Handling Interaction in Fuzzy Production Rule Reasoning

Daniel S. Yeung, *Fellow, IEEE*, Xi-Zhao Wang, *Senior Member, IEEE*, and Eric C. C. Tsang

***Abstract*—When fuzzy production rules are used to approximate reasoning, interaction exists among rules that have the same consequent. Due to this interaction, the weighted average model frequently used in approximate reasoning does not work well in many real-world problems. In order to model and handle this interaction, this paper proposes to use a nonadditive nonnegative set function to replace the weights assigned to rules having the same consequent, and to draw the reasoning conclusion based on an integral with respect to the nonadditive nonnegative set function, rather than on the weighted average model. Handling interaction in fuzzy production rule reasoning in this way can lead to a good understanding of the rules base and an improvement of reasoning accuracy. This paper also investigates how to determine from data the nonadditive set function that cannot be specified by a domain expert.**

***Index Terms*—Approximate reasoning, fuzzy integrals and measures, fuzzy production rules, interaction, rule-based reasoning.**

I. Introduction

FUZZY PRODUCTION rules (FPRs) are widely used in expert systems to represent fuzzy and uncertain concepts. FPRs are usually presented in the form of a fuzzy IF-THEN rule in which both the antecedent and the consequent are fuzzy concepts denoted by fuzzy sets. To effectively represent both the fuzziness and the uncertainty in FPRs, several knowledge parameters such as certainty factor, local weight, threshold value, and global weight have been incorporated into the FPRs ([1], [2], [27]). For example, one usually sees an FPR such as IF A AND B THEN C (CF) where A, B, and C are fuzzy sets and CF denotes the certainty factor.

FPR reasoning in modern expert systems is a very complicated process. Given a set of FPRs (usually called a knowledge base) and an observed fact, FPR reasoning is used to draw an approximate conclusion by matching the observed fact against the set of FPRs. Many researchers have investigated this fundamental issue in fuzzy reasoning. For instance, Yuan and Shaw in [30] proposed the use of operators (min, max) to model FPR reasoning, while Wang *et al.* in [26] extended the operators (min, max) to a generalized case. Yeung *et al.* [27] presented a weighted fuzzy production rule (WFPR) and proposed an improved method to compute the certainty factor of the consequent assertion and a better way to interpret the linguistic meaning of the consequent. Yeung and Tsang [28] compared the reasoning mechanism of the proposed WFPRs with other similarity-based fuzzy reasoning methods. Further, weighted fuzzy reasoning could be discussed by using weighted fuzzy Petri nets [4] or could be extended to multilevels or multistages [29], i.e., the consequents of some rules can be the antecedents of other rules. In this paper, however, we will consider only one-step FPR reasoning and will focus on the interaction that exists among the FPRs in the knowledge base.

One example to indicate the existence of interaction among rules is given as follows. Suppose we have three rules:

Rule 1: IF (Attribute A = Getting a fever) and (Attribute B = coming from a SARS infected area) THEN possibly SARS.
Rule 2: IF (Attribute C = Having a cough) and (Attribute D = Breathing difficulty) THEN possibly SARS.
Rule 3: IF (Attribute A = Getting a fever) and (Attribute E = Having much phlegm) THEN possibly SARS.

A person who satisfies both Rules 1 and 2 antecedents will be infected with SARS more likely than another person who only satisfies the antecedent of Rule 1 or Rule 2. This example may explain that Rules 1 and 2 are enhancing each other for leading to the SARS. However, a person who satisfies both Rules 2 and 3 antecedents will be infected with SARS less likely than another person who only satisfies the antecedent of Ruled 2 or 3. It is because (Having a cough) and (Having much phlegm) result in a nondry cough. This example may explain that Rules 2 and 3 are resisting each other for leading to the SARS. The enhancing-effect or resisting-effect is considered in this paper as the interaction among rules.

Why do we need interaction? The reason is simple.

1) It can help domain experts discover new knowledge existed among the rules. With respect to a given consequent (e.g., SARS), knowing and modeling the enhancing/resisting-effect among rules learned from data is helpful to maintaining the rule base (i.e., the knowledge maintenance).
2) By discovering the interaction among the rules and then applying it to fuzzy reasoning, the reasoning accuracy is expected to be improved. There have been many approaches to improve the reasoning accuracy. The improvement proposed in this paper has the more interpretability (i.e., the knowledge background of interactive effect among the rules).

Manuscript received June 5, 2003; revised January 8, 2004. This work was supported by a Hong Kong Polytechnic University Research Grant (GT-210), by the Natural Science Foundation of Hebei Province (603137, F2004000129), and by a key project of the Chinese Educational Ministry (03017). This paper was recommended by Associate Editor A. F. G. Skarmeta.

D. S. Yeung and E. C. C. Tsang are with the Department of Computing, Hong Kong Polytechnic University, Kowloon, Hong Kong, China (e-mail: csdaniel@comp.polyu.edu.hk).

X.-Z. Wang is with the Faculty of Mathematics and Computer Science, Hebei University, Baoding, Hebei, China (e-mail: wangxz@mail.hbu.edu.cn; csetsang@comp.polyu.edu.hk).

Digital Object Identifier 10.1109/TSMCB.2004.831460

The interaction among attributes has already been a clear concept [8], [10], [11], but the interaction among rules is firstly proposed in this paper. In fact, the interaction among rules is the interaction among attribute-values. For example, one can consider the interaction among attributes A, B, C, D, and E as shown in the above Rules 1–3. It can be elaborated by considering interaction among attribute-values, e.g., the interaction between $(\text{Attribute A} = \text{Getting a fever}) \wedge (\text{Attribute B} = \text{coming from a SARS infected area})$ and $(\text{Attribute C} = \text{Having a cough}) \wedge (\text{Attribute D} = \text{Breathing difficulty})$. Since the rules' antecedents consist of $\text{Attribute} = \text{attribute}-\text{vales}$, the interaction among the attribute values in this paper is called the interaction among the rules.

One popular model for representing the interaction is the fuzzy measure. In [8], the author proposed to represent the importance and interaction of features by fuzzy measures. Then, in [5], the authors proposed an interaction transform of set functions over a finite set, which leads to a new interaction index. In [11], the authors discussed the extension of pseudo-Boolean functions for the aggregation of interaction criteria. Marichal in [15], [16] showed that the discrete Choquet integral is an adequate aggregation operator that extends the weighted arithmetic mean by taking into consideration the interaction among criteria. He also showed that the mutual preferential independence of criteria reduces the Sugeno integral to a dictatorial aggregation.

To handle the interaction that exists among a set of rules having the same consequent, this paper proposes to use a nonadditive nonnegative set function to replace global weights (which are assigned to rules in a rule-set) to draw the reasoning conclusion of a set of rules based on an integral with respect to the nonadditive nonnegative set function, rather than on the weighted average model. Such a handling of interaction in FPR reasoning can lead to a well understanding to the rule base, and lead to an improvement of reasoning accuracy. Moreover, this paper also investigates how to determine from data the nonadditive set function that cannot be specified by experts.

This paper is organized as follows: Section II reviews the WFPR reasoning and states the globally WFPR reasoning. Section III outlines the Choquet integral, a kind of nonlinear integral with respect to a nonadditive set function, and discusses the interaction among a set of rules in the reasoning process. Section IV investigates how to determine the nonadditive set function from data and reports our numerical experimental results for determining the set functions and improving reasoning accuracy. Section V offers a number of remarks and concludes this paper.

II. Globally WFPR Reasoning

Basically, the reasoning addressed in this paper is similarity-based reasoning. We first review a definition of a similarity measure.

Definition 1: Let X be a universe of discourse and $\mathrm{F(X)}$ be the set of all fuzzy subsets defined on X. A mapping SM from $\mathrm{F(X)} \times \mathrm{F(X)}$ to $[0,1]$, is called a similarity measure if SM satisfies: 1) $\mathrm{SM(A,B)} = \mathrm{SM(B,A)}$ for any $\mathrm{A,B} \in \mathrm{F(X)}$ and 2) $\mathrm{SM(A,B)} = 1$ whenever $\mathrm{A} = \mathrm{B}$.

The similarity measure between two fuzzy subsets can be defined based on their membership functions. Discussions of similarity metrics can be found in many articles e.g., [20], [19], [12]. This paper does not discuss the details of similarity measures.

According to Zadeh's initial definition of Generalized Modus Ponens [6], the reasoning model is described as

$$\begin{aligned}&\text{A fuzzy production rule: IF `x is A' THEN `y is B'}\\&\text{A given fact : `x is A}*\text{'}\\&\text{A conclusion : `y is B}*\text{'}.\end{aligned}$$

This paper focuses on a type of globally WFPRs [27], which are specific to classification problems. A WFPR has a conjunctive form

$$\wedge_{j=1}^{n}(V_j = A_j) \Rightarrow (U = C), \mathrm{GW}$$

where $V_j (j = 1,2,\ldots,n)$ and U are variables; $A_j (j = 1,2,\ldots,n)$ and C are fuzzy values of these variables; GW is a nonnegative value denoting the global weight of the rule R; and $\wedge$ denotes the conjunction AND. WFPRs will degenerate to FPRs in commonsense when the global weight is ignored.

Consider a set of m WFPRs: $S = \{R_i, i = 1,2,\ldots,m\}$ and a given fact, the reasoning model which slightly modifies the Generalized Modus Ponens is described as

$$\begin{aligned}&\text{A set of WFPRs} : R_i : \wedge_{j=1}^{n_i}\left(V_j = A_j^{(i)}\right) \Rightarrow (U = C),\\&\qquad\qquad \mathrm{GW}(R_i) \quad i = 1,2,\ldots,m\\&\text{A given fact: } (V_j = B_j) \quad j = 1,2,\ldots,n\\&\text{A conclusion: } U = D, \quad \mathrm{CF}(D)\end{aligned}$$

where $\mathrm{GW}(R_i)$ represents the global weight assigned to the ith rule R_i and $\mathrm{CF}(D)$is the certainty factor of the conclusion. It is worth nothing that the m rules have the same consequent C.

How to draw the conclusion: $U = D, \mathrm{CF}(D)$? The following is a scheme to draw the conclusion and compute its certainty factor. We call the scheme globally WFPR reasoning.

Globally WFPR Reasoning Algorithm

Step (A1): For each rule R_i within S, the similarity between the proposition $A_j^{(i)}$and the observed attribute-value B_j, denoted by $\mathrm{SM}_j^{(i)}$, is defined below:

If B_j is a fuzzy set then $\mathrm{SM}_j^{(i)} = \mathrm{SM}(A_j^{(i)}, B_j)$ where SM is a given similarity measure.

If B_j is a real number then $\mathrm{SM}_j^{(i)} = A_j^{(i)}(B_j)$ where $A_j^{(i)}(\cdot)$ denotes its membership function.

The overall similarity $\mathrm{SM}^{(i)}$ is defined as

$$\mathrm{SM}^{(i)} = \mathrm{Min}_{1 \le j \le n_i} \mathrm{SM}_j^{(i)} \tag{1}$$

where n_i is the number of propositions of antecedent of the ith rule R_i.

Step (A2): Compute $M = \sum_{i=1}^{m} \mathrm{GW}(R_i)$ and

$$\mathrm{SM} = \sum_{i=1}^{m} \frac{\mathrm{GW}(R_i)}{M} \mathrm{SM}^{(i)}, \tag{2}$$

Step (A3): The conclusion's certainty factor $\mathrm{CF}(D)$ is given by

$$\mathrm{CF}(D) = \sum_{i=1}^{m} \left(\frac{\mathrm{GW}(R_i)}{M} \mathrm{SM}^{(i)} \right) \tag{3}$$

For classification problems, the value given in (3) denotes the degree of truth of the object belonging to class C. If there are K classes (corresponding to K sets of fuzzy rules), then the computed result in (3) refers to the degree of truth of some class, denoted by $x_k (k = 1, 2, \ldots, K)$. The normalized form of the inferred result is defined as $(d_1, d_2, \ldots, d_k)$, where $d_k = x_k / \mathrm{Max}_{1 \le j \le K} x_j (k = 1, 2, \ldots, K)$.

When a crisp inferred result is needed, one can take the consequent with maximum $d_k (1 \le k \le K)$. One problem is that the algorithm cannot give a crisp decision if there exists more than one maximum $d_k (1 \le k \le K)$. In that situation, we need another defuzzification method to determine the crisp decision. We now illustrate this reasoning mechanism and its shortcomings.

Example 1: Consider the following two sets of fuzzy rules.

Set1: $\{R_1, R_2, R_3, R_4\}$ having the same consequent Class1 and 4 global weights Gw1, Gw2, Gw3, and Gw4.

Set2: $\{R_5\}$with consequent Class 2 and global weight 1.

Suppose that we have two testing cases e1 and e2 with the crisp consequents Class1 and Class2. The degrees of similarity between rule antecedents and the two cases are as follows:

	R_1	R_2	R_3	R_4	R_5	Actual Consequent
e1	0.6	0.0	0.7	0.8	0.5	Class1
e2	0.6	0.7	0.0	0.8	0.5	Class2.

It is clear, if an appropriate global weight assignment is given, that the reasoning conclusion for the two cases will be correct. For instance, one can select $(\mathrm{Gw1}, \mathrm{Gw2}, \mathrm{Gw3}, \mathrm{Gw4}) = (0.2, 0.2, 0.4, 0.2)$, then the degrees of truth are

$$0.2 \cdot 0.6 + 0.2 \cdot 0 + 0.4 \cdot 0.7 + 0.2 \cdot 0.8 = 0.56 > 0.5$$

for e1 and

$$0.2 \cdot 0.6 + 0.2 \cdot 0.7 + 0.4 \cdot 0 + 0.2 \cdot 0.8 = 0.42 < 0.5.$$

for e2, respectively.

This simple example clearly shows that the global weight can be used to adjust the degrees of truth of some inferred consequents. If no global weights are used (e.g., the simple average or maximum operation is used), the inferred conclusions of Example1 cannot be correct. However, if some additional weight information including experts' experiences and previous learning results are incorporated into the reasoning process, it will become more complicated. For example, domain experts specify that if the rule R_2 has the same degree of importance as rule R_3, i.e., $\mathrm{Gw2} = \mathrm{Gw3}$, then it will be impossible to achieve the correct reasoning results for e1 and e2 by using the proposed globally weighted reasoning algorithm, since for both e1 and e2 the computed degrees of truth are the same value which is equal to $0.6\mathrm{Gw1}+0.7\mathrm{Gw2}+0.8\mathrm{Gw4}$. This is a drawback of the globally WFPR reasoning model. This paper proposes to overcome this problem through substituting a nonadditive nonnegative set function for global weights. It draws the reasoning conclusion by matching a new given fact against a set of rules based on an integral with respect to the nonadditive nonnegative set function, rather than on the weighted average model.

III. Interaction Represented by Nonadditive Set Function

Let X be a nonempty set and$P(X)$ be the power set of X. We use the symbol μ to denote a nonnegative set function defined on $P(X)$ with the properties $\mu(\oslash) = 0$ and $\mu(X) = 1$. When X is finite, μ is usually called a fuzzy measure if it satisfies monotonicity

$$A \subseteq B \Rightarrow \mu(A) \le \mu(B) \quad \text{for } A, B \in P(X).$$

For a nonnegative set function μ, there are some associated concepts. μ is said to be additive if $\mu(A \cup B) = \mu(A) + \mu(B)$ for $A \cap B = \oslash, A, B \in F(X)$; to be sub-additive if $\mu(A \cup B) \le \mu(A) + \mu(B)$ for $A, B \in F(X)$ and $A \cap B = \oslash$; and to be super-additive if $\mu(A \cup B) \ge \mu(A) + \mu(B)$ for $A, B \in F(X)$ and $A \cap B = \oslash$. Let $X = \{R_1, R_2, \ldots, R_n\}$ be a set of rules with the same consequent. We now consider A and B as two subsets of rules. If we consider $\mu(A)$ as the importance of subset A, then the additivity of the set function means that there is no interaction among the rules such that the joint importance of some rules is just the sum of their respective importance. However, this is not true in many real problems. In other words, most measures of importance are nonadditive [22]. Subadditivity and super-additivity are two special types of nonadditivity. Super-additivity means that the joint importance of two sets of rules is greater than or equal to the sum of their respective importance, which indicates that the two sets of rules enhance each other. Subadditivity means that the joint importance of two sets of rules is less than or equal to the sum of their respective importance, which indicates that the two sets of rules resist each other.

In [22], the concept of weights is extended to a nonnegative set function μ called an importance measure, and then the integral of a nonnegative function with respect to a nonadditive set function is defined, which generalizes the classical Lebesgue integral. In recent decades, a number of new kinds of nonlinear integrals have been introduced. Here we select the Choquet integral [9], a nonlinear integral with respect to a nonadditive set function or a fuzzy measure, because the Choquet integrals have been shown to be an adequate aggregation operator [15].

Definition 2: Let $X = \{x_1, x_2, \ldots, x_n\}, \mu$ be a fuzzy measure or a nonadditive set function defined on the power set of X, f be a function from X to $[0, 1]$. The Choquet integral of $f$with respect to μ is defined by

$$(C) \int_X f d\mu = \sum_{i=1}^{n} (f(x_i) - f(x_{i-1}))\mu(A_i)$$

where we assume without loss of generality that $0 = f(x_0) \le f(x_1) \le \cdots \le f(x_n)$ and $A_i = \{x_i, \ldots x_n\}$. If no confusion

TABLE I
VALUES OF SET FUNCTION μ IN EXAMPLE 2

Set	Value of μ
$\varnothing$	0
$\{x_1\}$	0.5
$\{x_2\}$	0.2
$\{x_1, x_2\}$	0.95
$\{x_3\}$	0.3
$\{x_1, x_3\}$	0.55
$\{x_2, x_3\}$	0.85
X	1

TABLE II
VALUES OF FUNCTION f IN EXAMPLE 2

x_i	$f(x_i)$
x_1	1.0
x_2	0.9
x_3	0.5

arises, we can omit the X and in short denote the Choquet integral by $(C)\int_{f d\mu}$.

Example 2: Let $X = \{x_1, x_2, x_3\}$. The values of μ and f are shown in Tables I and II, respectively. Here, μ is a fuzzy measure.

After reordering the values of the function$f$, we can use the formula given in definition 2 to compute the Choqet integral. It is $(C)\int f\, d\mu = 0.93$.

We now state our FPR reasoning algorithm based on fuzzy integral. Consider K sets of FPRs $S_k = \{R_i^{(k)}, i = 1, 2, \ldots, m_k\}, k = 1, 2, \ldots, K$. For each k $(1 \le k \le K)$, the kth set of FPRs S_k has the same consequent Class_k, and a nonnegative set function μ_k defined on $2^{S_k} = 2^{\{R_1^{(k)}, R_2^{(k)}, \ldots, R_{m_k}^{(k)}\}}$ where 2^{S_k} denotes the power set of S_k, i.e., the set of all subsets of S_k. The μ_k is used to represent the interaction among the rules within the set S_k. The μ_k here is assumed to be given in advance by domain experts and will be learned from data in Section IV. The FPR reasoning with interaction among rules is described as follows.

Algorithm of FPR Reasoning With Interaction

For $k = 1, 2, \ldots, K$ **DO**

Step (B1): For each rule $R_i^{(k)}$ within S_k, let $R_i^{(k)}$ have the following form:

$\wedge_{j=t(1)}^{t(n_{ki})}(V_j = A_{kj}^{(i)}) \Rightarrow (U = \text{Class}_k)$ where $\{t(1), t(2), \ldots t(n_{ki})\} \subset \{1, 2, \ldots, n\}, n$ is the number of all attributes. The observed fact is $\wedge_{j=1}^{n}(V_j = B_j)$ and the similarity between $A_{kj}^{(i)}$ and B_j, denoted by $SM_{kj}^{(i)}$, is determined similarly to Step (A1) of the Globally WFPR reasoning algorithm given in Section II.

Step (B2): Define an m_k-dimensional real vector, i.e., a discrete function SM_k on $S_k = \{R_i^{(k)}, i = 1, 2, \ldots, m_k\}$ by

$$\text{SM}_k = \left(\text{SM}_k^{(1)}, \text{SM}_k^{(2)}, \ldots, \text{SM}_k^{(m_k)}\right). \tag{4}$$

Step (B3): Evaluate the certainty factor of the conclusion $\text{CF}(\text{Class}_k)$ by

$$\text{CF}(\text{Class}_k) = (C)\int_{S_k} \text{SM}_k\, d\mu_k \tag{5}$$

where the integral is Choquet integral, defined in definition 2.

END DO

The final crisp classification result for the observed facts $\wedge_{j=1}^{n}(V_j = B_j)$ will be determined to be Class_{k_0} where $\text{CF}(\text{Class}_{k_0}) = \text{Max}_{1 \le k \le K}\text{CF}(\text{Class}_k)$.

Example 3: Let us continue considering Example 1 where the importance of rule 2 is assumed to be the same as that of rule 3. The other additional information for these four rules is given in Table III. It is noted that Table III specifies a nonadditive set function and the interaction among the four rules is included.

TABLE III
A SET FUNCTION

Set	Value	Set	Value
$\varnothing$	0.0	$\{R_2, R_3\}$	0.10
$\{R_1\}$	0.2	$\{R_2, R_4\}$	0.15
$\{R_2\}$	0.3	$\{R_3, R_4\}$	0.85
$\{R_3\}$	0.3	$\{R_1, R_2, R_3\}$	0.30
$\{R_4\}$	0.2	$\{R_1, R_2, R_4\}$	0.20
$\{R_1, R_2\}$	0.15	$\{R_1, R_3, R_4\}$	0.90
$\{R_1, R_3\}$	0.85	$\{R_2, R_3, R_4\}$	0.30
$\{R_1, R_4\}$	0.30	X	0.50

From Table III, one can see that the rule 2 has the resisting effect on other rules and the rule 3 has the enhanced effect on the other rules. $\mu(\{R_2\}) = \mu(\{R_3\}) = 0.3$ indicates that the rules 2 and 3 have the same individual impact on the consequent.

Let $\text{SM}_1 = (0.6\ 0.0\ 0.7\ 0.8)$, and $\text{SM}_2 = (0.6\ 0.7\ 0.0\ 0.8)$. We use the Choquet integral to compute the two values, which are regarded as the degrees of truth of consequent C1. The computed result is that $(C)\int \text{SM}_1 d\mu = 0.645$ and $(C)\int \text{SM}_2 d\mu = 0.155$. From rule 5, we know the degree of truth of either e1 or e2 belonging to C2 is 0.5. The above computed result shows that e1 corresponds to the consequent C1 and e2 to C2. This inferred result is correct.

The handling of interaction in FPR reasoning can lead to a reduction of the likely occurrence of an undesirable consequent or an improvement of reasoning accuracy. The set function must be determined before the Choquet integral is applied. However, in many real applications, the nonnegative nonadditive set function representing the interaction among the rules is very difficult to acquire. Usually it is given by domain experts. In the next section, we will discuss how to learn the set function from data.

IV. Learning Nonnegative Nonadditive-Additive Set Functions

We need to compute the integral values while handling the interaction among the rules. Before the integral values can be computed, the fuzzy measure or the nonadditive set function must be determined. This is a very difficult task. Some methodologies for determining fuzzy measures have been developed and generally discussed in [8], [23], [13], [24], [25], and [3]. In this section, with respect to our particular issues of handling interaction among rules, we make an attempt to determine the nonnegative set function by solving a linear programming problem.

Let us consider a classification problem with two classes and N training examples. Suppose that, using some learning techniques, we have already extracted M fuzzy rules from the N examples. The FPR form is IF (attribute-values) THEN (class), where the attribute-values are the intersection of a number of fuzzy subsets and the class is either C1 or C2.

Due to fuzziness and interaction among rules, the reasoning accuracy of the M rules to the N examples fails to attain 100%. That is, the extracted M rules may not entirely cover the N examples. The M rules are categorized in two groups, S_1 and S_2, one leading to the consequent C1 and the other leading to C2.

Let $S_1 = \{R_i, i = 1, 2, \ldots, m\}$ and $S_2 = \{R_i, i = m + 1, m + 2, \ldots, M\}$.

Moreover, the N examples are also classified into two parts, as follows:

$$T_1 = \{e_i, i = 1, 2, \ldots, n\}$$
$$T_2 = \{e_i, i = n + 1, n + 2, \ldots, N\}.$$

The actual classification of the examples within T_1 is C1, and within T_2 is C2.

By using the matching mechanism given in (1)–(3) where the global weights are ignored, we match each example e_i against both S_1 and S_2. The matching leads to two values x_{i1} and x_{i2} (the degrees of truth of e_i belonging to C1 and C2 respectively). Noting the definitions of S_1, S_2, T_1, and T_2, we hope that the following inequalities hold:

$$x_{i1} > x_{i2}, \quad \text{for } i = 1, 2, \ldots, n$$
$$x_{i1} < x_{i2}, \quad \text{for } i = n + 1, \quad n + 2, \ldots, N. \tag{6}$$

Due to the reasoning mechanism and the existence of interaction among rules, the above inequalities generally fails to be valid for all $i, (i = 1, 2, \ldots, N)$. There may be several reasons for this. One of them is the interaction. Since the interaction is reflected in the fuzzy measures (set functions), which are generally unknown, we may numerically determine them by using the optimization criterion of reasoning accuracy.

Let μ_1, μ_2 be two set functions defined on $S_1 = \{R_i, i = 1, 2, \ldots, m\}$ and $S_2 = \{R_i, i = m+1, m+2, \ldots, M\}$ respectively, subject to $0 \le \mu_1, \mu_2 \le 1$. Suppose that matching degree functions of e_i matching S_1 and S_2 are

$$f_{i1} = \left(\mathrm{SM}_i^{(1)}, \mathrm{SM}_i^{(2)}, \ldots, \mathrm{SM}_i^{(m)}\right)$$
$$f_{i2} = \left(\mathrm{SM}_i^{(m+1)}, \mathrm{SM}_i^{(m+2)}, \ldots, \mathrm{SM}_i^{(M)}\right) \tag{7}$$

where $\mathrm{SM}_i^{(j)}$ is the result of e_i matching the jth rule $R_j (i = 1, 2, \ldots, N; j = 1, 2, \ldots, M)$.

Then

$$x_{i1} = (C)\int_{S_1} f_{i1} d\mu_1 \quad \text{and } x_{i2} = (C)\int_{S_2} f_{i2} d\mu_2.$$

According to (6), we hope the following inequalities hold:

$$(C)\int_{S_1} f_{i1} d\mu_1 > (C)\int_{S_2} f_{i2} d\mu_2, \quad \text{for } i = 1, 2, \ldots, n \tag{8}$$

$$(C)\int_{S_1} f_{i1} d\mu_1 < (C)\int_{S_2} f_{i2} d\mu_2, \quad \text{for } i = n + 1, n + 2, \ldots, N \tag{9}$$

subject to $0 \le \mu_1, \mu_2 \le 1$.

Noting that $\mu_1(\phi) = 0$ and $\mu_2(\phi) = 0$, we can define a $(2^m + 2^{M-m} - 2)$-dimensional vector (representing all coefficients of the two nonadditive set functions) as follows:

$$\begin{aligned}(x_1, x_2, \ldots, x_K) = (&\mu_1(\{R_1\}), \mu_1(\{R_2\}), \ldots,\\ &\mu_1(\{R_m\}), \mu_1(\{R_1, R_2\}), \ldots, \mu_1(\{R_1, R_2, \cdots R_m\}),\\ &\mu_2(\{R_{m+1}\}), \mu_2(\{R_{m+2}\}), \ldots, \mu_2(\{R_M\}), \ldots,\\ &\mu_2(\{R_{m+1}, R_{m+2}\}), \cdots \mu_2(\{R_{m+1}, R_{m+2}, \ldots, R_M\}))\end{aligned} \tag{10}$$

where $K = 2^m + 2^{M-m} - 2$.

From Definition 2, we know that Choquet integrals given in inequalities (8) and (9) can be represented as linear combinations of $(x_1, x_2, \ldots, x_K)$. That is, we can rewrite the inequalities (8) and (9) as follows:

$$\sum_{j=1}^{K} a_{ij} x_j > 0, \quad i = 1, 2, \ldots, \quad n, n + 1, \ldots, N \tag{11}$$

$$\text{Subject to } 0 \le x_j \le 1, \quad j = 1, 2, \ldots, K \tag{12}$$

where $a_{ij} (i = 1, 2, \ldots, n, n + 1, \ldots, N; j = 1, 2, \ldots, K)$ are determined in terms of the values of $f_{i1} = (\mathrm{SM}_i^{(1)}, \mathrm{SM}_i^{(2)}, \ldots, \mathrm{SM}_i^{(m)}), f_{i2} = (\mathrm{SM}_i^{(m+1)}, \mathrm{SM}_i^{(m+2)}, \ldots, \mathrm{SM}_i^{(M)})$ and the definition of the Choquet integral.

Usually the problem of solving the system of inequalities (11) and (12) is transformed into the following linear programming problem:

$$\text{Minimize } (\xi_1 + \xi_2 + \cdots + \xi_N) \tag{13}$$

$$\text{Subject to } \sum_{j=1}^{K} a_{ij} x_j + \xi_i > 0 \quad i = 1, 2, \ldots, N \tag{14}$$

$$0 \le x_j \le 1, \quad j = 1, 2, \ldots, K \tag{15}$$

$$0 \le \xi_i, \quad i = 1, 2, \ldots, N. \tag{16}$$

The transformation is based on such a fact that, if the solution to linear programming problem (13)–(16) makes $(\xi_1 + \xi_2 + \cdots + \xi_N)$ zero, then inequalities (11) and (12) hold well. It is our desirable result. Generally, the minimum value of $(\xi_1 + \xi_2 + \cdots + \xi_N)$ may not be zero. In this case. we consider the solution of (13)–(16) as an approximate solution of inequalities (11) and (12).

In [7], [11], the authors selected the square error criterion to solve inequalities (8) and (9). By minimizing the quadratic error between the expected output and the actual output of a classifier, they transferred the problem of solving inequalities (8) and (9) into a quadratic program denoted as follows:

$$\begin{cases} \text{Minimize} & \frac{1}{2}x^T Dx + \Gamma^T x \\ \text{Subject to} & Ax + b \geq 0, \end{cases} \tag{17}$$

where the $(2^m + 2^{M-m} - 2)$-dimensional vector x is defined by (10), Γ and b are two constant vectors, and D and A are two constant matrices. Our initial experimental results show that the learning accuracy from solving quadratic program (17) is better than the learning accuracy of linear programming (13)–(16) but the computational cost of solving (17) is much higher than that of (13)–(16).

One may speculate as to whether learning so many $(2^m + 2^{M-m} - 2)$ parameters will lead to over-learning (over-fitting); in other words, the training accuracy is extra-high but the testing accuracy is lowered. The experiments in this section (Table IV) do not show an over-learning phenomenon.

If the interaction among rules is not considered, but only the degree of importance of each rule (i.e., the global weight for each rule) then the system of inequalities given in (6) can be further simplified. The task will be reduced from determining two set functions with $(2^m + 2^{M-m} - 2)$ parameters to finding a global weight vector with M parameters.

From (2) and (7), we have

$$x_{i1} = w_1 \mathrm{SM}_i^{(1)} + w_2 \mathrm{SM}_i^{(2)} + \cdots + w_m \mathrm{SM}_i^{(m)}$$
$$x_{i2} = w_{m+1} \mathrm{SM}_i^{(m+1)} + \cdots + w_M \mathrm{SM}_i^{(M)}$$

subject to $0 \leq w_j \leq 1, j = 1, 2, \ldots, M$.

According to (6), we hope the following inequalities hold:

$$\sum_{j=1}^{M} b_{ij} w_j > 0, \quad i = 1, 2, \ldots, N \tag{18}$$

where

$$b_{ij} = \begin{cases} \mathrm{SM}_i^{(j)}, & 1 \leq i \leq n, \quad 1 \leq j \leq m \\ -\mathrm{SM}_i^{(j)}, & 1 \leq i \leq n, \quad m+1 \leq j \leq M \\ -\mathrm{SM}_i^{(j)}, & n+1 \leq i \leq N, \quad 1 \leq j \leq m \\ \mathrm{SM}_i^{(j)}, & n+1 \leq i \leq N, \quad m+1 \leq j \leq M \end{cases}$$
$$\text{subject to } 0 \leq w_j \leq 1, \quad j = 1, 2, \ldots, M$$
$$\sum_{j=1}^{m} w_j = 1; \sum_{j=m+1}^{M} w_j = 1. \tag{19}$$

Therefore, by the introduction of relaxed variables $\xi_1, \xi_2, \ldots, \xi_N$, we can obtain from (18) and (19) the following linear programming problem:

$$\text{Minimize } (\xi_1 + \xi_2 + \cdots + \xi_N) \tag{20}$$
$$\text{Subject to } \sum_{j=1}^{M} b_{ij} w_j + \xi_i > 0, \quad i = 1, 2, \ldots, N$$
$$\sum_{j=1}^{m} w_j = 1, \sum_{j=m+1}^{M} w_j = 1$$
$$0 \leq w_j \leq 1 (j = 1, 2, \ldots, M)$$
$$0 \leq \xi_i (i = 1, 2, \ldots, N). \tag{21}$$

A. Experiments

The seven databases employed for experiments are obtained from various sources. Their features are briefly described below.

1) *Rice taste data:* This database was used by Nozaki [17] to verify a simple and powerful algorithm for fuzzy rule generation. It contains 105 cases with five numerical attributes. The classification attribute is continuous. In our experiments, cases are categorized into two classes according to positive and negative values of the classification attribute.

2) *Mango leaf data:* This set of data was used by Pal [18] to investigate automatic feature extraction based on fuzzy techniques. It provides information on different kinds of mango-leaf with 18 numerical attributes for 166 patterns (cases). It has three classes representing three kinds of mango. We consider the first class as the positive and the other two classes as the negative in our experiments.

3) *Thyroid gland data [21]:* This set of data contains 215 cases of three different kinds of thyroid gland. Each case consists of five numerical attributes. We consider the first class as the positive and the other two classes as the negative.

4) *Pima India diabetes data [21]:* This database contains 768 cases related to the diagnosis of diabetes (268 positive and 500 negative). It has eight numerical attributes.

5) *Glass Identification Database [21]:* This database has 214 instances related to seven classes of glass. Each instance has nine numerical attributes. In our experiments, we consider the first class as the positive and the other six classes as the negative.

6) *Auto-Mpg Data [21]:* This database has 398 instances and nine attributes. In our experiments, we only use the five numerical attributes and one integer-valued attribute. The mpg attribute is regarded as the class attribute. Moreover, since the attribute horsepower has six missing values, we only use 392 of 398 instances.

7) *Sonar Database [21]:* This database contains 208 patterns, 111 patterns belonging to metal class and 97 patterns belonging to rock class. Each pattern is a set of 60 numbers in the range 0.0–1.0.

We conduct our experiments as follows. Each database is randomly split into two parts. One part includes 90% of the cases in the database. This part is used for training while the remaining 10% is used for testing. The purpose of training is to extract a number of FPRs from the training set.

A fuzzy decision tree algorithm (Fuzzy ID3) is used to generate fuzzy rules for each selected training set. Since we need to extract FUZZY production rules from the training set, the C4.5 programs are not suitable for our task. A fuzzification for numerical attributes should be conducted before generating fuzzy decision trees. Here we opt for a fuzzy clustering algorithm based on Kohonen self-organized mapping [14] to generate cluster centers. By applying this algorithm to each selected database, we

TABLE IV
REASONING ACCURACY OF THREE METHODS WHERE THE FIRST IS THE FREQUENTLY USED (MAX-MIN) METHOD, THE SECOND IS THE GLOBALLY WEIGHTED FUZZY PRODUCTION RULE REASONING, AND THE THIRD IS THE PROPOSED FUZZY REASONING WITH HANDLING INTERACTION AMONG RULES

Databases	First training / testing accuracy	Second training / testing accuracy	Third training / testing accuracy
Rice taste data	0.9225 / 0.9007	0.9455 / 0.9188	0.9815 / 0.9808
Mango leaf data	0.8221 / 0.8178	0.8534 / 0.8384	0.9145 / 0.8838
Thyroid gland data	0.8907 / 0.8729	0.9256 / 0.9294	0.9667 / 0.9561
Auto-mpg data	0.8849 / 0.8910	0.9001 / 0.8895	0.9096 / 0.8975
Sonar signal data	0.8048 / 0.7193	0.8184 / 0.7217	0.9140 / 0.7839
Pima diabetes data	0.7416 / 0.7121	0.7577 / 0.7342	0.8531 / 0.7784
Glass identification	0.8041 / 0.7409	0.8741 / 0.7837	0.8860 / 0.7866

can fuzzify the numerical attributes where the number of linguistic terms for each attribute is assumed to be three. The membership function is selected to be normal, that is, each membership function f has the following form:

$$f(x) = \begin{cases} e^{-\frac{(x-c)^2}{r^2}}, & x > c \\ e^{-\frac{(x-c)^2}{l^2}}, & x \le c \end{cases} \tag{22}$$

where c is the center determined by Kohonen self-organized clustering algorithm, and r, l are determined by solving

$$e^{-\frac{(x-c_1)^2}{r^2}}\bigg|_{x=\frac{c_1+c_2}{2}} = e^{-\frac{(c_2-x)^2}{l^2}}\bigg|_{x=\frac{c_1+c_2}{2}} = \frac{1}{2}$$

for any two adjacent centers c_1 and $c_2 (c_1 < c_2)$. It results in $r = l = (c_2 - c_1)/(2\sqrt{\ln 2})$.

Suppose there are m cluster centers $c_1 < c_2 < \cdots < c_m (m \ge 2)$. Substituting $(c_j - c_{j-1})/(2\sqrt{\ln 2})$ for r and l in (22), we have m normal membership functions $f_1, f_2, \ldots, f_m$ expressed as

$$f_1(x) = \exp\left(-\frac{4\ln 2 \cdot (x-c_1)^2}{(c_2-c_1)^2}\right)$$

$$f_j(x) = \begin{cases} \exp\left(-\frac{4\ln 2\cdot(x-c_j)^2}{(c_j-c_{j-1})^2}\right), & x < c_j \\ \exp\left(-\frac{4\ln 2\cdot(x-c_j)^2}{(c_{j+1}-c_j)^2}\right), & x > c_j \end{cases}$$

$$j = 2, 3, \ldots, m-1$$

and

$$f_m(x) = \exp\left(-\frac{4\ln 2 \cdot (x-c_{m-1})^2}{(c_m-c_{m-1})^2}\right).$$

After fuzzification, we use the Fuzzy ID3 algorithm to generate a set of FPRs. Matching both the training set and the testing set against the extracted FPRs, we obtain the initial training accuracy and the initial testing accuracy. Here the reasoning mechanism is based on the frequently used operators (min, max).

Subsequently, we consider the degree of importance of individual rules (i.e., the global weights of individual rules), but we do not consider the interaction among rules. By solving the linear programming indicated by (20)–(21) to acquire the values of global weights, the FPRs extracted from each training set are revised to be globally WFPRs. Matching both the training set and the testing set against the globally WFPRs, we obtain the second training accuracy and the second testing accuracy. Here the reasoning mechanism is the globally weighted fuzzy reasoning algorithm given in Section II.

Finally, we consider the interaction among the extracted fuzzy rules. By solving the linear programming indicated by (13)–(16), we can acquire the two nonadditive set functions that indicate the interaction among the rules. The initial FPRs are now revised to be the two sets of FPRs with two nonadditive nonnegative set functions that indicate the interaction among the rules. By matching both the training set and the testing set against the set of fuzzy rules with two set functions, we obtain the third training accuracy and the third testing accuracy. Here, the reasoning mechanism is given by (c1)–(c3) in Section III.

The experiments are conducted repeatedly five times and their averaged results are given in Table IV. Analyzing Table IV, we can summarize the following experimental conclusions.

1) To a certain degree, the learning accuracy, i.e., the training and testing accuracy, of the seven selected databases improved both from method 1 to method 2 and from method 2 to method 3. The amount of accuracy-improvement depends on the concrete structure of databases.
2) Of the seven databases, the smallest improvement in learning-accuracy was in the Auto-mpg data. For this kind of database, both feature weight assignment and handling interaction among rules are not necessary. This means that there are almost no salient features and interactive rules.
3) The Sonar signal data and the Pima diabetes data show a very small improvement in learning-accuracy from method 1 to method 2 but a very significant improvement from method 2 to method 3. This implies that the rules extracted from the databases have a very strong interactive effect but that the effect of feature weight assignment is not important. In this situation the handling of interaction among rules seems to be extremely important.
4) Glass identification data shows that learning-accuracy very significantly improved from method 1 to method 2 but improved very little from method 2 to method 3. This implies that the rules extracted from the databases have little interactive effect but that the effect of feature weight

assignment is very important. In this situation the handling of interaction among rules can be replaced with the feature weight assignment.

5) Given that the difference between the training accuracy and testing accuracy of method 3 is not significant, it would seem that the proposed method 3 does not generate an overlearning phenomenon.

We now briefly illustrate how the set function could affect (positively or negatively) the interpretability of the fuzzy system. For the Sonar signal data, we find the reasoning accuracy of extracted fuzzy rules not very good. Of course, it depends on the data structure. By learning global weights which are assigned to the fuzzy rules, we expect to improve the reasoning accuracy. It indeed results in an accuracy improvement. However, the improvement is not significant (see Table IV). In this situation, we think that interaction exists among the rules having the same consequent and then learn the interaction expressed by a set function. We find that rules 2, 3, and 5 have the weight values 0.081, 0.219, and 0.116, respectively, but the set function take values 0.925 and 0.883 on $\{\mathrm{rule2}, \mathrm{rule3}\}$ and $\{\mathrm{rule3}, \mathrm{rule5}\}$ respectively. The two inequalities, i.e., $0.925 > 0.081 + 0.219$ and $0.883 > 0.219 + 0.116$, imply that rules 2, 3, and 5 have the very strong interaction which, for leading to Class1, is positive.

V. Conclusions and Remarks

When domain experts provide additional information for a set of fuzzy rules, there exists interaction among the set of fuzzy rules. This paper models the interaction by using a nonlinear integral with respect to a nonadditive set function that is a formulation of the additional information of domain experts. The interaction is handled by extending a vector of global weights to a nonadditive set function. As a result of the handling of interaction among the rules, the globally WFPR reasoning is extended to the nonlinear integral model, reasoning accuracy is improved, and the occurrence of undesirable reasoning consequents is reduced. This paper also discussed how to learn the nonadditive set function from data while the set function indicating interaction is hard to be given by domain experts. It demonstrates that the learning of nonadditive set functions can be achieved by solving a linear programming problem. We have the following remarks.

1) In the process of learning nonadditive set functions, the procedure is given by solving a linear programming problem. This means that little computational effort is added in the learning phase.

2) Our algorithms for determining the set function or the vector of global weights are derived in terms of a two-cluster classification problem, but they can easily be extended to the situation of classification problems of more than two clusters.

3) The FPR reasoning with handling interaction proposed in this paper is suitable for both real and nominal attributes.

4) In this paper, we use Choquet integrals with respect to a nonadditive set function to evaluate the value of overall similarity and, based on the overall similarity, the reasoning is conducted. The reasoning results together with the learned values of the nonadditive set function are dependent on the selection of nonlinear integrals. However, initial experiments show that the reasoning results are not sensitive to the selection of nonlinear integrals.

5) The consequent of an IF-THEN rule in this paper is assumed to be a crisp classification. Without much difficulty, it can be extended to the case that the consequent is a fuzzy set. For the extension, we briefly give an explanation. For a two-class problem, the crisp classification means that each case should be assigned a Boolean vector (0, 1) or (1, 0), but the fuzzy classification means that each case corresponds to a fuzzy vector (a, b) where $0 < \mathrm{a}, \mathrm{b} < 1$. In fuzzy case, the fundamental inequalities given in (6) are established according to $\mathrm{a} > \mathrm{b}$ or $\mathrm{b} > \mathrm{a}$. Similarly, the linear programming (13)–(16) can be derived.

6) The frequently used min-max reasoning mechanism is a special case of a globally weighted fuzzy reasoning scheme, but the globally weighted fuzzy reasoning is a special case of our proposed reasoning model with handling interaction among rules.

7) The nonadditive set function discussed in this paper can be regarded as a set of knowledge representation parameter. The refinement of these parameters, which is a process of solving a linear programming problem, is independent of the initial training algorithms.

8) The reasoning in this paper essentially belongs to similarity-based reasoning. The result of given facts matching against antecedents of rules is dependent on the selection of similarity measure.

9) The main problem of the proposed approach to interaction handling is too many parameters when the number of rules increases. At this stage, the approach is not yet appropriate for many rules due to the exponential complexity. As a further investigated topic, the authors would like to address how to effectively represent the interaction among the rules (e.g., fuzzy measures with some restricts or with some particular structure) such that the representation parameters can be reduced greatly.

References

[1] B. G. Buchanan and E. H. Shortliffe, *Rule-Based Expert Systems, The MYCIN Experiments of the Stanford Heuristic Programming Projects*. Reading, MA: Addison-Wesley, 1984.

[2] S. M. Chen, J. S. Ke, and J. F. Chang, "Knowledge representation using fuzzy Petri nets," *IEEE Trans. Knowledge Data Eng.*, vol. 2, pp. 311–319, Sept. 1990.

[3] T.-Y. Chen, J.-C. Wang, and G.-H. Tzeng, "Identification of general fuzzy measures by genetic algorithms based on partial information," *IEEE Trans. Syst., Man, Cybern. B*, vol. 30, pp. 517–528, Aug. 2000.

[4] S. M. Chen, "Weighted fuzzy reasoning using weighted fuzzy Petri nets," *IEEE Trans. Knowledge Data Eng.*, vol. 14, pp. 386–397, Feb. 2002.

[5] D. Denneberg and M. Grabisch, "Interaction transform of set functions over a finite set," *Inform. Sci.*, vol. 121, pp. 149–170, 1999.

[6] *The Handbooks of Fuzzy Sets Vol 7: Fundamentals of Fuzzy Sets*, D. Dubois and H. Prade, Eds., Kluwer, Boston, MA, 2000.

[7] M. Grabisch and J. M. Nicols, "Classification by fuzzy integrals—Performance and tests," *Fuzzy Sets Syst.*, vol. 65, pp. 255–271, 1994.

[8] M. Grabisch, "The representation of importance and interaction of features by fuzzy measures," *Pattern Recognit. Lett.*, vol. 17, pp. 567–575, 1996.

[9] ——, "The application of fuzzy integrals in multi-criteria decision making," *Eur. J. Oper. Res.*, vol. 89, pp. 445–456, 1996.

[10] ——, "Fuzzy integral for classification and feature extraction," in *Fuzzy Measures and Integrals*, M. Grabisch, T. Murofushi, and M. Sugeno, Eds. Berlin, Germany: Physica-Verlag, 2000, pp. 416–434.
[11] M. Grabisch, C. Labreuche, and J. Vansnick, "On the extension of Pseudo-Boolean functions for the aggregation of interacting bipolar criteria," *Eur. J. Oper. Res.*, vol. 148, pp. 28–47, 2003.
[12] L. K. Hyung, Y. S. Song, and K. M. Lee, "Similarity measure between sets and between elements," *Fuzzy Sets Syst.*, vol. 62, pp. 291–293, 1994.
[13] G. J. Klir, W. Zhenyuan, and D. Harmanec, "Constructing fuzzy measures in expert systems," *Fuzzy Sets Syst.*, vol. 92, no. 2, pp. 251–264, 1997.
[14] T. Kohonen, *Self-Organization and Associate Memory*. Berlin, Germany: Springer, 1988.
[15] J. L. Marichal, "An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria," *IEEE Trans. Fuzzy Syst.*, vol. 8, pp. 800–807, Dec. 2000.
[16] ——, "An axiomatic approach of the discrete Sugeno integral as a tool to aggregate interacting criteria in qualitative framework," *IEEE Trans. Fuzzy Syst.*, vol. 9, pp. 164–172, Feb. 2001.
[17] K. Nozaki, H. Ishibuchi, and H. Tanaka, "A simple but powerful heuristic method for generating fuzzy rules from numerical data," *Fuzzy Sets Syst.*, vol. 86, pp. 251–270, 1997.
[18] S. K. Pal, "Fuzzy set theoretic measures for automatic feature evaluation," *Inform. Sci.*, vol. 64, pp. 165–179, 1992.
[19] C. P. Pappis and N. I. Karacapilidis, "A comparative assessment of measures of similarity of fuzzy sets," *Fuzzy Sets Syst.*, vol. 56, pp. 171–174, 1993.
[20] I. B. Turksen and Z. Zhong, "An approximate analogical reasoning approach based on similarity measures," *IEEE Trans. Syst., Man, Cybern.*, vol. 18, pp. 1049–1056, Nov./Dec. 1988.
[21] UCI Repository of Machine Learning Databases and Domain Theories [Online]. Available: ftp://ftp.ics.uci.edu/pub/machine-learning-databases
[22] Z. Wang and G. J. Klir, *Fuzzy Measure Theory*. New York: Plenum, 1992.
[23] J. Wang and Z. Wang, "Using neural networks to determine Sugeno measures by statistics," *Neural Networks*, vol. 10, no. 1, pp. 183–195, 1997.
[24] W. Wang, Z. Wang, and G. J. Klir, "Genetic algorithms for determining fuzzy measure from data," *J. Intell. Fuzzy Syst.*, vol. 6, no. 2, pp. 171–183, 1998.
[25] Z. Wang, K.-S. Leung, and J. Wang, "A genetic algorithm for determining nonadditive set functions in information fusion," *Fuzzy Sets Syst.*, vol. 102, no. 3, pp. 463–469, 1999.
[26] X. Z. Wang, E. C. C. Tsang, and D. S. Yeung, "A comparative study on heuristic algorithms for generating fuzzy decision trees," *IEEE Trans. Syst., Man, Cybern. B*, vol. 31, pp. 215–226, Apr. 2001.
[27] D. S. Yeung and E. C. C. Tsang, "Weighted fuzzy production rules," *Fuzzy Sets Syst.*, vol. 88, pp. 299–313, 1997.
[28] ——, "A comparative study on similarity-based fuzzy reasoning methods," *IEEE Trans. Syst., Man, Cybern.*, vol. 27, pp. 216–226, Apr. 1997.
[29] ——, "A multilevel weighted fuzzy reasoning algorithm for expert systems," *IEEE Trans. Syst., Man, Cybern.*, vol. 28, pp. 149–158, Apr. 1998.
[30] Y. Yuan and M. J. Shaw, "Induction of fuzzy decision trees," *Fuzzy Sets Syst.*, vol. 69, pp. 125–139, 1995.

Daniel S. Yeung (F'04) received the Ph.D. degree in applied mathematics from Case Western Reserve University in 1974.

In the past, he has been an Assistant Professor of Mathematics and Computer Science at Rochester Institute of Technology, Rochester, NY, a Research Scientist with General Electric Corporate Research Center, and as a System Integration Engineer at TRW. He was the Chairman of the Department of Computing at The Hong Kong Polytechnic University, Hong Kong, for seven years, where he is currently a Chair Professor. His research interests include neural-network sensitivity analysis, data mining, Chinese computing, and fuzzy systems.

Dr. Yeung was the President of IEEE Hong Kong Computer Chapter (91 and 92), an Associate Editor for both IEEE TRANSACTIONS ON NEURAL NETWORKS and IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS–PART B: CYBERNETICS. He was the General Co-Chair of the 2002 and 2003 International Conference on Machine Learning and Cybernetics. He has recently been elected by the IEEE SMC Society as Vice President for Technical Activities.

Xi-Zhao Wang (SM'04) received the B.Sc. and M.Sc. degrees in mathematics from Hebei University, Baoding, China, in 1983 and 1992, respectively, and the Ph.D. degree in computer science from Harbin Institute of Technology, Harbin, China, in 1998.

From 1983 to 1998, he worked as a Lecturer, an Associate Professor, and a Full Professor in the Department of Mathematics, Hebei University. From 1998 to 2001, he worked as a Research Fellow at the Department of Computing, Hong Kong Polytechnic University, Kowloon. Since 2001, he has been the Dean and Professor of the Faculty of Mathematics and Computer Science, Hebei University. His main research interests include inductive learning with fuzzy representation, fuzzy measures and integrals, neuro-fuzzy systems and genetic algorithms, feature extraction, multiclassifier fusion, and applications of machine learning. He was the General Co-Chair of the 2002 and 2003 International Conferences on Machine Learning and Cybernetics.

Eric C. C. Tsang received the B.Sc. degree in computer studies from the City University of Hong Kong in 1990 and the Ph.D. degree in computing from the Hong Kong Polytechnic University, Kowloon, in 1996.

He is an Assistant Professor with the Department of Computing, Hong Kong Polytechnic University. His main research interests are in the area of fuzzy expert systems, fuzzy neural networks, machine learning, genetic algorithm and fuzzy support vector machine.