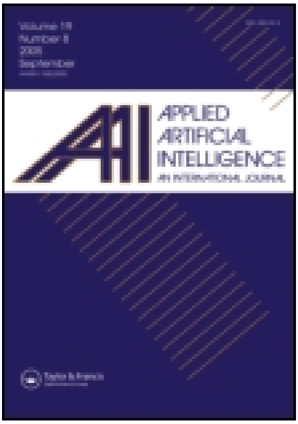

Applied Artificial Intelligence: An International Journal
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uaai20

A new method for generating fuzzy rules from numerical data for handling classification problems

Shyi-Ming Chen , Shao-Hua Lee & Chia-Hoang Lee

Published online: 30 Nov 2010.

Applied Artificial Intelligence, 15:645–664, 2001

0883-9514/01 $12.00 +.00

A NEW METHOD FOR GENERATING FUZZY RULES FROM NUMERICAL DATA FOR HANDLING CLASSIFICATION PROBLEMS

SHYI-MING CHEN
Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, Republic of China

SHAO-HUA LEE and CHIA-HOANG LEE
Department of Computer and Information Science, National Chiao Tung University, Hsinchu, Taiwan, Republic of China

Fuzzy classification is one of the important applications of fuzzy logic. Fuzzy classification Systems are capable of handling perceptual uncertainties, such as the vagueness and ambiguity involved in classification problems. The most important task to accomplish a fuzzy classification system is to find a set of fuzzy rules suitable for a specific classification problem. In this article, we present a new method for generating fuzzy rules from numerical data for handling fuzzy classification problems based on the fuzzy subsethood values between decisions to be made and terms of attributes by using the level threshold value α *and the applicability threshold value* β*, where* $\alpha \in [0, 1]$ *and* $\beta \in [0, 1]$*. We apply the proposed method to deal with the "Saturday Morning Problem," where the proposed method has a higher classification accuracy rate and generates fewer fuzzy rules than the existing methods.*

Fuzzy classification is one of the important applications of fuzzy logic (Zadeh, 1965, 1988). In a fuzzy classification system (Yoshinari, Pedrycz, & Hirota, 1993), a case can properly be classified by applying a set of fuzzy rules based on the linguistic terms (Zadeh, 1975) of its attributes. Fuzzy classification systems are capable of handling perceptual uncertainties,

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

Address correspondence to Professor Shyi-Ming Chen, Ph.D., Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, R.O.C. E-mail: smchen@et.ntust.edu.tw

such as the vagueness and ambiguity involved in classification problems (Yuan & Shaw, 1995). The most important task to accomplish a fuzzy classification system is to find a set of fuzzy rules suitable for a specific classification problem. Usually, we have two methods to complete this task. One approach is to obtain knowledge from experts and translate their knowledge directly into fuzzy rules. However, the process of knowledge acquisition and validation is difficult and time-consuming. It is very likely that an expert may not be able to express his or her knowledge explicitly and accurately. Another approach is to generate fuzzy rules through a machine-learning process (Castro & Zurita, 1997; Chen & Yeh, 1998; Chen, Lee, & Lee, 1999; Hayashi & Imura, 1990; Hong & Lee, 1996; Klawonn & Kruse, 1997; Nozaki, Ishibuchi, & Tanaka, 1997; Wang & Mendel, 1992; Wu & Chen, 1999, Yuan & Shaw, 1995; Yuan & Zhuang, 1996), in which knowledge can be automatically extracted or induced from sample cases or examples. In Castro and Zurita (1997) an inductive learning algorithm in fuzzy systems is presented. In Chen and Yeh (1998) we have presented a method for generating fuzzy rules from relational database systems for estimating null values. In Hayashi and Imura (1990) a method to automatically extract fuzzy if-then rules from a trained neural network is presented. In Hong and Lee (1996) an algorithm to induce fuzzy rules and membership functions from training examples is presented. In Klawonn and Kruse (1997) a method for constructing a fuzzy controller from data is presented. In Nozaki, Ishibuchi, and Tanaka (1997) a heuristic method for generating fuzzy rules from numerical data is presented. In Wang and Mendel (1992) an algorithm for generating fuzzy rules by learning from examples is presented. In Wu and Chen (1999), we have presented a method for constructing membership functions and fuzzy rules from training examples. In Yuan and Zhuang (1996) a genetic algorithm for generating fuzzy classification rules from training examples is presented.

A commonly used machine-learning method is the induction of decision trees (Quinlan, 1994) for a specific problem. The method of decision trees induction has been expanded to induce fuzzy decision trees proposed by Yuan and Shaw (1995), where fuzzy entropy is used to lead the search of the most effective decision nodes. However, the method presented in Yuan and Shaw (1995) has some drawbacks, i.e., (1) it generates too many fuzzy rules and (2) its classification accuracy rate is not good enough.

In this article, we present a new method based on the filtering of the fuzzy subsethood values (Kosko, 1986; Yuan & Shaw, 1995) between decisions to be made and terms of attributes by the level threshold value α and the applicability threshold value β for generating fuzzy rules from the numerical data in a more efficient manner, where $\alpha \in [0, 1]$ and $\beta \in [0, 1]$. We apply the proposed method to deal with the Saturday Morning Problem (Yuan & Shaw, 1995), where the proposed method has higher classification accuracy and generates fewer fuzzy rules than the one presented in (Yuan & Shaw, 1995).

FUZZY SET THEORY

In 1965, Zadeh proposed the theory of fuzzy sets (1965). Let U be a universe of discourse, where $U = \{u_1, u_2, \ldots, u_n\}$. A fuzzy set A of the universe of discourse U can be represented by

$$A = \sum_{i=1}^{n} \mu_A(u_i)/u_i$$
$$= \mu_A(u_1)/u_1 + \mu_A(u_2)/u_2 + \cdots + \mu_A(u_n)/u_n, \quad (1)$$

where μ_A is the membership function of the fuzzy set A, $\mu_A(u_i)$ indicates the degree of membership of u_i in the fuzzy set A, $\mu_A(u_i) \in [0, 1]$, the symbol "+ " means the union operator, the symbol "/" represents the separator, and $1 \leq i \leq n$.

Definition 2.1. Let A and B be two fuzzy sets of the universe of discourse U with membership functions μ_{A} and μ_{B}, respectively. The union of the fuzzy sets A and B is defined by

$$\mu_{A \cup B}(u) = \max\{u_A(u),\ \mu_B(u)\}, \quad \forall u \in U. \quad (2)$$

The intersection of A and B, denoted as $A \cap B$, is defined by

$$\mu_{A \cap B}(u) = \min\{\mu_A(u),\ \mu_B(u)\}, \quad \forall u \in U. \quad (3)$$

The complement of A, denoted as $\bar{A}$, is defined by

$$\mu_{\bar{A}}(u) = 1 - \mu_A(u), \quad \forall u \in U. \quad (4)$$

Definition 2.2. Let A and B be two fuzzy sets defined on the universe of discourse U with membership functions μ_{A} and μ_{B}, respectively. The fuzzy subsethood $\mathrm{S(A, B)}$ (Kosko, 1986; Yuan & Shaw, 1995) measures the degree in which A is a subset of B:

$$S(A, B) = \frac{M(A \cap B)}{M(A)} = \frac{\sum_{u \in U} \mathrm{Min}(\mu_A(u), \mu_B(u))}{\sum_{u \in U} \mu_A(u)}, \quad (5)$$

where $S(A, B) \in [0, 1]$.

A REVIEW OF YUAN AND SHAW'S METHOD FOR FUZZY RULES GENERATION

In the following, we briefly review Yuan and Shaw's method for fuzzy rules generation (1995). In a fuzzy classification problem, a collection of cases $U = \{u\}$ is represented by a set of attributes $A = \{A_1, \ldots, A_k\}$, where U is called the object space (Yuan & Shaw, 1995). Each attribute A_k depicts some important feature of a case and is usually limited to a small set of discrete linguistic terms $T(A_k) = \{T_1^k, \ldots, T_{s_k}^k\}$. In other words, $T(A_k)$ is the domain of the attribute A_k. Each case u in U is classified into a class C_i, where C_i is a member of classes C and $C = \{C_1, \ldots, C_L\}$. In our discussions, both cases and classes are fuzzy. The class C_i of C, $i = 1, \ldots, L$, is a fuzzy set defined on the universe of cases U. The membership function $\mu_{c_i}(u)$ assigns a degree to which u belongs to class C_i. The attribute A_k is a linguistic variable that takes linguistic values from $T(A_k) = \{T_1^k, \ldots, T_{s_k}^k\}$. The linguistic values T_j^k are also fuzzy sets defined on U. The membership value $\mu_{T_j^k}(u)$ depicts the degree to which case u's attribute A_k is T_j^k. A fuzzy classification rule (or abbreviated into fuzzy rule) can be written in the form

$$IF\ (\mathrm{A}_1 \text{ is } \mathrm{T}_{i_1}^1)\ AND \ldots AND\ (\mathrm{A}_\mathrm{k} \text{ is } \mathrm{T}_{i_k}^k)\ THEN\ (\mathrm{C} \text{ is } \mathrm{C}_\mathrm{j}). \tag{6}$$

Using a machine-learning method from a training set of cases whose class is known can induce a set of classification rules. An example of a small training data set of the Saturday Morning Problem (Yuan & Shaw, 1995) with fuzzy membership values is shown in Table 1. In the Saturday Morning Problem, a case is a Saturday morning's weather which can have four attributes:

Attribute = {Outlook, Temperature, Humidity, Wind},

and each attribute has linguistic values

Outlook = {Sunny, Cloudy, Rain},
Temperature = {Hot, Mild, Cool},
Humidity = {Humid, Normal},
Wind = {Windy, Not windy}.

The classification result (i.e., Plan) is the sport to be taken on that weekend day,

Plan = {Volleyball, Swimming, Weightlifting}.

The fuzzy decision tree induction method presented in Yuan and Shaw (1995) consists of the following steps:

TABLE 1 A Small Data Set for the Saturday Morning Problem (Yuan & Shaw, 1995)

	Outlook			Temperature			Humidity		Wind		Plan		
Case	Sunny	Cloudy	Rain	Hot	Mild	Cool	Humid	Normal	Windy	Not-windy	Volleyball	Swimming	W-lifting
1	0.9	0.1	0	1	0	0	0.8	0.2	0.4	0.6	0	0.8	0.2
2	0.8	0.2	0	0.6	0.4	0	0	1	0	1	1	0.7	0
3	0	0.7	0.3	0.8	0.2	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	0.2	0.8	0.3	0.7	0.9	0.1	0
5	0	0.1	0.9	0.7	0.3	0	0.5	0.5	0.5	0.5	0	0	1
6	0	0.7	0.3	0	0.3	0.7	0.7	0.3	0.4	0.6	0.2	0	0.8
7	0	0.3	0.7	0	0	1	0	1	0.1	0.9	0	0	1
8	0	1	0	0	0.2	0.8	0.2	0.8	0	1	0.7	0	0.3
9	1	0	0	1	0	0	0.6	0.4	0.7	0.3	0.2	0.8	0
10	0.9	0.1	0	0	0.3	0.7	0	1	0.9	0.1	0	0.3	0.7
11	0.7	0.3	0	1	0	0	1	0	0.2	0.8	0.4	0.7	0
12	0.2	0.6	0.2	0	1	0	0.3	0.7	0.3	0.7	0.7	0.2	0.1
13	0.9	0.1	0	0.2	0.8	0	0.1	0.9	1	0	0	0	1
14	0	0.9	0.1	0	0.9	0.1	0.1	0.9	0.7	0.3	0	0	1
15	0	0	1	0	0	1	1	0	0.8	0.2	0	0	1
16	1	0	0	0.5	0.5	0	0	1	0	1	0.8	0.6	0

1. fuzzification of the training data,
2. induction of a fuzzy decision tree,
3. conversion of the decision tree into a set of rules,
4. application of the fuzzy rules for classification.

Using the data shown in Table 1, the generated fuzzy decision tree is shown in Figure 1. From the fuzzy decision tree shown in Figure 1, we can enumerate the number of routes from root to leaf. Each route can be converted into a rule, where the condition part represents the attributes on the passing branches from the root to the leaf and the conclusion part represents

A. Fuzzy decision tree

Temperature? (G (Temperature) = 0.48)

Hot (G (Hot) = 0.45): Outlook? (G (Outlook | Hot) = 0.42)

Sunny: Swimming (S = 0.85)

Cloudy: Swimming (S = 0.72)

Rain: Weight-lifting (S = 0.73)

Mild (G (Mild) = 0.83): Wind? (G (Wind | Mild) = 0.36)

Windy: Weight-lifting (S = 0.81)

Not-windy: Volleyball (S = 0.78)

Cool(G (Cool) = 0.20): Weight-lifting (S = 0.88)

Note: G is the classification ambiguity measure at the decision node.

S is the classification truth level at the leaf.

B. Fuzzy rules converted from the fuzzy decision tree

Rule 1: ***IF*** Temperature is Hot ***AND*** Outlook is Sunny ***THEN*** Swimming (S = 0.85)

Rule 2: ***IF*** Temperature is Hot ***AND*** Outlook is Cloudy ***THEN*** Swimming (S = 0.72)

Rule 3: ***IF*** Temperature is Hot ***AND*** Outlook is Rain ***THEN*** Weight-lifting (S = 0.73)

Rule 4: ***IF*** Temperature is Mild ***AND*** Wind is Windy ***THEN*** Weight-lifting (S = 0.81)

Rule 5: ***IF*** Temperature is Mild ***AND*** Wind is Not-windy ***THEN*** Volleyball (S = 0.81)

Rule 6: ***IF*** Temperature is Cool ***THEN*** Weight-lifting (S = 0.88)

Note: Rule 3 can be simplified to Rule 3':

Rule 3': ***IF*** Outlook is Rain ***THEN*** Weight-lifting (S = 0.89)

FIGURE 1. The induced fuzzy decision tree and fuzzy rules of Yuan and Shaw's method (1995).

TABLE 2 Learning Result from the Small Training Data Set (Yuan & Shaw, 1995)

	Classification Known in Training Data			Classification with Learned Rules		
Case	Volleyball	Swimming	W-lifting	Volleyball	Swimming	W-lifting
1	0.0	0.8	0.2	0.0	0.9	0.0
2	1.0	0.7	0.0	0.4	0.6	0.0
3	0.3	0.6	0.1	0.2	0.7	0.3
4	0.9	0.1	0.0	0.7	0.3	0.3
5	0.0	0.0	1.0	0.3	0.1	0.9
6	0.2	0.0	0.8	0.3	0.0	0.7
7	0.0	0.0	1.0	0.0	0.0	1.0
8	0.7	0.0	0.3	0.2	0.0	0.8*
9	0.2	0.8	0.0	0.0	1.0	0.0
10	0.0	0.3	0.7	0.1	0.0	0.7
11	0.4	0.7	0.0	0.0	0.7	0.0
12	0.7	0.2	0.1	0.7	0.0	0.3
13	0.0	0.0	1.0	0.0	0.2	0.8
14	0.0	0.0	1.0	0.3	0.0	0.7
15	0.0	0.0	1.0	0.0	0.0	1.0
16	0.8	0.6	0.0	0.5	0.5	0.0†

* Wrong classification
† Cannot distinguish between two or more classes.

the class at the leaf with the highest classification truth level. The generated fuzzy rules after conversion from the fuzzy decision tree are also shown in Figure 1. Yuan and Shaw (1995) pointed out that Rule 3: "**IF** *Temperature* is *Hot* **AND** *Outlook is Rain* **THEN** *Weightlifting*" can be simplified into Rule 3′: "**IF** *Outlook is Rain* **THEN** *Weightlifting*." The truth level of Rule 3′ is 0.89 and is not less than 0.73 (the truth level of the original Rule 3). With the generated six fuzzy rules in Figure 1, the classification results for the training data shown in Table 1 are shown in Table 2. Among 16 training cases, 13 cases (except cases 2, 8, 16) are correctly classified. The classification accuracy of Yuan and Shaw's method is $\frac{13}{16} \times 100\% = 81\%$.

A NEW METHOD FOR GENERATING FUZZY RULES FROM NUMERICAL DATA

In the following, we present a new method for generating fuzzy rules from numerical data. The data set we use to introduce the concepts of fuzzy rules generation is shown in Table 3. In Table 3, we have nine cases with three attributes for each case and three kinds of decisions for each plan:

Attribute= {A, B, C}

TABLE 3 A Small Data Set for Illustrating the Proposed Fuzzy Rules Generation Method

	A			B			C		Plan		
Case	A1	A2	A3	B1	B2	B3	C1	C2	X	Y	Z
1	0.3	0.7	0	0.2	0.7	0.1	0.3	0.7	0.1	0.9	0
2	1	0	0	1	0	0	0.7	0.3	0.8	0.2	0
3	0	0.3	0.7	0	0.7	0.3	0.6	0.4	0	0.2	0.8
4	0.8	0.2	0	0	0.7	0.3	0.2	0.8	0.6	0.3	0.1
5	0.5	0.5	0	1	0	0	0	1	0.6	0.8	0
6	0	0.2	0.8	0	1	0	0	1	0	0.7	0.3
7	1	0	0	0.7	0.3	0	0.2	0.8	0.7	0.4	0
8	0.1	0.8	0.1	0	0.9	0.1	0.7	0.3	0	0	1
9	0.3	0.7	0	0.9	0.1	0	1	0	0	0	1

and each attribute has linguistic terms

$$A = \{A1,\ A2,\ A3\}$$
$$B = \{B1,\ B2,\ B3\}$$
$$C = \{C1,\ C2\}.$$

The classification is the decision to be made on a case with attributes A_i, B_j, and C_k, respectively, to carry out one of the plans X, Y, or Z:

$$\text{Plan} = \{X, Y, Z\}.$$

We want to generate fuzzy classification rules from the given numerical data in Table 3. The generated fuzzy classification rules are in the form of formula (6). For example,

Rule 1: **IF** A is A1 **THEN** Plan is X,
Rule 2: IF B is *NOT* B3 **AND** C is C2 **THEN** Plan is Y,
Rule 3: IF MF(Rule 1)< β **AND** MF(Rule 2)< β **THEN** Plan is Z,

are the rules that satisfied our purpose, where MF means "membership function value" (Yuan & Shaw, 1995), β is an applicability threshold value, and $\beta \in [0, 1]$.

All the attributes and classifications are vague by nature, since they represent a human's cognition and desire. For example, peoples' feeling of cool, mild, and hot are vague and there are no definite boundaries between them. Assume that attributes "A," "B," and "C" stand for some attributes of weather, respectively, and assume that "X," "Y," and "Z" stand for sport plans of "volleyball," "Swimming," and "Weightlifting" for a special day, respectively. Although there are distinctions between the sport plans such as "Swimming" or "Volleyball," the classification when it is interpreted as the desire to play can still be vague. For example, the weather can be excellent or just okay for swimming. The classification has the same situation. For example, the weather could be suitable for both swimming and playing volleyball and one may feel that it is difficult to select between the two.

As shown in Table 3, we have nine cases, where each case has three attributes to describe it. For each attribute, we have two or three terms to choose. In addition to the attribute part, we have to decide on a plan. One of decisions "X," "Y," or "Z" is the plan to be decided for a specific case. The value accompanying each term or decision of plan is in the range [0, 1]. For each case, we can decide which decision of plan (with the highest possibility value) is most likely to be chosen. For example, in Case 6, the possibility to choose decision "X" is 0, to choose decision "Y" is 0.7, to choose decision "Z" is 0.3, and the final decision is plan "Y."

From the possibility values of decisions "X," "Y," and "Z," for each case, we can decide which decision to make for a specific case. If we divide the nine cases into three subgroups according to the classification results, i.e., "X," "Y," and "Z," we can get another table as shown in Table 4. As Table 4 depicts, there are three instances for "X," three instances for "Y," and three instances for "Z," respectively. After carefully examining the table, it seems that there are close relationships between classification results (decision of plan for that subgroup) and some terms of the attributes. Making use of the fuzzy subsethood concept (Kosko, 1986; Yuan & Shaw, 1995), we can get information about the relationship between the decision of the plan and every distinct term of the attributes.

In each subgroup, we calculate the fuzzy subsethood values between decisions of that subgroup and every term of each attribute. After the computations of subsethood values, we can get a set of subsethood values for each decision. In this set of values, the larger the value, the closer the relationship between the decision of the plan and the term. For each subgroup, we can attain the most important factors that result in the decision of the

TABLE 4 Three Subgroups According to the Decision to be Made

		A			B			C		Plan		
Subgroup	Case	A1	A2	A3	B1	B2	B3	C1	C2	X	Y	Z
Subgroup_1	2	1	0	0	1	0	0	0.7	0.3	0.8	0.2	0
	4	0.8	0.2	0	0	0.7	0.3	0.2	0.8	0.6	0.3	0.1
	7	1	0	0	0.7	0.3	0	0.2	0.8	0.7	0.4	0
Subgroup_2	1	0.3	0.7	0	0.2	0.7	0.1	0.3	0.7	0.1	0.9	0
	5	0.5	0.5	0	1	0	0	0	1	0.6	0.8	0
	6	0	0.2	0.8	0	1	0	0	1	0	0.7	0.3
Subgroup_3	3	0	0.3	0.7	0	0.7	0.3	0.6	0.4	0	0.2	0.8
	8	0.1	0.8	0.1	0	0.9	0.1	0.7	0.3	0	0	1
	9	0.3	0.7	0	0.9	0.1	0	1	0	0	0	1

plan of that subgroup. We can use these terms to form the condition part of the classification rule for that decision of the plan. The consequent part of the rule is the decision of the plan for that subgroup.

From Table 4, we can see that there are three subgroups of cases. In each subgroup, the decision to be made is fixed. To find the closeness between the decision and each term of the three attributes, we first calculate the subsethood values for them. The meaning of fuzzy subsethood value is defined by using formula (5), A is a subset of B, defined by

$$\mathrm{S(A, B)} = \frac{M(A \cap B)}{M(A)}.$$

Take Subgroup_1 as an example ("X" is the decision of Subgroup_1), the denominator and the numerator of the subsethood formula for S(X, A1) are as follows:

$$\begin{aligned} \mathrm{M(X)} &= 0.8 + 0.6 + 0.7 = 2.1, \\ \mathrm{M(X \cap A1)} &= \mathrm{Min}(0.8, 1) + \mathrm{Min}(0.6, 0.8) + \mathrm{Min}(0.7, 1) \\ &= 0.8 + 0.6 + 0.7 \\ &= 2.1. \end{aligned}$$

The value of S(X, A1) is

$$\begin{aligned} \mathrm{S(X, A1)} &= \mathrm{M(X \cap A1)/M(X)} \\ &= 2.1/2.1 \\ &= 1, \end{aligned}$$

where $\mathrm{S(X, A1)}$ stands for the subsethood of "X" to "A1" of "A" in Subgroup_1.

Using the same formula (i.e., formula (5)), we can compute all the subsethood values as summarized in Figure 2. From Figure 2, we can find that some terms are closely related to the decision to be made in that subgroup and some are not. We need a standard to distinguish close or not close enough between the decision and terms of attributes. We use the level threshold value α as the standard to measure close enough or not on fuzzy subsethood values between the decision of the subgroup and all terms of attributes, where $\alpha \in [0, 1]$. Assume that the value we assigned to the level threshold α is 0.9. For each attribute, we can select at most one term. If there are two or more terms belonging to the same attribute which have a fuzzy subsethood value not less than 0.9, the one with the largest fuzzy subsethood value will be chosen. If there

Subgroup_1(X):

A:

S(X, A1) = 1	S(X, A2) = 0.1	S(X, A3) = 0

B:

S(X, B1) = 0.71	S(X, B2) = 0.43	S(X, B3) = 0.14

C:

S(X, C1) = 0.52	S(X, C2) = 0.76

Subgroup_2(Y):

A:

S(Y, A1) = 0.33	S(Y, A2) = 0.58	S(Y, A3) = 0.29

B:

S(Y, B1) = 0.42	S(Y, B2) = 0.58	S(Y, B3) = 0.04

C:

S(Y, C1) = 0.13	S(Y, C2) = 0.92

Subgroup_3(Z):

A:

S(Z, A1) = 0.14	S(Z, A2) = 0.64	S(Z, A3) = 0.29

B:

S(Z, B1) = 0.32	S(Z, B2) = 0.61	S(Z, B3) = 0.14

C:

S(Z, C1) = 0.82	S(Z, C2) = 0.25

FIGURE 2. The list of the fuzzy subsethood values for small data set.

are two terms with subsethood values not less than 0.9 at the same time, the term which is the original term of the attribute will have privilege over the one which is a complemented term of the same attribute.

Referring to Figure 2 the fuzzy subsethood values (including those for the complement terms) not less than the level threshold value α, where $\alpha = 0.9$, in Subgroup_1 are S(X, A1) = 1, S(X, **NOT** A2) = 0.9, and S(X, **NOT** A3) = 1. Because "A1," "A2," and "A3" are all terms of attribute "A," only one of them will be chosen. In this condition, "A1" is the only original term that belongs to attribute "A," and it is the one we choose among them. From this term we can generate the first fuzzy rule as follows:

Rule 1: IF A is A1 **THEN** Plan is X.

Likewise, the fuzzy subsethood values that are not less than 0.9 in Subgroup_2 are S(Y, **NOT** B3)= 0.96 and S(Y, C2)= 0.92. The generated fuzzy rule is as follows:

Rule 2: IF B is **NOT** B3 **AND** C is C2 **THEN** Plan is Y.

For a rule to be generated, there must be at least one original term not less than α. From Subgroup_3, we can see that the subsethood values are quite average. In this condition, no term is outstanding among them (no term has a value not less than 0.9 in Subgroup_3). This means that for decision "Z," those terms of attributes are average and no terms are representative enough. Thus, Rule 3 is unable to be generated at this time.

We use MF(Rule i)= MF(condition part of Rule i), where $1 \leq i \leq 2$ and MF means "membership function value" (Yuan & Shaw, 1995). If we want to classify Case 3 of Table 3, then we can get

$$\begin{aligned}
&\text{MF(condition part of Rule 1)} = \text{MF(A1)} = 0,\\
&\text{MF(condition part of Rule 2)} = \text{MF}(\textbf{NOT}\ \text{B3} \cap \text{C2}) = (1 - 0.3) \cap 0.4 = 0.4,\\
&\text{MF(Rule 1)} = \text{MF(condition part of Rule 1)} = 0,\\
&\text{MF(Rule 2)} = \text{MF(condition part of Rule 2)} = 0.4.
\end{aligned}$$

Because both membership values of the existing rules are not high enough to choose decision "X" or decision "Y," it is very possible that decision "Z" is more appropriate than the other two decisions. In this situation, we need another applicability threshold value β, where $\beta \in [0, 1]$, to judge the applicability of the existing rules. The existing rules are applicable to a case if MF(Rule i) $\geq \beta$, where $i \in \{1, \ldots, n\}$ and n is the number of existing rules.

As an alternative, we can conclude that a case that is not well classified by Rule 1 and Rule 2 will be classified into the plan with decision "Z." Thus, the third fuzzy rule is generated as follows:

Rule 3: IF MF(Rule 1) $< \beta$ **AND** MF(Rule 2) $< \beta$ **THEN** Plan is Z,

where MF(Rule i)= MF(condition part of Rule i), where $1 \leq i \leq 2$ and MF means "membership function value" (Yuan & Shaw, 1995), and β is an applicability threshold value that MF(Rule 1) or MF(Rule 2) must exceed if that rule is applicable to a case, where $\beta \in [0, 1]$.

To apply the generated fuzzy rules to each case of the data sets shown in Table 3, we must assign the applicability threshold value β in advance, where $\beta \in [0, 1]$. For each case, calculate MF(Condition part of Rule 1) and MF(Condition part of Rule 2), respectively, and then assign MF(Rule 1) = MF(Condition part of Rule 1), MF(Rule 2) = MF(Condition part of Rule 2). The applicability threshold value β is used to compare MF(Rule 1) and MF(Rule 2), respectively, for the specified case. If both MF(Rule 1) and MF(Rule 2) are less than β, then we let MF(Rule 3)=1. Otherwise, we let MF(Rule 3) = 0.

In the example of Table 3, the classification results of Rule 1, Rule 2, and Rule 3 are "X," "Y," and "Z," respectively. The possibility values of the classification result for a specific case with respect to "X," "Y," and "Z" are represented by "Plan(X)," "Plan(Y)," and "Plan(Z)" for a specific case, we can assign

$$\begin{aligned}
\text{Plan(X)} &= \text{MF(Rule 1)},\\
\text{Plan(Y)} &= \text{MF(Rule 2)},\\
\text{Plan(Z)} &= \text{MF(Rule 3)}.
\end{aligned} \tag{7}$$

The generated fuzzy rules at the level threshold value $\alpha = 0.9$ are listed as follows:

Rule 1: F A is A1 **THEN** Plan is X.

Rule 2: IF B is **NOT** B3 **AND** C is C2 **THEN** Plan is Y.

Rule 3: IF MF(Rule 1) $< \beta$ **AND** MF(Rule 2) $< \beta$ **THEN** Plan is Z.

Assume that the applicability threshold value β in the explained example is 0.6 (i.e., $\beta = 0.6$), then

1. From Case 1 of Table 3, we can get

$$\begin{aligned}
\text{MF(condition part of Rule 1)} &= \text{MF(A is A1)} = 0.3, \\
\text{MF(condition part of Rule 2)} &= \text{MF(B is } \textbf{NOT} \text{ B3 } \textbf{AND} \text{ C is C2)} \\
&= \text{MF(B is } \textbf{NOT} \text{ B3} \cap \text{C is not C2)} \\
&= \text{MF(B is } \textbf{NOT} \text{ B3)} \cap \text{MF(C is C2)} \\
&= \text{Min}\{(1 - 0.1), 0.7\} \\
&= 0.7,
\end{aligned}$$

$$\begin{aligned}
\text{MF(Rule 1)} &= \text{MF(condition part of Rule 1)} = 0.3, \\
\text{MF(Rule 2)} &= \text{MF(condition part of Rule 2)} = 0.7.
\end{aligned}$$

Because MF(Rule 1) $< \beta$ and MF(Rule 2)$> \beta$, where $\beta = 0.6$, thus MF(Rule 3) = 0.

From formula (7), the possibility values of the decisions of plan for Case 1 are

$$\begin{aligned}
\text{Plan(X)} &= \text{MF(Rule 1)} = 0.3, \\
\text{Plan(Y)} &= \text{MF(Rule 2)} = 0.7, \\
\text{Plan(Z)} &= \text{MF(Rulea3)} = 0,
\end{aligned}$$

and we fill Plan(X), Plan(Y), and Plan(Z) (i.e., 0.3, 0.7, and 0) into the last three columns of Case 1 in Table 5. Because Plan(Y) is the one with the highest possibility value among the values of Plan(X), Plan(Y), and Plan(Z), the decision to be made for Case 1 is "Y."

TABLE 5 Results After Applying the Generated Fuzzy Rules to Table 3

	A			B			C		Plan		
Case	A1	A2	A3	B1	B2	B3	C1	C2	X	Y	Z
1	0.3	0.7	0	0.2	0.7	0.1	0.3	0.7	0.3	0.7	0
2	1	0	0	1	0	0	0.7	0.3	1	0.3	0
3	0	0.3	0.7	0	0.7	0.3	0.6	0.4	0	0.4	1
4	0.8	0.2	0	0	0.7	0.3	0.2	0.8	0.8	0.7	0
5	0.5	0.5	0	1	0	0	0	1	0.5	1	0
6	0	0.2	0.8	0	1	0	0	1	0	1	0
7	1	0	0	0.7	0.3	0	0.2	0.8	1	0.8	0
8	0.1	0.8	0.1	0	0.9	0.1	0.7	0.3	0.1	0.3	1
9	0.3	0.7	0	0.9	0.1	0	1	0	0.3	0	1

2. From Case 2 of Table 3, we can get

$$\begin{aligned}
\text{MF(condition part of Rule 1)} &= \text{MF(A is A1)} = 1,\\
\text{MF(condition part of Rule 2)} &= \text{MF(B is } \textbf{NOT} \text{ B3 } \textbf{AND} \text{ C is C2)}\\
&= \text{MF(B is } \textbf{NOT} \text{ B3} \cap \text{C is C2)}\\
&= \text{MF(B is } \textbf{NOT} \text{ B3)} \cap \text{MF(C is C2)}\\
&= \text{Min}\{(1-0), 0.3\} = 0.3,
\end{aligned}$$

$$\begin{aligned}
\text{MF(Rule 1)} &= \text{MF(condition part of Rule 1)} = 1,\\
\text{MF(Rule 2)} &= \text{MF(condition part of Rule 2)} = 0.3.
\end{aligned}$$

Because MF(Rule 1) $> \beta$ and MF(Rule 2) $< \beta$, where $\beta = 0.6$, thus MF(Rule 3) = 0.

From formula (7), the possibility values of the decisions of plan for Case 2 are

$$\begin{aligned}
\text{Plan(X)} &= \text{MF(Rule1)} = 1,\\
\text{Plan(Y)} &= \text{MF(Rule 2)} = 0.3,\\
\text{Plan(Z)} &= \text{MF(Rule3)} = 0,
\end{aligned}$$

and we fill Plan(X), Plan(Y), and Plan(Z) (i.e., 1, 0.3, and 0) into the last three columns of Case 2 in Table 5. Because Plan(X) is the one with the highest possibility value among the values of Plan(X), Plan(Y), and Plan(Z), the decision to be made for Case 2 is "X."

3. From Case 3 of Table 3, we can get

$$\begin{aligned}
\text{MF(condition part of Rule 1)} &= \text{MF(A is A1)} = 0,\\
\text{MF(condition part of Rule 2)} &= \text{MF(B is }\textbf{NOT}\text{ B3 }\textbf{AND}\text{ C is C2)}\\
&= \text{MF(B is }\textbf{NOT}\text{ B3} \cap \text{C is C2)}\\
&= \text{MF(B is }\textbf{NOT}\text{ B3} \cap \text{MF(C is C2)}\\
&= \text{Min}\{(1-0.3), 0.4\}\\
&= 0.4,
\end{aligned}$$

$$\begin{aligned}
\text{MF(Rule 1)} &= \text{MF(condition part of Rule 1)} = 0,\\
\text{MF(Rule 2)} &= \text{MF(condition part of Rule 2)} = 0.4.
\end{aligned}$$

Because MF(Rule 1) $< \beta$ and MF(Rule 2) $< \beta$, where $\beta = 0.6$, thus MF(Rule 3) = 1.

From equation (7), the possibility values of the decisions of plan for Case 3 are

$$\begin{aligned}
\text{Plan(X)} &= \text{MF(Rule1)} = 0,\\
\text{Plan(Y)} &= \text{MF(Rule2)} = 0.4,\\
\text{Plan(Z)} &= \text{MF(Rule3)} = 1,
\end{aligned}$$

and we fill Plan(X), Plan(Y), and Plan(Z) (i.e., 0, 0.4, and 1) into the last three columns of Case 3 in Table 5. Because Plan(Z) is the one with the highest possibility value among the values of Plan(X), Plan(Y), and Plan(Z), the decision to be made for Case 3 is "Z."

The other cases are treated in a similar way. We summarize the result in Table 5.

Based on the generated fuzzy classification rules, the classification results for the training data in Table 3 are shown in Table 5. Among nine training cases, all cases are correctly classified. The classification accuracy rate is 100%.

EXPERIMENT RESULTS

In the following, we use an example (Yuan & Shaw, 1995) (i.e., the Saturday Morning Problem) to illustrate the fuzzy rules generation process.

Example. Assume that the small data set we use here is the same as Yuan and Shaw (1995) and shown in Table 1. From Table 1, we can see that there are four attributes for each case and there are three kinds of sport for each plan:

Attribute = {Outlook, Temperature, Humidity, Wind),

and each attribute has terms shown as follows:

Outlook {Sunny, Cloudy, Rain),

Temperature = {Hot, Cool, Mild),

Humidity = {Humid, Normal),

Wind = {Windy, Not-windy).

The classification result is the sport plan to be played on the weekend day:

Plan = {Volleyball, Swimming, Weightlifting}.

Assume that the values for level threshold value α and applicability threshold value β are 0.9 and 0.6, respectively (i.e., $\alpha = 0.9$ and $\beta = 0.6$). From Table 1, we divide the 16 cases into three subgroups according to the sport plan with the highest possibility value in each case. The result of the division is as follows (refer to Table 6):

1. Subgroup_1 with "Volleyball" as the activity to be taken: Cases 2, 4, 8, 12, and 16.
2. Subgroup_2 with "Swimming" as the activity to be taken: Cases 1, 3, 9, and 11.
3. Subgroup_3 with "Weightlifting" as the activity to be taken: Cases 5, 6, 7, 10, 13, 14, and 15.

According to formula (5), the calculations for subsethood for all three subgroups are shown in Figure 3.

According to the previous discussions, there are three fuzzy rules to be generated for $\alpha = 0.9$ and $\beta = 0.6$ which are summarized as follows:

Rule 1: **IF** Outlook is **NOT** Rain **AND** Humidity is Normal **AND** Wind is Not-windy
THEN Plan is Volleyball.

Rule 2: **IF** Outlook is **NOT** Rain **AND** Temperature is Hot
THEN Plan is Swimming.

Rule 3: **IF** MF(Rule 1) $< \beta$ **AND** MF(Rule 2) $< \beta$
THEN Plan is Weightlifting.

TABLE 6 Three Subgroups According to the Sport to be Taken

		Outlook			Temperature			Humidity		Wind		Plan		
Sub-Group	Case	Sunny	Cloudy	Rain	Hot	Mild	Cool	Humid	Normal	Windy	Not-windy	Volleyball	Swimming	W-lifting
Subgroup_1	2	0.8	0.2	0	0.6	0.4	0	0	1	0	1	1	0.7	0
	4	0.2	0.7	0.1	0.3	0.7	0	0.2	0.8	0.3	0.7	0.9	0.1	0
	8	0	1	0	0	0.2	0.8	0.2	0.8	0	1	0.7	0	0.3
	12	0.2	0.6	0.2	0	1	0	0.3	0.7	0.3	0.7	0.7	0.2	0.1
	16	1	0	0	0.5	0.5	0	0	1	0	1	0.8	0.6	0
Subgroup_2	1	0.9	0.1	0	1	0	0	0.8	0.2	0.4	0.6	0	0.8	0.2
	3	0	0.7	0.3	0.8	0.2	0	0.1	0.9	0.2	0.8	0.3	0.6	0.1
	9	1	0	0	1	0	0	0.6	0.4	0.7	0.3	0.2	0.8	0
	11	0.7	0.3	0	1	0	0	1	0	0.2	0.8	0.4	0.7	0
Subgroup_3	5	0	0.1	0.9	0.7	0.3	0	0.5	0.5	0.5	0.5	0	0	1
	6	0	0.7	0.3	0	0.3	0.7	0.7	0.3	0.4	0.6	0.2	0	0.8
	7	0	0.3	0.7	0	0	1	0	1	0.1	0.9	0	0	1
	10	0.9	0.1	0	0	0.3	0.7	0	1	0.9	0.1	0	0.3	0.7
	13	0.9	0.1	0	0.2	0.8	0	0.1	0.9	1	0	0	0	1
	14	0	0.9	0.1	0	0.9	0.1	0.1	0.9	0.7	0.3	0	0	1
	15	0	0	1	0	0	1	1	0	0.8	0.2	0	0	1

Subgroup_1(Volleyball):
Outlook:
S(Volleyball, Sunny) = 0.49 S(Volleyball, Cloudy) = 0.54
S(Volleyball, Rain) = 0.07
Temperature:
S(Volleyball, Hot) = 0.34 S(Volleyball, Mild) = 0.61
S(Volleyball, Cool) = 0.17
Humidity:
S(Volleyball, Humid) = 0.17 S(Volleyball, Normal) = 0.98
Wind:
S(Volleyball, Windy) = 0.15 S(Volleyball, Not-windy) = 0.95

Subgroup_2(Swimming):
Outlook:
S(Swimming, Sunny) = 0.79 S(Swimming, Cloudy) = 0.35
S(Swimming, Rain) = 0.10
Temperature:
S(Swimming, Hot) = 1 S(Swimming, Mild) = 0.07
S(Swimming, Cool) = 0
Humidity:
S(Swimming, Humid) = 0.76 S(Swimming, Normal) = 0.41
Wind:
S(Swimming, Windy) = 0.52 S(Swimming, Not-windy) = 0.76

Subgroup_3(Weight-lifting):
Outlook:
S(Weight-lifting, Sunny) = 0.25 S(Weight-lifting, Cloudy) = 0.34
S(Weight-lifting, Rain) = 0.46
Temperature:
S(Weight-lifting, Hot) = 0.14 S(Weight-lifting, Mild) = 0.4
S(Weight-lifting, Cool) = 0.54
Humidity:
S(Weight-lifting, Humid) = 0.37 S(Weight-lifting, Normal) = 0.66
Wind:
S(Weight-lifting, Windy) = 0.65 S(Weight-lifting, Not-windy) = 0.4

FIGURE 3. The list of the fuzzy subsethood values for the Saturday Morning Problem.

TABLE 7 Learning Result of the Saturday Morning Problem with the Generated Fuzzy Rules

	Classification Known in Training Data			Classification with Learned Rules		
Case	Volleyball	Swimming	W-lifting	Volleyball	Swimming	W-lifting
1	0.0	0.8	0.2	0.2	1	0
2	1.0	0.7	0.0	1	0.6	0
3	0.3	0.6	0.1	0.7	0.7	0*
4	0.9	0.1	0.0	0.7	0.3	0
5	0.0	0.0	1.0	0.1	0.1	1
6	0.2	0.0	0.8	0.3	0	1
7	0.0	0.0	1.0	0.3	0	1
8	0.7	0.0	0.3	0.8	0	0
9	0.2	0.8	0.0	0.3	1	0
10	0.0	0.3	0.7	0.1	0	1
11	0.4	0.7	0.0	0	1	0
12	0.7	0.2	0.1	0.7	0	0
13	0.0	0.0	1.0	0	0.2	1
14	0.0	0.0	1.0	0.3	0	1
15	0.0	0.0	1.0	0	0	1
16	0.8	0.6	0.0	1	0.5	0

*Cannot distinguish between two or more classes.

TABLE 8 A Comparison of the Number of Generated Fuzzy Rules and Accuracy Rate Between the Yuan and Shaw's (1995) Method and the Proposed Method

	Yuan and Shaw's Method (1995)	The Proposed Method (for $\alpha = 0.9$ and $\beta = 0.6$)
Number of rules	6	3
Accuracy rate	81.25%	93.75%

Based on the previous discussions, we can apply the generated fuzzy rules to Table 1. The classification results of the application of the generated fuzzy rules are summarized in Table 7. From Table 7, we can see that among 16 training cases, 15 cases (except Case 3) are correctly classified. The classification accuracy rate is $\frac{15}{16} \times 100\% = 93.75\%$.

A comparison of the number of generated fuzzy rules and accuracy rate between the proposed method and Yuan and Shaw's method (1995) is listed in Table 8.

From Table 8, we can see that the accuracy rate of the proposed method is better than that of Yuan and Shaw's method under $\alpha = 0.9$ and $\beta = 0.6$. The number of rules generated by the proposed method is less than the number of rules generated by Yuan and Shaw's method.

CONCLUSIONS

We have presented a new method for generating fuzzy rules from numerical data for handling fuzzy classification problems based on the fuzzy subsethood values between the decisions to be made and terms of attributes of subgroups by using the level threshold value α and the applicability threshold value β, where $\alpha \in [0, 1]$ and $\beta \in [0, 1]$. We also apply the proposed method to deal with the Saturday Morning Problem (Yuan & Shaw, 1995). The proposed method has the following advantages:

1. The proposed method gets a better accuracy rate than the one presented in Yuan and Shaw (1995). From Table 8, we can see that the accuracy rate of the proposed method is 93.75% (for $\alpha = 0.9$ and $\beta = 0.6$), while the accuracy rate of the Yuan and Shaw's method is 81.25%.
2. The proposed method generates fewer fuzzy rules than the one presented in Yuan and Shaw (1995). From the experimental results, we can see that the number of fuzzy rules generated by the proposed method is 3, but the number of fuzzy rules generated by Yuan and Shaw's method is 6.
3. The proposed method needs less calculations than the one presented in Yuan and Shaw (1995).

REFERENCES

Castro, J. L., and J. M. Zurita. 1997. An inductive learning algorithm in fuzzy systems. *Fuzzy Sets and Systems* 89(2):193–203.

Chen, S. M., and M. S. Yeh. 1998. Generating fuzzy rules from relational database systems for estimating null values. *Cybernetics and Systems: An International Journal* 29(6):363–376.

Chen, S. M., S. H. Lee, and C. H. Lee. 1999. Generating fuzzy rules from numerical data for handling fuzzy classification problems. *Proceedings of the 1999 National Computer Symposium*, Taipei, Taiwan, Republic of China, Vol. 2, 336–343.

Hayashi, Y., and A. Imura. 1990 Fuzzy neural expert system with automated extraction of fuzzy if-then rules from a trained neural network. *Proceedings of the 1990 First International Symposium on Uncertainty Modeling and Analysis*, University of Maryland, College Park, MD.

Hong, T. P., and C. Y. Lee. 1996. Induction of fuzzy rules and membership functions from training examples. *Fuzzy Sets and Systems* 84(1):33–47.

Klawonn, F., and R. Kruse. 1997. Constructing a fuzzy controller from data. *Fuzzy Sets and Systems* 85(2):177–193.

Kosko, B. 1986. Fuzzy entropy and conditioning. *Information Sciences* 40(2):165–174.

Nozaki, K., H. Ishibuchi, and H. Tanaka. 1997. A simple but powerful heuristic method for generating fuzzy rules from numerical data. *Fuzzy Sets and Systems* 86(3):251–270.

Quinlan, J. R. 1994. Decision trees and decision making. *IEEE Transactions on Systems, Man, and Cybernetics* 20(2):339–346.

Spirtes, P., C. Glymour, and R. Scheines. 1993. *Causation, Prediction, and Search*. New York: Springer-Verlag.

Wang, L. X., and J. M. Mendel. 1992. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man, and Cybernetics* 22(6):1414–1427.

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

Yoshinari, Y., W. Pedrycz, and K. Hirota. 1993. Construction of fuzzy models through clustering techniques. *Fuzzy Sets and Systems* 54(2):157–165.

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

Yuan, Y., and H. Zhuang. 1996. A genetic algorithm for generating fuzzy classification rules. *Fuzzy Sets and Systems* 84(1):1–19.

Zadeh, L. A. 1965. Fuzzy sets. *Information and Control* 8:338–353.

Zadeh, L. A., 1988. Fuzzy logic. *IEEE Computer* 21(4):83–91.

Zadeh, L. A. 1975. The concept of a linguistic variable and its approximate reasoning – I. *Information Sciences* 8(3):199–249.