A_1, A_2, ... , A_m, where A has the minimum "entropy" using formula (3). Sprout the tree from the node T as shown

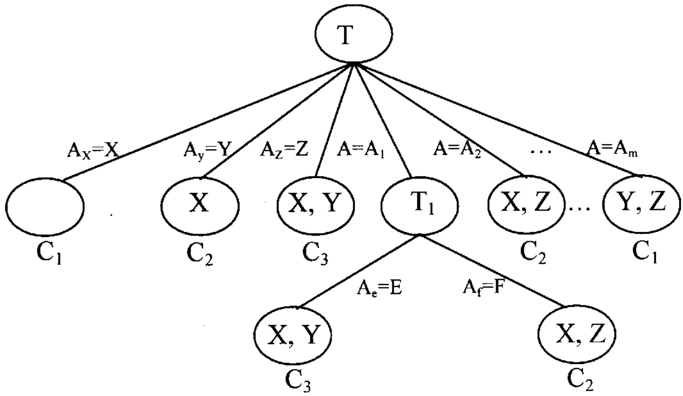

Figure 3. An example of the constructed fuzzy decision.

Table 5. Generated fuzzy classification rules

Rule 1: IF A_x is X THEN C is C_1
Rule 2: IF A_y is Y and A_x is $\overline{X}$ THEN C is C_2
⋮
Rule k: IF A_z is Z and A_x is $\overline{X}$ and A_y is $\overline{Y}$ THEN C is C_3
Rule k + 1: IF A is A_1 and A_e is E and A_x is $\overline{X}$ and A_y is $\overline{Y}$ THEN C is C_3
Rule k + 2: IF A is A_1 and A_f is F and A_x is $\overline{X}$ and A_z is $\overline{Z}$ THEN C is C_2
Rule k + 3: IF A is A_2 and A_x is $\overline{X}$ and A_z is $\overline{Z}$ THEN C is C_2
⋮
Rule n: IF A is A_m and A_y is $\overline{Y}$ and A_z is $\overline{Z}$ THEN C is C_1

in Figure 2 and let each T_i be a new T, where node T_i contains the instances whose attribute A contains the linguistic term A_i and $1 \leq i \leq m$. Go to Step 2.

Figure 3 shows an example of a fuzzy decision tree, where we assume that X is a disadvantageous linguistic term to C_2 denoted by $\underset{C_2}{\otimes}$, X and Y are disadvantageous linguistic terms to C_3 denoted by $\underset{C_3}{(X,Y)}$, X and Z are disadvantageous linguistic terms to C_2 denoted by $\underset{C_2}{(X,Z)}$ and Y and Z are disadvantageous linguistic terms to C_1 denoted by $\underset{C_1}{(Y,Z)}$. From the constructed fuzzy decision tree shown in Figure 3, one can generate the fuzzy classification rules shown in Table 5 from the root node to the leaf nodes of the fuzzy decision tree, where X, Y, Z, E, and F are linguistic terms of the attributes A_x, A_y, A_z, A_e, and A_f, respectively,

Table 6. A small training data set (Yuan et al., 1995)

	Outlook			Temperature			Humidity		Wind		Plan		
Case	sunny	cloudy	rain	hot	mild	cool	humid	normal	windy	not windy	volleyball	swimming	weightlifting
1	0.9	0.1	0.0	1.0	0.0	0.0	0.8	0.2	0.4	0.6	0.0	0.8	0.2
2	0.8	0.2	0.0	0.6	0.4	0.0	0.0	1.0	0.0	1.0	1.0	0.7	0.0
3	0.0	0.7	0.3	0.8	0.2	0.0	0.1	0.9	0.2	0.8	0.3	0.6	0.1
4	0.2	0.7	0.1	0.3	0.7	0.0	0.2	0.8	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	0.0	0.0	1.0
6	0.0	0.7	0.3	0.0	0.3	0.7	0.7	0.3	0.4	0.6	0.2	0.0	0.8
7	0.0	0.3	0.7	0.0	0.0	1.0	0.0	1.0	0.1	0.9	0.0	0.0	1.0
8	0.0	1.0	0.0	0.0	0.2	0.8	0.2	0.8	0.0	1.0	0.7	0.0	0.3
9	1.0	0.0	0.0	1.0	0.0	0.0	0.6	0.4	0.7	0.3	0.2	0.8	0.0
10	0.9	0.1	0.0	0.0	0.3	0.7	0.0	1.0	0.9	0.1	0.0	0.3	0.7
11	0.7	0.3	0.0	1.0	0.0	0.0	1.0	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	0.3	0.7	0.3	0.7	0.7	0.2	0.1
13	0.9	0.1	0.0	0.2	0.8	0.0	0.1	0.9	1.0	0.0	0.0	0.0	1.0
14	0.0	0.9	0.1	0.0	0.9	0.1	0.1	0.9	0.7	0.3	0.0	0.0	1.0
15	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.8	0.2	0.0	0.0	1.0
16	1.0	0.0	0.0	0.5	0.5	0.0	0.0	1.0	0.0	1.0	0.8	0.6	0.0

Table 7. Reduced training data set under the significance level $\alpha = 0.5$

Case	Outlook	Temperature	Humidity	Wind	Plan
1	{sunny}	{hot}	{humid}	{not windy}	{swimming}
2	{sunny}	{hot}	{normal}	{not windy}	{volleyball, swimming}
3	{cloudy}	{hot}	{normal}	{not windy}	{swimming}
4	{cloudy}	{mild}	{normal}	{not windy}	{volleyball}
5	{rain}	{hot}	{humid, normal}	{windy, not windy}	{weightlifting}
6	{cloudy}	{cool}	{humid}	{not windy}	{weightlifting}
7	{rain}	{cool}	{normal}	{not windy}	{weightlifting}
8	{cloudy}	{cool}	{normal}	{not windy}	{volleyball}
9	{sunny}	{hot}	{humid}	{windy}	{swimming}
10	{sunny}	{cool}	{normal}	{windy}	{weightlifting}
11	{sunny}	{hot}	{humid}	{not windy}	{swimming}
12	{cloudy}	{mild}	{normal}	{not windy}	{volleyball}
13	{sunny}	{mild}	{normal}	{windy}	{weightlifting}
14	{cloudy}	{mild}	{normal}	{windy}	{weightlifting}
15	{rain}	{cool}	{humid}	{windy}	{weightlifting}
16	{sunny}	{hot, mild}	{normal}	{not windy}	{volleyball, swimming}

and C_1, C_2, and C_3 are class terms of the classification attribute C, where "$\overline{X}$" means the complement of X (i.e., not X).

Example 3.5: Assume that the significance level α given by the user is 0.5, using the training data set presented in Yuan and Shaw (1995) as shown in Table 6. By Step 1 of the proposed algorithm, one can reduce Table 6 into Table 7 under the significance level $\alpha = 0.5$, where "plan" is the classification attribute and "volleyball," "swimming," and "weightlifting" are the class terms of the classification attribute "plan."

Assume that the disadvantageous threshold value $\beta = 0.8$, the candidate threshold value $\gamma = 0.8$, and the criteria threshold value $\theta = 1.0$; then, initially, the root node contains the instances of the reduced training data set under the significance level $\alpha = 0.5$ as shown in Table 7.

(i) First iteration: Based on Step 2 of the proposed algorithm, we can see that

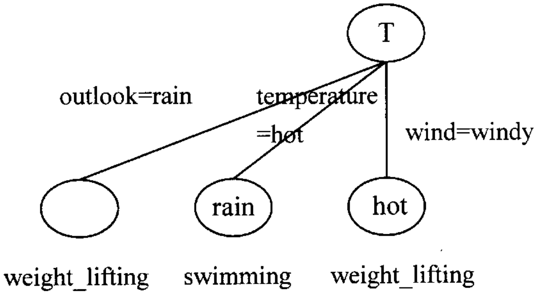

Figure 4. Partially constructed fuzzy decision tree.

{rain, humid, windy} are disadvantageous linguistic terms to the class term "volleyball";
{rain, cool} are disadvantageous linguistic terms to the class term "swimming";
{hot} is a disadvantageous linguistic term to the class term "weightlifting".

Based on Step 3 of the proposed algorithm, one can get the set S of linguistic terms to produce classifications under the candidate threshold value $\gamma = 0.8$, where

S = {rain, hot, windy, cool}.

Since "rain," "hot," and "windy" can produce classifications (i.e., they have the degrees of correctness of the classification not less than the criteria threshold value θ, where $\theta = 1.0$) after applying Step 4, and Step 5 of the proposed algorithm one can partially construct the fuzzy decision tree shown in Figure 4, where the linguistic terms printed in the leaf nodes of the partially constructed fuzzy decision tree are the disadvantageous linguistic terms to the corresponding class term, i.e., "rain" is a disadvantageous linguistic term with respect to the class term "swimming"; "hot" is a disadvantageous linguistic term with respect to the class term "weightlifting."

(ii) Second iteration: After applying the second iteration of the proposed algorithm (i.e., Step 2, Step 3, Step 4, Step 5, and Step 6), we have S = {humid, normal, sunny, hot, mild, not windy}. In this case,

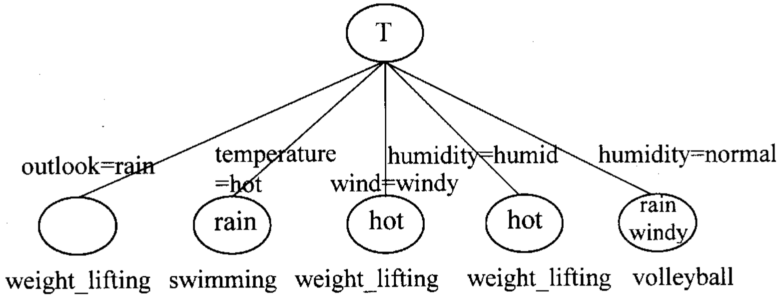

Figure 5. A constructed fuzzy decision tree.

"humid" and "normal" produce classifications (i.e., they have the degree of correctness of the classification not less than the criteria threshold value θ, where $\theta = 1.0$). The final constructed fuzzy decision tree is shown in Figure 5, where the linguistic terms printed in the leaf nodes of the constructed fuzzy decision tree are the disadvantageous linguistic terms to the corresponding class term.

Based on the constructed fuzzy decision tree shown in Figure 5, we can get the fuzzy classification rules shown as follows:

Rule 1: **IF** outlook is *rain* **THEN** plan is *weightlifting*.
Rule 2: **IF** temperature is *hot* **AND** outlook is $\overline{rain}$ **THEN** plan is *swimming*.
Rule 3: **IF** wind is *windy* **AND** temperature is $\overline{hot}$ **THEN** plan is *weightlifting*.
Rule 4: **IF** humidity is *humid* **AND** temperature is $\overline{hot}$ **THEN** plan is *weightlifting*.
Rule 5: **IF** humidity is *normal* **AND** outlook is $\overline{rain}$ **AND** wind is $\overline{windy}$ **THEN** plan is volleyball.

where $\overline{rain}$ is the complement of *rain* (i.e., NOT *Rain*), $\overline{windy}$ is the complement of *windy* (i.e., NOT *windy*), and $\overline{hot}$ is the complement of *hot* (i.e., NOT *hot*), respectively.

One can apply the generated fuzzy classification rules to deal with the classification problems. A fuzzy classification rule has the following form:

IF A is A_i AND B is B_j THEN C is C_k,

where A and B are attributes of an instance, A_i and B_j are linguistic terms of A and B, respectively; C_k is a class term of the classification attribute C. Then, one can forecast the value of the term C_k of the classification attribute C. Let a and b be the membership values of the instance in A_i and B_j, respectively. Let c be the forecasted value of C_k, where $c = \min(a, b)$. When two or more fuzzy classification rules have the same conclusion (i.e., they all conclude that "IF conditions THEN C is C_k"), then one generates several forecasted values with respect to C_k, and one takes the largest one.

Based on the generated fuzzy classification rules described above, one can use them to get the forecasted membership values of each class term of the classification attribute. For example, by applying the above five fuzzy classification rules to case 1 shown in Table 6, we can get the forecasted membership values for case 1 with respect to the class terms "volleyball," "swimming," and "weightlifting," respectively, shown as follows:

(1) Based on Rule 5 and Table 6, we can see that

  (i) The degree of membership of "humidity is normal" for case 1 is 0.2.
  (ii) The degree of membership of "outlook is rain" for case 1 is 0.0. Thus, the degree of membership of "outlook is $\overline{\text{rain}}$" for case 1 is equal to $1.0 - 0.0 = 1.0$.
  (iii) The degree of membership of "wind is windy" for case 1 is 0.4. Thus, the degree of membership of "outlook is $\overline{\text{windy}}$" for case 1 is equal to $1.0 - 0.4 = 0.6$.

  Thus, the forecasted membership value for case 1 with respect to the class term "volleyball" can be evaluated as follows:

$$\begin{aligned} &\min(0.2,\ 1.0-0.0,\ 1.0-0.4) \\ &= \min\ (0.2,\ 1.0,\ 0.6) \\ &= 0.2. \end{aligned}$$

(2) Based on Rule 2 and Table 6, one can see that

(i) The degree of membership of "temperature is hot" for case 1 is 1.0.

(ii) The degree of membership of "outlook is rain" for case 1 is 0.0. Thus, the degree of membership of "outlook is $\overline{\text{rain}}$" for case 1 is equal to 1.0−0.0 = 1.0.

Thus, the forecasted membership value for case 1 with respect to the class term "swimming" can be evaluated as follows:

min(1.0, 1.0−0.0)
= min (1.0, 1.0)
= 1.0.

(3) (a) Based on Rule 1 and Table 6, one can see that the degree of membership of "outlook is rain" for case 1 is 0.0.

(b) Based on Rule 3 and Table 6, we can see that

(i) The degree of membership of "wind is windy" for case 1 is 0.4.

(ii) The degree of membership of "temperature is hot" of case 1 is 1.0. Thus, the degree of membership of "temperature is $\overline{hot}$" for case 1 is equal to 1.0−1.0 = 0.0.

(c) Based on Rule 4 and Table 6, one can see that

(i) The degree of membership of "humidity is humid" for case 1 is 0.8.

(ii) The degree of membership of "temperature is hot" for case 1 is 1.0. Thus, the degree of membership of "temperature is $\overline{hot}$" for case 1 is equal to 1.0−1.0 = 0.0.

From (a), (b), and (c), the forecasted membership value for case 1 with respect to the class term "weightlifting" can be evaluated as follows:

max[0.0, min (0.4, 1.0−1.0), min(0.8, 1.0−1.0)]
= max [0.0, min(0.4, 0.0), min (0.8, 0.0)]
= max [0.0, 0.0, 0.0]
= 0.0.

Similarly, one can get the forecasted membership values of "volleyball," "swimming," and "weightlifting" through Case 2 to Case 16 shown as follows:

Case 2: volleyball = 1.0, swimming = 0.6, weightlifting = 0.0,
Case 3: volleyball = 0.7, swimming = 0.7, weightlifting = 0.3,
Case 4: volleyball = 0.7, swimming = 0.3, weightlifting = 0.3,
Case 5: volleyball = 0.1, swimming = 0.1, weightlifting = 0.9,
Case 6: volleyball = 0.3, swimming = 0.0, weightlifting = 0.7,
Case 7: volleyball = 0.3, swimming = 0.0, weightlifting = 0.7,
Case 8: volleyball = 0.8, swimming = 0.0, weightlifting = 0.2,
Case 9: volleyball = 0.3, swimming = 1.0, weightlifting = 0.0,
Case 10: volleyball = 0.1, swimming = 0.0, weightlifting = 0.9,
Case 11: volleyball = 0.0, swimming = 1.0, weightlifting = 0.0,
Case 12: volleyball = 0.7, swimming = 0.0, weightlifting = 0.3,
Case 13: volleyball = 0.0, swimming = 0.2, weightlifting = 0.8,
Case 14: volleyball = 0.3, swimming = 0.0, weightlifting = 0.7,
Case 15: volleyball = 0.0, swimming = 0.0, weightlifting = 1.0,
Case 16: volleyball = 1.0, swimming = 0.5, weightlifting = 0.0.

In the next section, experimental results of the proposed method will be compared with that of Yuan and Shaw's method (1995) by assigning the parameters of the significance level α, the disadvantageous threshold value β, the candidate threshold value γ, and the criteria threshold value θ to construct fuzzy decision trees and generate fuzzy classification rules, where $\alpha \in [0,1]$, $\beta \in [0,1]$, $\gamma \in [0,1]$, $\theta \in [0,1]$, and $\gamma \leq \theta$. The Turbo C version 3.0 has been used on a PC/AT to implement the proposed algorithm for constructing fuzzy classification trees and generating fuzzy classification rules. By using the generated fuzzy classification rules, one can forecast the membership values of the class terms of the classification attributes.

EXPERIMENT ANALYSIS

In the following, it will be assumed that the significance level $\alpha = 0.5$, the disadvantageous threshold value $\beta = 0.8$, the candidate threshold value $\gamma = 0.8$, and the criteria threshold value $\theta = 1.0$. Comparative results between the proposed method and Yuan and Shaw's method (1995) concerning the number of edges in the constructed fuzzy decision tree, the number of decision nodes involving the computations of the entropy of the attributes, and the classification accuracy rate will also be discussed.

Yuan and Shaw (1995) used the data shown in Table 6 to generate six fuzzy classification rules shown below under the significance level = 0.5 and the truth level = 0.7. It is obvious that Yuan and Shaw's method generated more fuzzy classification rules than the proposed method. One can see that the proposed method only generated five fuzzy decision rules, while Yuan and Shaw's method generated six fuzzy classification rules. In the following, the classification results of Yuan and Shaw's method (1995) and the classification results of the proposed method are compared as shown in Table 8 and Table 9. As mentioned in Yuan and Shaw (1995), the classification accuracy rate is the ratio of the number of instances (cases) in the training data set, which are correctly classified to the number of total instances in the training data set. As described above, one can see that the classification accuracy rate of the proposed method is 0.9375, while the classification accuracy rate of Yuan and Shaw's method is 0.8125.

The comparison between Yuan and Shaw's method and the proposed method concerning the number of edges in the constructed fuzzy decision tree, the number of decision nodes in the fuzzy decision trees that involve the computations of the entropy of the attribute, and the classification accuracy rate are shown in Table 10.

From Table 10, one can see that Yuan and Shaw's method has the classification accuracy rate equal to 0.8125 (under the parameters that the significance level is 0.5 and the truth level threshold is 0.7), while the proposed method has the classification accuracy rate equal to 0.9375 (under the significance level $\alpha = 0.5$, the disadvantageous threshold value $\beta = 0.8$, the candidate threshold value $\gamma = 0.8$, and the criteria threshold value $\theta = 1.0$). Furthermore, one also can see that the proposed method generates less edges and decision nodes in the constructed fuzzy decision tree than the one presented in (Yuan and Shaw, 1995).

CONCLUSIONS

In this article, a new method has been presented for constructing fuzzy decision tree and generating fuzzy classification rules from the constructed fuzzy decision trees, using the compound analysis techniques. An experiment has also been made to compare the proposed method with Yuan and Shaw's method (1995). From the experimental results, one can see that the proposed method has the following advantages:

Table 8. Compare Yuan and Shaw's classification results with known classification results

Classification Known in Training Data (Yuan et al., 1995) (see Table 6)

Case	Plan: Volleyball	Plan: Swimming	Plan: Weightlifting
1	0.0	0.8	0.2
2	1.0	0.7	0.0
3	0.3	0.6	0.1
4	0.9	0.1	0.0
5	0.0	0.0	1.0
6	0.2	0.0	0.8
7	0.0	0.0	1.0
8	0.7	0.0	0.3
9	0.2	0.8	0.0
10	0.0	0.3	0.7
11	0.4	0.7	0.0
12	0.7	0.2	0.1
13	0.0	0.0	1.0
14	0.0	0.0	1.0
15	0.0	0.0	1.0
16	0.8	0.6	0.0

Classification Results with Learned Rules by Yuan and Shaw's Method (Yuan et al., 1995)

Case	Plan: Volleyball	Plan: Swimming	Plan: Weightlifting
1	0.0	0.9	0.0
2	0.4	0.6	0.0 (#)
3	0.2	0.7	0.3
4	0.7	0.3	0.3
5	0.3	0.1	0.9
6	0.3	0.0	0.7
7	0.0	0.0	1.0
8	0.2	0.0	0.8 (#)
9	0.0	1.0	0.0
10	0.1	0.0	0.7
11	0.0	0.7	0.0
12	0.7	0.0	0.3
13	0.0	0.2	0.8
14	0.3	0.0	0.7
15	0.0	0.0	1.0
16	0.5	0.5	0.0 (*)

(#) Wrong classification; (*) Cannot distinguish between two or more classes.

Table 9. Compare our classification results with known classification results

	Classification Known in Training Data (Yuan et al., 1995) (see Table 6)		
	Plan		
Case	Volleyball	Swimming	Weightlifting
1	0.0	0.8	0.2
2	1.0	0.7	0.0
3	0.3	0.6	0.1
4	0.9	0.1	0.0
5	0.0	0.0	1.0
6	0.2	0.0	0.8
7	0.0	0.0	1.0
8	0.7	0.0	0.3
9	0.2	0.8	0.0
10	0.0	0.3	0.7
11	0.4	0.7	0.0
12	0.7	0.2	0.1
13	0.0	0.0	1.0
14	0.0	0.0	1.0
15	0.0	0.0	1.0
16	0.8	0.6	0.0
	Classification Results with Learned Rules by the Proposed Method		
	Plan		
Case	Volleyball	Swimming	Weightlifting
1	0.2	1.0	0.0
2	1.0	0.6	0.0
3	0.7	0.7	0.3 (*)
4	0.7	0.3	0.3
5	0.1	0.1	0.9
6	0.3	0.0	0.7
7	0.3	0.0	0.7
8	0.8	0.0	0.2
9	0.3	1.0	0.0
10	0.1	0.0	0.9
11	0.0	1.0	0.0
12	0.7	0.0	0.3
13	0.0	0.2	0.8
14	0.3	0.0	0.7
15	0.0	0.0	1.0
16	1.0	0.5	0.0

(*) Cannot distinguish between two or more classes.

Table 10. Comparison between Yuan and Shaw's method and the proposed method

	Yuan and Shaw's Method (Yuan et al., 1995)	The Proposed Method
Number of branch edges	8	5
Number of decision nodes	3	2
Classification accuracy rate	0.8125	0.9375

(1) The proposed method could get a better classification accuracy rate than the one presented in Yuan and Shaw (1995). From Table 10, one can see that the classification accuracy rate of the proposed method is 0.9375, and the classification rate of Yuan and Shaw's method is 0.8125. Furthermore, the proposed method generates less branch edges and decision nodes than the one presented in Yuan and Shaw (1995).

(2) The proposed method generated less fuzzy classification rules. From the illustrated example, one can see that the proposed method generated five fuzzy classification rules but Yuan and Shaw's method generated six fuzzy classification rules.

REFERENCES

Boyen, X., and L. Wehenkel. 1999. Automatic induction of fuzzy decision trees and its application to power system security assessment. *Fuzzy Sets Syst.* 102(1):3–19.

Chang, R. L. P., and T. Pavlidis. 1977. Fuzzy decision tree algorithms. *IEEE Trans. Syst., Man, Cybern* 7(1):28–35.

Chen, S. M. 1988. A new approach to handling fuzzy decision making problems. *IEEE Trans. Syst., Man, Cybern.* 18(6):1012–1016.

Chen, S. M., and M. S. Yeh. 1997. Generating fuzzy rules from relational database systems for estimating null values. *Cybernetics and Systems: An International Journal* 28(8):695–723.

Hart, A. 1995. The role of induction in knowledge elicitation, *Expert Systems* 2(1):24–28.

Hunt, E. B., J. Marin, and P. J. Stone. 1966. *Experience in induction*. New York: Academic Press.

Jeng, B., and T. P. Liang. 1993. Fuzzy indexing and retrieval in case-based systems. In *Proc. 1993 Pan Pacific Conference on Information Systems*, Taiwan, R. O. C., 258–266.

Lin, S. Y., and S. M. Chen. 1996. A new method for generating fuzzy rules from fuzzy decision trees. In *Proc. 7th International Conference on Information Management*, Chungli, Taoyuan, Taiwan, R. O. C., 358–364.

Mingers, J. 1989. An empirical comparison of pruning methods for decision tree induction. *Machine Learning* 4(2):227–243.

Quinlan, J. R. 1979. Discovering rules by induction from large collection of examples. In *Expert systems in the micro electronic age*, D. Michie, ed. Edinburgh: Edinburgh University Press.

Quinlan, J. R. 1986. Introduction of decision trees. *Machine Learning* 1(1):81–106.

Quinlan, J. R. 1990. Decision trees and decisionmaking. *IEEE Trans. Syst., Man, Cybern.* 20(2):339–346.

Safavian, S. R., and D. Landgrebe. 1991. A survey of decision tree classifier methodology. *IEEE Trans. Syst., Man, Cybern.* 21(3):660–674.

Sun, S. W. 1995. *A fuzzy approach to decision trees*. M.S. Thesis, Institute of Computer and Information Science, National Chiao Tung University, Hsinchu, Taiwan, R. O. C.

Wu, T. P., and S. M. Chen. 1999. A new method for constructing membership functions and fuzzy rules from training examples. *IEEE Trans. Syst., Man, Cybern.—Part B* 29(1):25–40.

Yasdi, R. 1991. Learning classification rules from database in the context of knowledge acquisition and representation. *IEEE Trans. Syst., Man, Cybern.* 3(3):293–306.

Yeh, M. S., and S. M. Chen. 1995. An algorithm for generating fuzzy rules from relational database systems. In *Proc. 6th International Conference on Information Management*, Taipei, Taiwan, R. O. C., 219–226.

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

Zadeh, L. A. 1965. Fuzzy sets. *Inf. Contr.* 8:338–353.

Zadeh, L. A. 1975. The concepts of a linguistic variable and its application to approximate reasoning (I). *Information Sciences* 8:199–249.