Classification of Heterogeneous Fuzzy Data by Choquet Integral With Fuzzy-Valued Integrand

Rong Yang, Zhenyuan Wang, Pheng-Ann Heng, *Member, IEEE*, and Kwong-Sak Leung, *Senior Member, IEEE*

Abstract—As a fuzzification of the Choquet integral, the defuzzified choquet integral with fuzzy-valued integrand (DCIFI) takes a fuzzy-valued integrand and gives a crisp-valued integration result. In this paper, the DCIFI acts as a projection to project high-dimensional heterogeneous fuzzy data to one-dimensional crisp data to handle the classification problems involving different data forms, such as crisp data, interval values, fuzzy numbers, and linguistic variables, simultaneously. The nonadditivity of the signed fuzzy measure applied in the DCIFI can represent the interaction among the measurements of features towards the discrimination of classes. Values of the signed fuzzy measure in the DCIFI are considered to be unknown parameters which should be learned before the classifier is used to classify new data. We have implemented a genetic algorithm (GA)-based adaptive classifier-learning algorithm to optimally learn the signed fuzzy measure values and the classified boundaries simultaneously. The performance of our algorithm has been tested both on synthetic and real data. The experimental results are satisfactory and outperform those of existing methods, such as the fuzzy decision trees and the fuzzy-neuro networks.

Index Terms—Choquet integral, classification, data mining, fuzzification, heterogeneous fuzzy data, signed fuzzy measure.

I. INTRODUCTION

HETEROGENEOUS fuzzy data are ubiquitous in practical databases. They are represented in diverse data forms, such as crisp data, fuzzy numbers, interval values, linguistic variables, or vectors with components in any combination of the aforementioned forms. To handle these heterogeneous fuzzy data efficiently and effectively is an essential task to data engineers. Currently, there are mainly two approaches to deal with them. One is to preprocess the heterogeneous fuzzy data so that they can be expressed by a set of real numbers and managed by traditional data mining methods for crisp data [8], [10], [18]. However, this indirect method is restricted to those applications where the heterogeneous data can be expressed simply by several real numbers. Another approach utilizes a linear parametric

model to handle the heterogeneous data directly [20]. Obviously, it is insufficient for systems having intrinsic nonlinear correlations. Due to the aforementioned reasons, a comprehensive model, which can not only handle the heterogeneous fuzzy data directly, but also elicit the nonlinear relationships among features, is expected.

The Choquet integral [4], [5], [26], [24] with respect to fuzzy measure or signed fuzzy measure [17], [19] has performed successfully as a nonlinear aggregation tool in information fusing and data mining for crisp data-bases. The nonadditivity of the signed fuzzy measure provides an effective representation to describe the interaction among the contributions from the predictive attributes to the objective attribute. Recently, the Choquet integral has been applied to classification problem [6], [7], [16] and has obtained quite encouraging results. In [29], a weighted Choquet integral with respect to signed fuzzy measure has been used as a projection tool to project high-dimensional crisp data to a virtual variable on a real axis so that the classification in high-dimensional space is simplified to that in one-dimensional space.

In this paper, the original Choquet integral of a real-valued function defined on a set of attributes with respect to a signed fuzzy measure is generalized for allowing a fuzzy-valued function to be its integrand. This fuzzification of Choquet integral is called the defuzzified Choquet integral with fuzzy-valued integrand (DCIFI) as it has a fuzzy valued integrand but a non-fuzzy valued integration result. A numerical method with the relevant algorithm is developed to estimate the integration values of the DCIFI. The nonadditivity property of the signed fuzzy measure of the DCIFI describes the respective contributions from the feature attributes towards the discrimination. Thus, the DCIFI is considered to be a more powerful and flexible tool to classify data, especially the data involving heterogeneous fuzzy data. It can project high-dimensional heterogeneous fuzzy data of different classes to their corresponding virtual variables on a real axis. Similarly, the class boundaries, which separate different classes in high-dimensional space, can also be projected on the real axis as a virtual boundary. By these schemes, the classification problem for high-dimensional heterogeneous fuzzy data is simplified to that for one-dimensional crisp data. A DCIFI projection classifier is identified by the values of the signed fuzzy measure, which are regarded as the unknown parameters. They are optimally determined according to a set of training data through a GA-based adaptive classifier-learning algorithm (GACA). The optimization is to project all heterogeneous fuzzy data onto a real axis of virtual variables with the best classifying boundaries such that the total misclassification rate is minimized.

R. Yang is with the Department of Automatic Science, the College of Engineering and Technology, Shen Zhen University, Nan Shan, Shen Zhen, P. R. China (e-mail: ryang@szu.edu.cn).

Z. Wang is with the Department of Mathematics, the University of Nebraska, Omaha, NE 68182 USA (e-mail: zhenyuanwang@mail.unomaha.edu).

P.-A. Heng and K.-S. Leung are with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong (e-mail: pheng@cse.cuhk.edu.hk; ksleung@cse.cuhk.edu.hk).
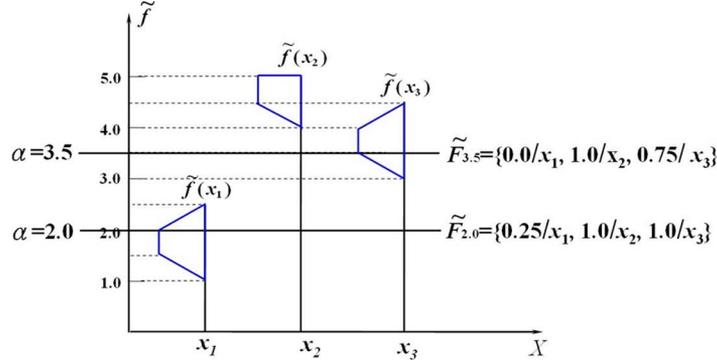
Fig. 1. α-cut of a real-valued function.

The paper is organized as follows. In Section II, relevant fundamental concepts on original Choquet integral are summarized. The defuzzified Choquet integral with fuzzy-valued integrand is derived in Section III, where a numerical method with the relevant algorithm is developed to estimate the integration value of the DCIFI. The DCIFI projection classifier is illustrated in Section IV with the GACA adopted for optimally determining unknown parameters from the training data sets. Both experiments on synthetic and real data are conducted to evaluate the performance of the DCIFI projection classifier. These experimental results are shown in Section VI. Finally, conclusions are given in Section VII.

II. Choquet Integral

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a nonempty finite set of feature attributes and $\mathcal{P}(X)$ be the power set of X. A signed fuzzy measure [25], μ, is a mapping from $\mathcal{P}(X)$ to $(-\infty, \infty)$ satisfying $\mu(\oslash) = 0$. The set function μ is nonadditive in general, so it is also called nonadditive set function. A fuzzy measure is considered to be a specialization of signed fuzzy measure with additional monotonic property, that is

$$\mu(A) \leq \mu(B) \, \forall A \subset B$$

on $\mathcal{P}(X)$. In this paper, we always assume μ as a signed fuzzy measure. To be convenient, $\mu(\{x_{j_1}, x_{j_2}, \ldots, x_{j_m}\})$ is denoted by μ_j, where $j = \sum_{k=1}^{m} 2^{j_k-1}$ and $\{j_1, j_2, \ldots, j_m\}$ is a subset of $\{1, 2, \ldots, n\}$. More explicitly, $\mu_0 = \mu(\oslash) = 0$, $\mu_1 = \mu(\{x_1\})$, $\mu_2 = \mu(\{x_2\})$, $\mu_3 = \mu(\{x_1, x_2\})$, $\mu_4 = \mu(\{x_3\})$, $\mu_5 = \mu(\{x_1, x_3\})$, $\mu_6 = \mu(\{x_2, x_3\})$, $\mu_7 = \mu(\{x_1, x_2, x_3\})$, ..., and $\mu_{2^n-1} = \mu(\{x_1, x_2, \ldots, x_n\})$.

Definition 2.1: Let $f : X \to (-\infty, \infty)$ be a real-valued function. The Choquet integral of f is defined as

$$\int f d\mu = \int_{-\infty}^{0} [\mu(F_\alpha) - \mu(X)] d\alpha + \int_{0}^{\infty} \mu(F_\alpha) d\alpha$$

where $F_\alpha = \{x | f(x) \geq \alpha\}$, for any $\alpha \in (-\infty, \infty)$, is the α-cut of f, represented as a crisp set of X.

For example, let $X = \{x_1, x_2, x_3\}$, and a real-valued function f is defined on X by $f(x_1) = 2.0$, $f(x_2) = 1.0$, $f(x_3) = 3.0$, then the α-cuts of f at $\alpha = 0.5$, 1.5, and 2.5 are crisp sets of X, described by $F_{0.5} = \{x_1, x_2, x_3\}$, $F_{1.5} = \{x_1, x_3\}$, and $F_{2.5} = \{x_3\}$ respectively, as shown in Fig. 1.

To calculate the value of the Choquet integral of a given real-valued function f, usually the values of f, i.e., $f(x_1), f(x_2), \ldots, f(x_n)$, should be sorted in a nondecreasing order so that $f(x_1') \leq f(x_2') \leq \ldots \leq f(x_n')$, where $(x_1', x_2', \ldots, x_n')$ is a certain permutation of $\{x_1, x_2, \ldots, x_n\}$. Then, the value of the Choquet integral is obtained by

$$\int f d\mu = \sum_{i=1}^{n} [f(x_i') - f(x_{i-1}')] \cdot \mu(\{x_i', x_{i+1}', \ldots, x_n'\}) \tag{1}$$

where $f(x_0') = 0$.

To be convenient, Wang [24] has proposed a new algorithm to calculate the value of a Choquet integral with real-valued integrand by the product of two vectors as

$$\int f d\mu = \sum_{j=1}^{2^n-1} z_j \mu_j \tag{2}$$

in which

$$z_j = \begin{cases} \min_{i:\text{frc}(\frac{j}{2^i})\in[\frac{1}{2},1)} f(x_i) & - \max_{i:\text{frc}(\frac{j}{2^i})\in[0,\frac{1}{2})} f(x_i) \\ & \text{if it is } < 0 \text{ or } j = 2^n - 1 \\ 0, & \text{otherwise} \\ & \text{for } j = 1, 2, \ldots, 2^n - 1 \end{cases}$$

with a convention that the maximum on the empty set is zero, where $\text{frc}((j)/(2^i))$ is the fractional part of $(j)/(2^i)$. In the previous formula, if we express j in the binary form $j_n j_{n-1} \ldots j_1$, then $\{i | \text{frc}((j)/(2^i)) \in [(1)/(2,1))\} = \{i | j_i = 1\}$ and $\{i | \text{frc}((j)/(2^i)) \in [0, (1)/(2))\} = \{i | j_i = 0\}$.

III. Defuzzified Choquet Integral With Fuzzy-Valued Integrand

The original Choquet integral only supports crisp-valued integrands so they only can deal with crisp-valued data, and are helpless when facing with fuzzy information. To extend the advantages of Choquet integral to fuzzy domain such that it can manage fuzzy information, fuzzifications of orginal Choquet integral have been investigated [28], [30]. Such fuzzifications can support fuzzy-valued integrand. They are regarded as generalizations of the original Choquet integral since they are able to handle diverse forms of information, includng crisp data, interval values, fuzzy numbers and linguistic variables. The fuzzi-
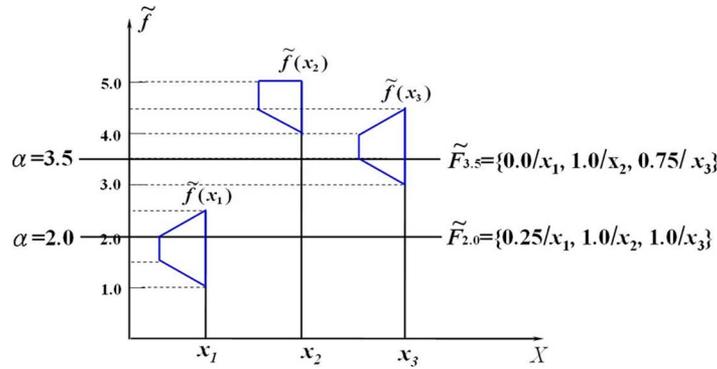
Fig. 2. α-cut of a fuzzy-valued function in Example 3.1.

fications of Choquet integral can have their integration results fuzzified or defuzzified. In this paper, we only focus on the model with fuzzy valued integrand and non-fuzzy valued integration result, which is called the DCIFI and defined as follows.

Definition 3.1: Let $\tilde{f} : X \to \mathcal{N}$ be a fuzzy-valued function defined on a universal set $X = \{x_1, x_2, \ldots, x_n\}$ and μ be a signed fuzzy measure defined on $\mathcal{P}(X)$, the power set of X, the DCIFI of $\tilde{f}$ is defined as

$$\int \tilde{f} d\mu = \int_{-\infty}^{0} [\mu(\tilde{F}_\alpha) - \mu(X)] d\alpha + \int_{0}^{\infty} \mu(\tilde{F}_\alpha) d\alpha$$

where $\tilde{F}_\alpha$ is the α-cut of the fuzzy-valued function $\tilde{f}$, and $\mathcal{N}$ denotes the set of all fuzzy numbers.

Obviously, the way to compute the value of the Choquet integral given in Section II cannot be directly applied for computing the DCIFI since the range of the fuzzy-valued function is not full-ordered, and therefore, the values of function $\tilde{f}$ at variant attributes cannot be rearranged in a nondecreasing order. However, we still can derive a calculation scheme of the DCIFI according to the fuzzy set theory and relevant properties of the Choquet integral. Actually, from the definition of the DCIFI, we can see that the calculation of the DCIFI can be rendered down into two subproblems.

1) How to get $\tilde{F}_\alpha$ for a fuzzy-valued function $\tilde{f}$?
2) How to get the value of $\mu(\tilde{F}_\alpha)$?

The following subsections aim to answer these questions, respectively.

A. The α-Cut of a Fuzzy-Valued Function

Let $\mathcal{F}(X)$ be the class of all fuzzy subsets of X. It is also called the *fuzzy power set* of X. Any fuzzy subset of X, $\tilde{A}$, can be expressed as

$$\tilde{A} = \{d_1/x_1, d_2/x_2, \cdots, d_n/x_n\}$$

where d_i is the degree of the membership of $\tilde{A}$ at x_i, $i = 1, 2, \ldots, n$. Let $\tilde{f}$ be a fuzzy-valued function defined on X. Function $\tilde{f}$ can be expressed as $(m_1, m_2, \ldots, m_n)$, where m_i is the membership function of $\tilde{f}(x_i)$ at x_i, $i = 1, 2, \ldots, n$.

Definition 3.2: For any given $\alpha \in (-\infty, \infty)$, the α-cut of a fuzzy-valued function $\tilde{f} = (m_1, m_2, \ldots, m_n)$, denoted $\tilde{F}_\alpha$,

is a fuzzy subset of X, whose membership function $m_{\tilde{F}_\alpha}$ has a degree of membership

$$m_{\tilde{F}_\alpha}(x_i) = \frac{\int_{\alpha}^{\infty} m_i(t) dt}{\int_{-\infty}^{\infty} m_i(t) dt} \tag{3}$$

at attribute x_i if $\int_{-\infty}^{\infty} m_i(t) dt \neq 0$, $i = 1, 2, \ldots, n$. When $\tilde{f}(x_i)$ is a crisp number, then $\int_{-\infty}^{\infty} m_i(t) dt = 0$. In this case, the degree of membership at x_i, denoted by $m_{\tilde{F}_\alpha}(x_i)$, is equal to 1 if $\tilde{f}(x_i) \geq \alpha$, or 0 if $\tilde{f}(x_i) < \alpha$.

Example 3.1: Let $X = \{x_1, x_2, x_3\}$ and let a fuzzy-valued function $\tilde{f}$ assign each element of X a trapezoidal fuzzy number, denoted by four parameters $[a_l a_b a_c a_r]$, that is, $\tilde{f}(x_1) = [1.0 1.5 2.0 2.5]$, $\tilde{f}(x_2) = [4.0 4.5 5.0 5.0]$, and $\tilde{f}(x_3) = [3.0 3.5 4.0 4.5]$. Then, we have

$$\tilde{F}_\alpha = (0.25/x_1, 1.0/x_2, 1.0/x_3) \text{ when } \alpha = 2.0$$

while

$$\tilde{F}_\alpha = (0.0/x_1, 1.0/x_2, 0.75/x_3) \text{ when } \alpha = 3.5.$$

as shown in Fig. 2.

B. The Choquet Extension of μ

Let X be a universal set, we can derive the signed fuzzy measure $\tilde{\mu}$ defined on $\mathcal{F}(X)$ based on the signed fuzzy measure μ defined on $\mathcal{P}(X)$.

Definition 3.3: Let μ be a signed fuzzy measure defined on $\mathcal{P}(X)$, the signed fuzzy measure $\tilde{\mu}$ is a set function mapping from the fuzzy power set of X, $\mathcal{F}(X)$, to $(-\infty, \infty)$. For any fuzzy set $\tilde{A} \in \mathcal{F}(X)$ with membership function $m_{\tilde{A}} : X \to [0, 1]$, we have

$$\tilde{\mu}(\tilde{A}) = \int m_{\tilde{A}} d\mu \tag{4}$$

where the integral is a Choquet integral with real-valued function, i.e., the membership function of $\tilde{A}$, $m_{\tilde{A}}$.

Here, for any crisp subset $A \in \mathcal{P}(X)$, we have $\tilde{\mu}(A) = \int \chi_A d\mu = \mu(A)$, where

$$\chi_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A \end{cases}$$

is the characteristic function of A. Thus, $\tilde{\mu}$ coincides with μ on $\mathcal{P}(X)$, that is, $\tilde{\mu}$ is an extension of μ from $\mathcal{P}(X)$ onto $\mathcal{F}(X)$ and called the *Choquet extension of* μ.

In this paper, we will simply use μ to replace $\tilde{\mu}$ on $\mathcal{F}(X)$ without any confusion.

Example 3.2: Let $X = \{x_1, x_2, x_3\}$ and a signed fuzzy measure μ be given as

$\mu_0 = \mu(\oslash) = 0 \quad \mu_1 = \mu(\{x_1\}) = 1 \quad \mu_2 = \mu(\{x_2\}) = -1$
$\mu_3 = \mu(\{x_1, x_2\}) = 3 \quad \mu_4 = \mu(\{x_3\}) = 2$
$\mu_5 = \mu(\{x_1, x_3\}) = -1 \quad \mu_6 = \mu(\{x_2, x_3\}) = 4$
$\mu_7 = \mu(\{x_1, x_2, x_3\}) = 5$

For fuzzy set $\tilde{A} = \tilde{F}_{2.0} = (0.25/x_1, 1.0/x_2, 1.0/x_3)$ and $\tilde{B} = \tilde{F}_{3.5} = (0.0/x_1, 1.0/x_2, 0.75/x_3)$ in Example 3.1, we have

$$\mu(\tilde{F}_{2.0}) = \int m_{\tilde{F}_{2.0}} d\mu$$
$$= m_{\tilde{A}}(x_1) \cdot \mu_7 + [m_{\tilde{A}}(x_2) - m_{\tilde{A}}(x_1)] \cdot \mu_6$$
$$= 0.25 \cdot 5 + (1.0 - 0.25) \cdot 4$$
$$= 4.25.$$
$$\mu(\tilde{F}_{3.5}) = \int m_{\tilde{F}_{3.5}} d\mu$$
$$= m_{\tilde{A}}(x_3) \cdot \mu_6 + [m_{\tilde{A}}(x_2) - m_{\tilde{A}}(x_3)] \cdot \mu_2$$
$$= 0.75 \cdot 4 + (1.0 - 0.75) \cdot (-1)$$
$$= 2.75.$$

C. Calculation of DCIFI

Now, we can calculate the value of the DCIFI. Obviously, it is rather difficult to express $\mu(\tilde{F}_\alpha)$ in an explicit form involving only fundamental functions of α, and by which, to compute the precise value of the DCIFI. However, we can numerically calculate it approximately. Before illustrating the algorithm, some concepts and properties are introduced.

The support set of a fuzzy number $\tilde{a}$, denoted by $\tilde{a}_{0+}$, is defined by $\tilde{a}_{0+} = \{t | m_{\tilde{a}}(t) > 0\}$, which is a crisp subset of the universe of discourse of the membership function of $\tilde{a}$. We denote the left and the right terminals of the support set of $\tilde{a}$ by $\tilde{a}_l$ and $\tilde{a}_r$, respectively. For example, a trapezoidal fuzzy number $\tilde{a} = [1.01.52.02.5]$ has $\tilde{a}_l = 1.0$ and $\tilde{a}_r = 2.5$; a normal fuzzy number $\tilde{b}$ has $\tilde{b}_l = -\infty$ and $\tilde{b}_r = \infty$.

A fuzzy-valued function $\tilde{f}$ assigns each element x_i in the universal set a fuzzy number $\tilde{f}(x_i)$, represented by its membership function $m_{\tilde{f}(x_i)}(t)$, $i = 1, 2, \ldots, n$. Now, we denote the left and the right terminals of the support set of $\tilde{f}(x_i)$ as $\tilde{f}(x_i)_l$ and $\tilde{f}(x_i)_r$, respectively.

Theorem 3.1: For a universal set X, let μ be a signed fuzzy measure on $\mathcal{F}(X)$ and $\tilde{f}$ be a fuzzy-valued function on X. Then,

$$\int \tilde{f} d\mu = \int (\tilde{f} - q) d\mu + q \cdot \mu(X)$$

for any real constant q.

Here, $(\tilde{f} - q)$ is also a fuzzy-valued function with its values represented by $\tilde{f}(x_i) - q$, $i = 1, 2, \ldots, n$, The subtraction between the fuzzy number $\tilde{f}(x_i)$ and the crisp number q is refer to [13]. The proof of this theorem is provided in the Appendix.

Using Theorem 3.1, we can write

$$\int \tilde{f} d\mu = \int_0^\infty \mu(\tilde{G}_\alpha) d\alpha + q \cdot \mu(X)$$
$$= \int_0^{r-q} \mu(\tilde{G}_\alpha) d\alpha + q \cdot \mu(X)$$

where $\tilde{G}_\alpha$ is the α-cut of function $\tilde{g} = \tilde{f} - q$, $q = \min_{1 \le i \le n} \tilde{f}(x_i)_l$, and $r = \max_{1 \le i \le n} \tilde{f}(x_i)_r$.

Now, we can numerically calculate the approximate value of the DCIFI through the following algorithm.

1) Input attributes' number n in X, subintervals' number K (with default value $K = 100$) required in the approximate computing, function's values $\tilde{f}(x_i)$ for $i = 1, 2, \ldots, n$, and the values of the signed fuzzy measure μ_j, $j = 1, 2, \ldots, 2^n - 1$.
2) Find $q = \min_{1 \le i \le n} \tilde{f}(x_i)_l$, $r = \max_{1 \le i \le n} \tilde{f}(x_i)_r$. If $q = -\infty$ or $r = \infty$, then take $\tilde{f}(x_i)_l$ and $\tilde{f}(x_i)_r$ as the left and right terminal of $\tilde{f}(x_i)|_{\alpha=\varepsilon}$, $i = 1, 2, \ldots, n$, respectively. Here, the ε is a very small positive real value defined by user with default value 10^{-3}. Then reset $q = \min_{1 \le i \le n} \tilde{f}(x_i)_l$, $r = \max_{1 \le i \le n} \tilde{f}(x_i)_r$, and set $\delta = (r - q)/K$.
3) Replace $\tilde{f}(x_i)$ by $\tilde{f}(x_i) - q$.
4) Initialize $\alpha = 0$ and $S = \mu_{2^n - 1}/2$.
5) $\alpha + \delta \Rightarrow \alpha$.
6) Whether $\alpha > (r - q)$? If yes, $\delta \cdot (S - \Delta S/2) + q \cdot \mu_{2^n - 1} \Rightarrow S$, output S as an approximate value of $\int \tilde{f} d\mu$, and stop; otherwise, continue.
7) Find $c_i = m_{F_\alpha}(x_i)$ by (3), $i = 1, 2, \ldots, n$.
8) Regarding $h = (c_1, c_2, \ldots, c_n)$ as a function on X, calculate $\Delta S = \int h d\mu$ by scheme shown in (2).
9) $S + \Delta S \Rightarrow S$ and go to 5).

We can see now, given a signed fuzzy measure, the value of the DCIFI is a crisp real number. Though the information on the fuzziness is compressed, applying such an aggregation tool in data mining is usually more convenient than giving a fuzzy number. In next section, using the DCIFI as a projection tool, a complex classification problem of heterogeneous fuzzy data can be optimally simplified to a classification problem of crisp data.

IV. DCIFI PROJECTION CLASSIFIER

In classification, an observation is denoted by an n-dimensional vector $(f(x_1), f(x_2), \ldots, f(x_n))$, whose components $f(x_i)$ are measurements of the feature attributes x_i, $i = 1, 2, \ldots, n$. We assume that there exist m groups or classes in the n-dimensional space, denoted by $C_1, C_2, \ldots, C_m$, and associated with each observation is a categorical variable Y that denotes the class or group membership. For example, if $Y = j$, then the observation belongs to C_j, $j \in \{1, 2, \ldots, m\}$. To design the classifier, we are usually given a set of training data with observations of known classes, represented as $\{(f_k(x_1), f_k(x_2), \ldots, f_k(x_n), Y_k), k = 1, 2, \ldots, l\}$. The training data set is used to set up internal parameters of the
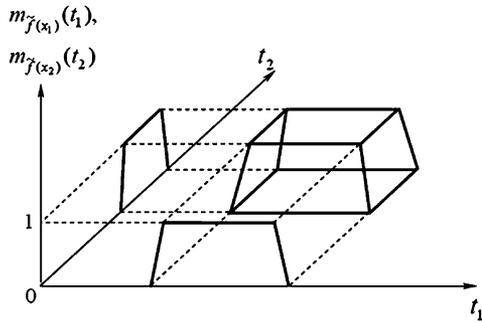
Fig. 3. Typical 2-dimensional heterogeneous fuzzy data.

classifier. Here, the positive integer l is the number of samples in the training data set. Once a classifier has been devised, we may estimate the class belongingness for any new observation.

When the measurements of feature attributes of an observation are heterogeneous fuzzy data, such as crisp data, fuzzy data, interval values, or linguistic variables, they are denoted by an n-dimensional fuzzy data vector $(\tilde{f}(x_1), \tilde{f}(x_2), \ldots, \tilde{f}(x_n))$. Such an n-dimensional fuzzy data vector can be visualized as a fuzzy point, which is not a single point but a special fuzzy subset in the n-dimensional space. Each coordinate value of a fuzzy point is a fuzzy number. A typical 2-dimensional heterogeneous fuzzy data $(\tilde{f}(x_1), \tilde{f}(x_2))$ is shown in Fig. 3. It is depicted as a frustum of a prism with height as 1. It has two coordinates which are represented by two different trapezoidal fuzzy numbers with their membership functions shown on the $m_{\tilde{f}(x_1)}(t_1)$-t_1 and the $m_{\tilde{f}(x_2)}(t_2)$-t_2 planes in Fig. 3, respectively.

Remember that the DCIFI takes a fuzzy-valued function as its integrand and gives a crisp value as its integration result. It can be regarded as a projection from the feature space onto the real axis. Under such a scheme, any fuzzy point $(\tilde{f}(x_1), \tilde{f}(x_2), \ldots, \tilde{f}(x_n))$, denoted simply by $(\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_n)$ in the feature space, is regarded as a fuzzy-valued function $\tilde{f}$ defined on $X = \{x_1, x_2, \ldots, x_n\}$, and furthermore, projected onto a virtual variable, denoted by $\hat{Y}$, on the real axis through a DCIFI defined by

$$\hat{Y} = \int \tilde{f} \, d\mu. \tag{5}$$

Fig. 4 illustrates the DCIFI projection of some heterogeneous fuzzy data in the two-dimensional space. Here, all heterogeneous fuzzy data are distributed into two classes. Each class has three observations. Each observation is identified by its fuzzy-valued coordinates $\tilde{f}(x_1)$ and $\tilde{f}(x_2)$. By certain DCIFI projection, each observation has been projected onto a virtual point (denoted by the black dots in Fig. 4) on the real axis L. It is natural to assume that there exists a boundary in the two-dimensional space, on which each point can be projected onto an identical virtual point (denoted by the white dot in Fig. 4), called the virtual boundary, on the real axis by the same DCIFI projection. According to this assumption, a classification problem of n-dimensional heterogeneous fuzzy data can be simplified to that of one-dimensional real data.

Good performance of the DCIFI projection classifier is expected due to the use of the signed fuzzy measure and the relevant nonlinear integral which can handle heterogeneous fuzzy data, since the nonadditivity of the signed fuzzy measure reflects
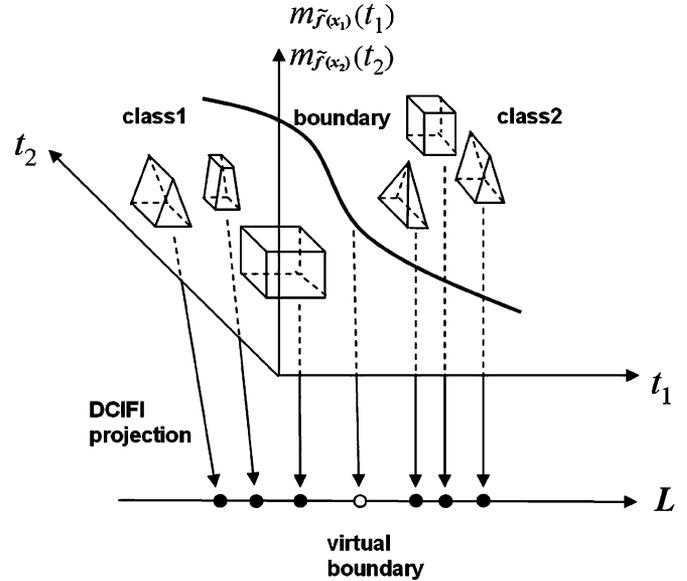
Fig. 4. DCIFI projection for 2-dimensional heterogeneous fuzzy data.

the importance of feature attributes, as well as their inherent interaction, toward the discrimination of the fuzzy points. In fact, the global contribution of several feature attributes to the decision of classification is not just the simple sum of the contribution of each feature to the decision. A combination of the feature attributes may have a mutually restraining or a complementary synergy effect on their contributions toward the classification decision. So, the signed fuzzy measure defined on the power set of all feature attributes is a proper representation of the respective importance of the feature attributes and the interaction among them, and a relevant DCIFI is a good fusion tool to aggregate information in different forms coming from the feature attributes for the classification.

V. GA-BASED ADAPTIVE CLASSIFIER-LEARNING ALGORITHM VIA DCIFI PROJECTION PURSUIT

Now, based on the DCIFI, we want to find an appropriate aggregation formula that projects the n-dimensional feature space onto the real axis, L, such that each fuzzy point $\tilde{f} = (\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_n)$ becomes a value of the virtual variable that is optimal with respect to classification. In such way, each classification boundary is just a point on the real axis L.

The classification task by the DCIFI projection classifier can be divided into two parts.
1) The DCIFI projection classifier depends on the signed fuzzy measure μ, so how to determine the values of μ is the first problem we are facing with.
2) Once the values of μ are retrieved, the DCIFI projection classifier is established. To classify new data, boundaries on the real axis L should be determined.

The following two subsections focus on the above two problems, respectively.

A. Boundaries Determination

A DCIFI projection classifier is identified by the signed fuzzy measure μ. Once the values of μ are given, the n-dimensional classification problem of heterogeneous fuzzy data is reduced
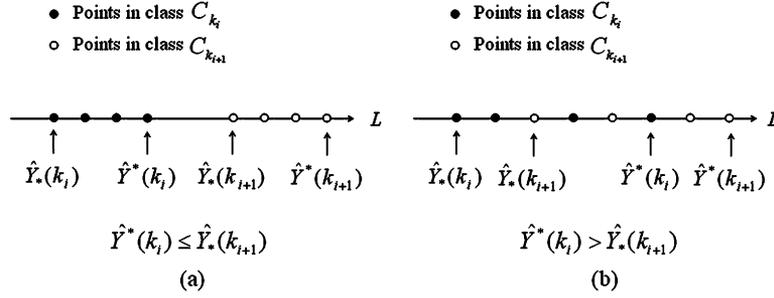
Fig. 5.   Illustration of virtual projection axis L when determining the boundary of a pair of successive classes C_{k_i} and $C_{k_{i+1}}$. (a) When $\hat{Y}^*(k_i) \leq \hat{Y}_*(k_{i+1})$. (b) When $\hat{Y}^*(k_i) > \hat{Y}_*(k_{i+1})$.

to a one-dimensional classification problem of crisp data on the virtual variable. The m classes of records in the original training data set are now projected to be m classes on the projection axis L. We can still use symbol C_k, $k = 1, 2, \ldots, m$, to denote these classes. The center c_k of each class C_k on L is the medium of the values of the virtual variables corresponding to the points in class C_k. The center c_k, expressed as a real number, is a numericalization of class C_k. After arranging $\{c_k | k = 1, 2, \ldots, m\}$ and, therefore, $\{C_k | k = 1, 2, \ldots, m\}$, in an increasing order as $(c_{k_1}, c_{k_2}, \ldots, c_{k_m})$ and $(C_{k_1}, C_{k_2}, \ldots, C_{k_m})$, where $(k_1, k_2, \ldots, k_m)$ is a permutation of $\{1, 2, \ldots, m\}$, we carry out a point-wise search for the best classification boundary between each pair of successive classes one by one under the criterion of minimizing the misclassification rate which is defined as the number of misclassified records (points) in the training set divided by data size l. The following algorithm is devoted to determining the boundaries of successive classes which have been rearranged according to the ascending order of their centers:

For $i = 1, 2, \ldots, m$

Find $\hat{Y}^*(k_i)$, the farthest right (largest) point of C_{k_i} and

$\hat{Y}_*(k_{i+1})$, the farthest left (smallest) point of $C_{k_{i+1}}$.

If $\hat{Y}^*(1) \leq \hat{Y}_*(k_{i+1})$ (as shown in Fig. 5(a))

$b_i = (1/2)(\hat{Y}^*(k_i) + \hat{Y}_*(k_{i+1}))$, where b_i is the boundary

between class C_{k_i} and $C_{k_{i+1}}$.

Else if $\hat{Y}^*(1) > \hat{Y}_*(k_{i+1})$ (as shown in Fig. 5(b))

b_i is the average of the collection points which satisfy

the following three conditions:

1) are members of class C_{k_i} and $C_{k_{i+1}}$;

2) are between $\hat{Y}^*(k_i)$ and $\hat{Y}_*(k_{i+1})$; and

3) have property "possessing the lowest number

of misclassified points if being a

classification boundary."

End If

End For

Thus, $b_1, b_2, \ldots, b_{m-1}$ are the best classification boundaries for the DCIFI projection classifier with respect to the given signed fuzzy measure μ. The corresponding global classification rate is the sum of the numbers of misclassified points in these $m - 1$ pairs of successive classes divided by l.

B. GA-Based Adaptive Classifier-Learning Algorithm

Here we discuss the optimization of the signed fuzzy measure μ under the criterion of minimizing the corresponding global misclassification rate, and then obtain an optimal DCIFI projection classifier. The optimizing process is just a "pursuit" for searching an appropriate projection direction. It is performed by the GACA. The optimization is also a data-driven process, where a set of training data set in the form of

$$\begin{matrix} \tilde{f}_{11} & \tilde{f}_{12} & \cdots & \tilde{f}_{1n} & \hat{Y}_1 \\ \tilde{f}_{21} & \tilde{f}_{22} & \cdots & \tilde{f}_{2n} & \hat{Y}_2 \\ \vdots & & & & \vdots \\ \tilde{f}_{l1} & \tilde{f}_{l2} & \cdots & \tilde{f}_{ln} & \hat{Y}_l \end{matrix}$$

are needed. Here, $\tilde{f}_{ji}$ denotes the fuzzy value of the ith feature at the jth observation and $\hat{Y}_j$ denotes the class tag of the jth observation, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, l$.

In the GACA, each individual of chromosome represents a DCIFI projection which is identified by the values of a signed fuzzy measure μ. Since real coding method is employed, each individual of chromosome consists of $2^n - 1$ genes. Each gene is represented by a real value between 0 and 1. The population in the GACA consists of p individuals of chromosome. The misclassification rate is adopted for estimating the fitness value of each individual of chromosome (i.e., the DCIFI projection). The probability of an individual of chromosome in the population being chosen as a parent to produce offspring depends on its fitness value. The optimization in the GACA is performed under the criterion of minimizing the misclassification rate. Fig. 6 shows the flow chart of the GACA.

It starts off from an initialized population. Individuals of chromosome in the population are decoded into their corresponding signed fuzzy measures to further determine their corresponding DCIFI projections. For a DCIFI projection, each observation in the training data set can be projected onto its virtual point on the real axis. According to the class tags provided by the training data, we can pursue the best virtual boundaries of the DCIFI projection being considered using the boundaries determination approach presented in the Subsection V-A. Then, cooperated with the training set, we can derive the misclassification rate of the
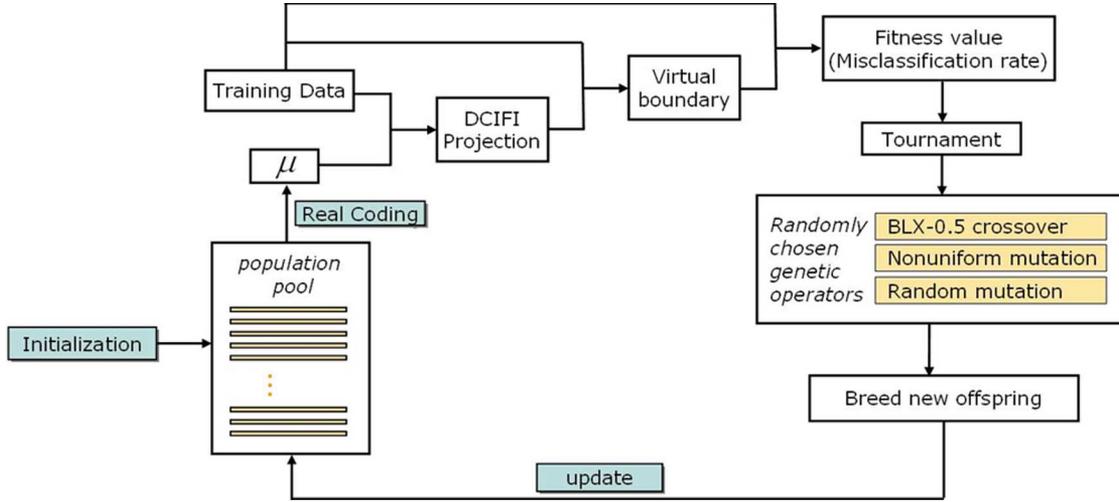
Fig. 6.   Flowchart of the GACA.
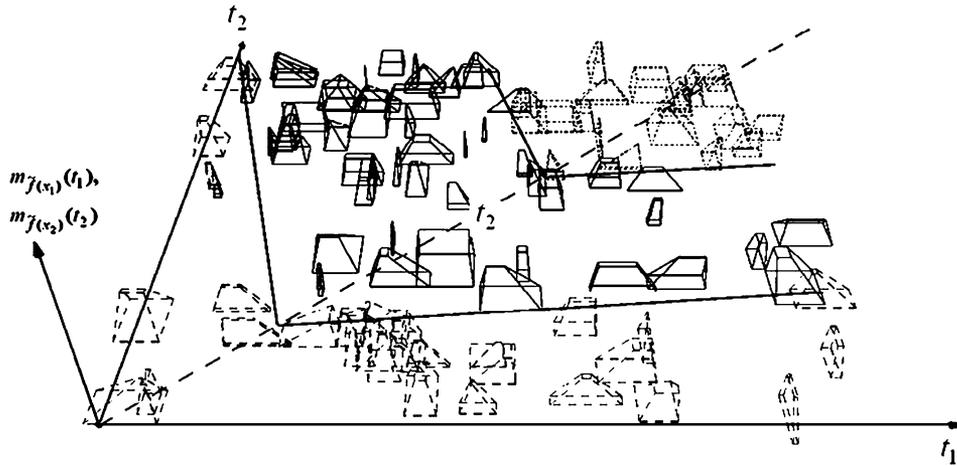
Fig. 7.   Training data and the trained classification boundaries in two-dimensional synthetic data test.

current DCIFI projection, which also represents the fitness value of the corresponding individual in the population. After that, a tournament selection is performed. Better individuals have more chance to produce offspring by some randomly chosen genetic operators [9]. The newly created offspring update the population. This process repeats until we get zero misclassification rate or the generation number exceeds the preset maximum number of generations.

To maintain the diversity of the searching space of our genetic algorithm, a special set of operations is used when the best fitness value remains unchanged for several consecutive generations (default value is 20). At that time, original population is divided into three parts by ascending order on fitness values. The individuals of chromosome in the first part are kept, while those in the second part create new offspring by random mutation [9], and those in the third part are replaced by new randomly created individuals of chromosome. Then, the population is updated and the iteration is continued.

After determining the signed fuzzy measure μ and the respective classification boundaries $b_1, b_2, \ldots, b_{m-1}$ from the training data, any new observation of the feature attributes

TABLE I
PRESET AND RETRIEVED VALUES OF THE SIGNED FUZZY MEASURE AND BOUNDARIES IN 2-D TEST

Parameters	Preset	Retrieved
$\mu(\{x_1\})$	-0.1	-0.105981
$\mu(\{x_2\})$	0.2	0.189793
$\mu(\{x_1, x_2\})$	1.0	1.0
b_1	0.2	0.201631
b_2	0.6	0.598042

$\tilde{f} = (\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_n)$ can be classified by calculating its corresponding value of the virtual variable

$$\hat{Y} = \int \tilde{f} \, d\mu$$

and checking its location relative to the classification boundaries in the order of $b_1, b_2, \ldots, b_{m-1}$ one by one. If $\hat{Y} \leq b_1$, then $\tilde{f}$ is classified into class C_{k_1}; if $\hat{Y} \in (b_{j-1}, b_j]$, then $\tilde{f}$ is classified into class C_{k_j}, $j = 2, 3, \ldots, m-1$; otherwise, $\tilde{f}$ is classified into class C_{k_m}.
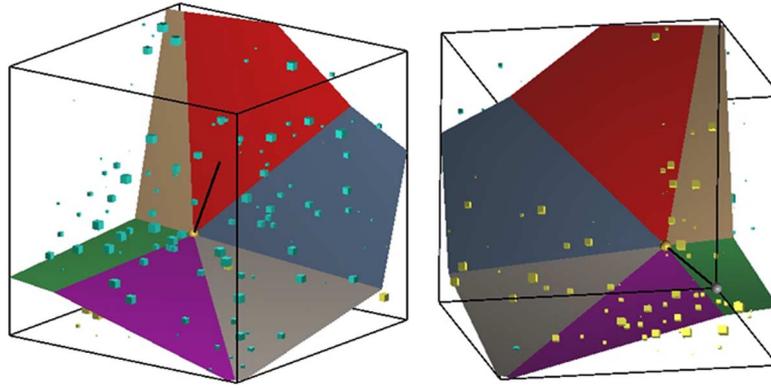
Fig. 8.   Artificial data and the classification boundaries in 3-dimensional test—from two view directions.

VI. Experimental Results

To evaluate the performance of the DCIFI projection classifier, a series of experiments both on synthetic and real data sets have been conducted. Comparisons are also implemented between the DCIFI projection classifier and two existing data mining methods, the fuzzy decision trees [31] and the fuzzy-neuro network [22], respectively.

A. Experiments on Synthetic Data

Two synthetic data sets, one containing two-dimensional heterogeneous fuzzy data distributed in three classes, and the other containing three-dimensional heterogeneous fuzzy data distributed in two classes, are generated and used to verify the efficiency and the effectiveness of the DCIFI and the GACA. To evaluate the performance of the GACA on recovering the classifier parameters, the classifier parameters, including the values of the signed fuzzy measure and the virtual boundaries, are preset. The preset DCIFI projection constructs normally distributed heterogeneous fuzzy data for each class which is separated by the preset virtual boundaries. Then, using the created training data sets, our GACA should recover the preset values of the parameters and obtain a low misclassification rate. The procedure to construct the synthetic training data sets is detailed as follows.

Assume that the data set has n feature attributes $\{x_1, x_2, \ldots, x_n\}$, m classes $\{C_1, C_2, \ldots, C_m\}$, and l records with l_j records for class C_j, $j = 1, 2, \ldots, m$. Here, $l = \sum_{j=1}^{m} l_j$. Each sample in the created data sets has the form of

$$\{(\tilde{f}(x_1), \tilde{f}(x_2), \ldots, \tilde{f}(x_n) \quad \text{class tag}\}.$$

The following algorithm creates the heterogeneous fuzzy data (with trapezoidal fuzzy number in each dimension) which are distributed in a unit hypercube in the n-dimensional space and classified into m classes.

1) Preset the values of the signed fuzzy measure μ by assigning $\mu_1, \mu_2, \ldots, \mu_{2^n-1}$ and the virtual boundaries $b_1, b_2, \ldots, b_{m-1}$.
2) Create the center of a fuzzy point in the n-dimensional space, represented as a vector $(c_1, c_2, \ldots, c_n)$. Each coordinate c_i, $i = 1, 2, \ldots, n$, of the center is a real

TABLE II
PRESET AND RETRIEVED VALUES OF THE SIGNED FUZZY MEASURE AND BOUNDARY IN 3-D TEST

Parameters	Preset	Retrieved
$\mu(\{x_1\})$	0.1	0.105585
$\mu(\{x_2\})$	0.2	0.181064
$\mu(\{x_1, x_2\})$	0.3	0.318546
$\mu(\{x_3\})$	0.05	0.0539669
$\mu(\{x_1, x_3\})$	0.25	0.246499
$\mu(\{x_2, x_3\})$	0.9	0.907981
$\mu(\{x_1, x_2, x_3\})$	1.0	1.0
boundary	0.23	0.239537

number generated by a random number generator with the uniform distribution in $[0, 1]$. Create a fuzzy point $(\tilde{f}(x_1), \tilde{f}(x_2), \ldots, \tilde{f}(x_n))$, where $\tilde{f}(x_i)$ is a randomly generated trapezoidal fuzzy number with its support set as $[c_i - r_i, c_i + r_i]$, $i = 1, 2, \ldots, n$. Here, r_i is a random value between 0.0 and 0.05.

3) For each observation $(\tilde{f}(x_1), \tilde{f}(x_2), \ldots, \tilde{f}(x_n))$, calculate the corresponding value of the DCIFI, denoted by $\hat{Y}$, with respect to the preset μ.

4) Create a random number, $\xi \in [0, 1]$, with the uniform distribution. In case $\hat{Y} \leq b_1$, if $\xi \leq e^{-(\hat{Y}-a_1)^2/2\sigma_1^2}$, then assign class C_1 to the right part of record, otherwise, abandon this record; in case $\hat{Y} \in (b_{j-1}, b_j]$, if $\xi \leq e^{-(\hat{Y}-a_j)^2/2\sigma_j^2}$, then assign class C_j to the right part of record, $j = 2, \ldots, m-1$, otherwise, abandon this record; in case $\hat{Y} > b_{m-1}$, if $\xi \leq e^{-(\hat{Y}-a_m)^2/2\sigma_m^2}$, then assign class C_m to the right part of record, otherwise, abandon this record. Here, the normal distribution $N(a_j, \sigma_j^2)$ are used to control the distribution of data in class C_j, $j = 1, 2, \ldots, m$.

5) Repeat steps 2)–4) until l_j records of class C_j, $j = 1, 2, \ldots, m$, have been created.

The first test is conducted on a classification problem of two feature attributes and three classes, that is, $X = \{x_1, x_2\}$, $C = \{C_1, C_2, C_3\}$. Totally 100 records are provided in the training data set, where 20 records for C_1, 50 records for C_2, and 30 records for C_3. The preset parameters to generate the training data are as follows: $\mu(\{x_1\}) = -0.1$, $\mu(\{x_2\}) = 0.2$, $\mu(\{x_1, x_2\}) = 1.0$, $b_1 = 0.2$ and $b_2 = 0.6$. Each record in the training data set presents a fuzzy point in the two-dimensional

TABLE III
PERFORMANCE COMPARISON BETWEEN DCIFI AND FUZZY ID3

Method	Training/Testing	Iris	Pima	Cancer	Sleep State
Fuzzy ID3	Training	0.9659	0.7519	0.9884	0.9526
	Testing	0.9701	0.7239	0.9222	0.9123
DCIFI Projection	Training	0.9711	0.7651	0.9893	0.9768
	Testing	0.9767	0.7591	0.9561	0.9342

space. Here, the fuzzy point is described by a 2-tuple vector whose elements are trapezoidal fuzzy number represented by their membership functions. Fig. 7 shows the sample data, where each frustum of a prism denotes a two-dimensional fuzzy point (with dashed contours for data of C_1, solid contours for data of C_2, and dashdotted contours for data of C_3). Setting $p = 20$ as the population size and running the GACA with the whole sample data, after three generations, zero misclassification rate is achieved, and we obtain a trained DCIFI projection classifier with the classification boundaries (solid lines in Fig. 7). Here, the dashed line starting from the origin shows the virtual real axis to which the two-dimensional heterogeneous fuzzy data are projected by the DCIFI. The values of the signed fuzzy measure and boundaries in the retrieved DCIFI projection classifier are rather close to the preset ones. That is to say, the GACA can retrieve the values of parameters well and perform the classification task successfully. The comparison of the preset and the retrieved values of parameters is listed in Table I.

The second test considers a 3-dimensional case. The classification samples have 3 feature attributes, $X = \{x_1, x_2, x_3\}$, and are classified into two classes $C = \{C_1, C_2\}$. 200 records are generated by the preset DCIFI parameters as: $\mu(\{x_1\}) = 0.1$, $\mu(\{x_2\}) = 0.2$, $\mu(\{x_1, x_2\}) = 0.3$, $\mu(\{x_3\}) = 0.05$, $\mu(\{x_1, x_3\}) = 0.25$, $\mu(\{x_2, x_3\}) = 0.9$, $\mu(\{x_1, x_2, x_3\}) = 1.0$, and $b_1 = 0.23$, where 80 records are for C_1 and 120 records are for C_2.

Setting $p = 30$ as the population size and running the GACA with the whole sample data, after 50 generations, we obtain the trained DCIFI projection classifier with misclassification rate 0. The values of the signed fuzzy measure in the retrieved DCIFI projection are rather close to their corresponding preset values. This experiment also confirms that our GACA can retrieve the values of the classifier parameters accurately. The comparison of the preset and the retrieved values of parameters is listed in Table II.

Fig. 8 illustrates the distribution of the training data and the classification boundary in three-dimensional feature space from two different viewing directions. The three-dimensional fuzzy data are represented by cubes in the graph. The lengths on three dimensions of a cube denote the ranges of support sets of the membership functions which represent the feature attributes of an observation. The blue cubes are of class C_1, while the yellow cubes are of class C_2. The classification boundary is a broken plane with six pieces that divide the feature space into two parts. These pieces of broken planes have a common vertex $(0.239537, 0.239537, 0.239537)$ on the virtual axis L (denoted by the black line in graph) that passes through the origin and points to point $(1, 1, 1)$. Fig. 8 also reveals the ability of the DCIFI projection classifier on classifying data which are separated by boundaries with irregular shape.
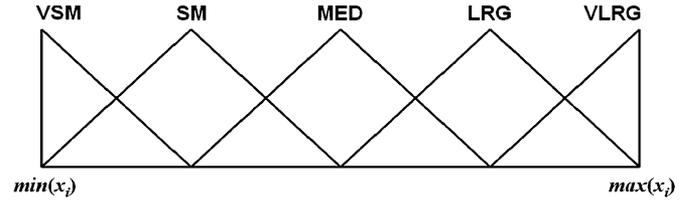
Fig. 9.   Five linguistic terms.

TABLE IV
THE ESTIMATED VALUES OF THE SIGNED FUZZY MEASURE AND THE VIRTUAL
BOUNDARY IN TWO-EMITTER IDENTIFICATION PROBLEM

Parameters	Estimated values
$\mu(\{x_1\})$	0.504087
$\mu(\{x_2\})$	0.476912
$\mu(\{x_1, x_2\})$	0.568434
$\mu(\{x_3\})$	0.394032
$\mu(\{x_1, x_3\})$	0.487458
$\mu(\{x_2, x_3\})$	0.503144
$\mu(\{x_1, x_2, x_3\})$	1.0
boundary	6.88557

B. Experiment and Comparison With Fuzzy Decision Tree on Real Data

We select four famous bench-mark data sets available in UCI [3] and compare our results with those by fuzzy decision tree induction based on ID3 [23]. Fuzzy ID3 is a fuzzy version of the crisp ID3 [21]. which is a popular and powerful heuristic method for generating crisp decision trees. As an extension of its counterpart, fuzzy ID3 is an important way of learning from fuzzy examples [10], [11].

The four data sets are summarized as follows.

1) Iris data set: 150 samples, three classes, four feature attributes.
2) Pima diabetes data set: 768 samples, two classes, eight feature attributes.
3) Breast cancer data set: 683 samples, two classes, nine feature attributes.
4) Sleep state data set [15]: 1236 samples, six classes, eleven feature attributes.

Since all these benchmark data sets are crisp values, fuzzification on the training/testing data sets are preprocessed. First, we need to find the minimum and maximum values of each attribute in the considered data set. The domain of each attribute is discretized into some linguistic terms. In our experiment, we use five linguistic terms shown in Fig. 9 [that is, "Very small (VSM)," "Small (SM)," "Medial (MED)," "Large (LRG)," and "Very large (VLRG)"]. Then, in the data set, each feature attribute of an observation is replaced by one linguistic term with the highest membership values respect to its original crisp value.

TABLE V
TESTING RESULTS ON TWO-EMITTER IDENTIFICATION PROBLEM WITH/WITHOUT NOISE

Error Deviation Level (%)	Total Average Accuracy (%)		
	NN by NVTBP algorithm	NN by CVTBP algorithm	DCIFI projection
15	99.71	91.04	100
13	99.90	93.75	100
11	99.91	94.85	100
9	99.91	95.49	100
7	99.91	95.83	100
5	99.91	96.03	100
3	99.91	96.15	100
1	99.91	96.23	100
0	99.91	96.26	100

The accuracies of both methods on each data set are determined by running ten-fold cross validation. For comparison, we present the average classification accurcies both on the training and testing data sets for each methods in Table III.

From the comparison, we can draw the following conclusion. Our DCIFI projection classifier outperforms the fuzzy ID3 algorithm in terms of testing and training accuracy. It is due to the fact that the DCIFI projection classifier utilizes the nonadditive set function, that is, the signed fuzzy measure, so that it can present the nonlinear relationships among the feature attributes for the determination of the classification more sufficiently.

C. Application on Emitters Identification

It is a high-priority problem in military operation to identify and track unique mobile transmitters for targeting. A powerful emitter identification function is necessary to warn of immediate threat with enough information to take evasive action. In military operation, such identification is accessed by Radio Frequency (RF), Pulse Width (PW), and Pulse Repetition Interval (PRI) of the collected pulse descriptor words. They form the feature attributes of an observation recognition problem, denoted by x_1, x_2, and x_3, respectively. The values of these features vary in interval ranges in accordance with a specific radar emitter. Shieh *et al.* proposed a fuzzy-neuro network to identify the emitters in [22], where an interval activation function is applied so that the network can process interval data. Two back propagation learning algorithms, NVTBP and CVTBP algorithms, were derived to tune the weights of neural network, and furthermore, to classify the observations. In our experiments, the DCIFI projection classifier is also implemented to identify different types of emitters, and its performance is compared to that of the fuzzy-neuro network. We use both the two-emitters and the three-emitters identification problems to test and compare the performance of the DCIFI projection classifier and those of the neural network approaches [22]. The training and testing data sets are the same as those in [22], where the data in training set are interval values while the data in testing set are crisp values. To evaluate the robustness of the proposed methods, a measurement distortion is also used as in [22] to simulate the adding of noise to the testing data. To perform the testing at different levels of adding noise, an Error Deviation Level (EDL) is defined in [22] by

$$\text{EDL}_i(\%) = \frac{\xi_{ji}}{x_{ji}} \times 100\%$$

TABLE VI
THE ESTIMATED VALUES OF THE SIGNED FUZZY MEASURE AND THE VIRTUAL BOUNDARIES IN THREE-EMITTER IDENTIFICATION PROBLEM

Parameters	Estimated values
$\mu(\{x_1\})$	0.488003
$\mu(\{x_2\})$	0.434324
$\mu(\{x_1, x_2\})$	0.479056
$\mu(\{x_3\})$	0.490667
$\mu(\{x_1, x_3\})$	0.454789
$\mu(\{x_2, x_3\})$	0.507754
$\mu(\{x_1, x_2, x_3\})$	1.0
boundary0	6.48158
boundary1	10.2373

for $i = 1, 2, 3$, and $j = 1, 2, \ldots, l$, where l is the number of observations. Here, x_{ji} denotes the values of attribute x_i of j-th observation in the testing data set, and ξ_{ji} is a small alteration added to the values of x_{ji}. The noisy testing data are obtained by adding random noise ξ_{ji} to each original testing observation, denoted by $(x_{j1} \pm \xi_{j1}, x_{j2} \pm \xi_{j2}, x_{j3} \pm \xi_{j3})$ with different EDL's (from 0% to 15%).

First, we consider the two-emitters identification problem with the input data corrupted by adding noise. For the DCIFI projection classifier, it is a three attributes and two classes problem. We set the population size p as 30, and the maximum number of generations as 1000. 10 training samples are used to train the DCIFI projection classifier and the neural network approaches respectively. The estimated values of the signed fuzzy measure and the virtual boundary are listed in Table IV.

Nine sets of 80 testing samples with different EDLs (from 0% to 15%) are generated and used to test the performance of the considered identification approaches. The experimental results on average accuracy are compared in Table V.

In the second experiment, we consider the three-emitters identification problem with the input data corrupted by adding noise. For the DCIFI projection classifier, it is a three attribute and three classes problem. We set the population size p as 30, and the maximum number of generations as 1000. 15 training samples are used to train the DCIFI projection classifier and the neural network approaches respectively. The estimated values of the signed fuzzy measure and the virtual boundary are listed in Table VI. 120 testing samples with different EDLs (from 0% to 15%) are used to train and test the performance of DCIFI projection classifier and the neural network approaches, respectively. The comparison results on average accuracy are shown in Table VII.

TABLE VII
TESTING RESULTS ON THREE-EMITTER IDENTIFICATION PROBLEM WITH/WITHOUT NOISE

Error Deviation Level (%)	Total Average Accuracy (%)		
	NN by NVTBP algorithm	NN by CVTBP algorithm	DCIFI projection
15	75.75	72.21	80.83
13	79.16	73.10	85.83
11	80.49	73.76	85.00
9	84.09	76.17	91.67
7	89.44	80.25	94.17
5	96.04	85.96	99.17
3	99.44	89.19	100
1	99.80	90.63	100
0	99.84	91.08	100

The comparison results shown in Tables V and VII indicate that the proposed DCIFI projection not only has higher identification capability, but also relatively more robust to noise than the neural network approaches.

VII. CONCLUSION

We have presented a new model, called the DCIFI, which uses a fuzzification of the Choquet integral for solving the classification problem involving heterogeneous fuzzy data, including crisp data, fuzzy numbers, interval values, and linguistic variables. The DCIFI acts as a projection tool which can map a high-dimensional heterogeneous fuzzy data to a crisp virtual value on a real axis, so that the classification problem in high dimensional heterogeneous fuzzy datum space is simplified to that in one dimensional crisp data space. A GACA has been used for searching the relevant optimal parameters in a DCIFI projection classifier, as well as the virtual boundaries of the projection images on the one-dimensional real axis. It has been shown that the DCIFI projection classifier achieves good performance on the classification problem though it requires long training time on estimating the internal parameters when the number of feature attributes or the number of classes is large. Moreover, this classifier is very informative and powerful for dealing with heterogeneous fuzzy data sets with strong interaction among their feature attributes towards the classification. The performance of the DCIFI projection classifier has been confirmed by a series of experiments both on synthetic and real data. It has also compared favorably with two existing approaches.

APPENDIX

The proof of Theorem 3.1 is summarized as follows.

Proof: Let $\tilde{g} = \tilde{f} - q$. Then $\tilde{g}$ is also a fuzzy-valued function and its α-cut, G_α, satisfies $G_\alpha = F_{\alpha+q}$ or, equivalently, $G_{\alpha-q} = F_\alpha$, for any real number α. Thus, denoting $\alpha - q$ by β, we have

$$\int \tilde{f}d\mu = \int_{-\infty}^{0} [\mu(F_\alpha) - \mu(X)]d\alpha$$
$$+ \int_{0}^{\infty} \mu(F_\alpha)d\alpha$$
$$= \int_{-\infty}^{0} [\mu(G_{\alpha-q}) - \mu(X)]d\alpha$$
$$+ \int_{0}^{\infty} \mu(G_{\alpha-q})d\alpha$$

$$= \int_{-\infty}^{0} [\mu(G_{\alpha-q}) - \mu(X)]d(\alpha - q)$$
$$+ \int_{0}^{\infty} \mu(G_{\alpha-q})d(\alpha - q)$$
$$= \int_{-\infty}^{-q} [\mu(G_\beta) - \mu(X)]d\beta$$
$$+ \int_{-q}^{\infty} \mu(G_\beta)d\beta$$
$$= \int_{-\infty}^{-q} [\mu(G_\beta) - \mu(X)]d\beta + \int_{-q}^{0} \mu(G_\beta)d\beta$$
$$+ \int_{0}^{\infty} \mu(G_\beta)d\beta - \int_{-q}^{0} \mu(X)d\beta$$
$$+ \int_{-q}^{0} \mu(X)d\beta$$
$$= \int_{-\infty}^{0} [\mu(G_\beta) - \mu(X)]d\beta$$
$$+ \int_{0}^{\infty} \mu(G_\beta)d\beta$$
$$+ \int_{-q}^{0} \mu(X)d\beta$$
$$= \int \tilde{g}d\mu + q \cdot \mu(X)$$
$$= \int (\tilde{f} - q)d\mu + q \cdot \mu(X).$$

REFERENCES

[1] S. Auephanwiriyakul, J. M. Keller, and P. D. Gader, "Generalized Choquet fuzzy integral fusion," *Inform. Fusion*, vol. 3, pp. 69–85, 2002.
[2] H. Bandemer and W. Nather, *Fuzzy Data Analysis*. London, U.K.: Kluwer, 1992.
[3] C. L. Blake and C. J. Merz, *UCI Repository of Machine Learning Databases*. Irvine, CA: Dept. Inform. Comput. Sci., Univ. California, 1998 [Online]. Available: http://www.ics.uci.edu/~mlearn/MLRepository.html
[4] G. Choquet, "Theory of capacities," *Annales de l'Institut Fourier*, vol. 5, pp. 131–295, 1954.
[5] D. Denneberg, *Non-Additive Measure and Integral*. Boston, MA: Kluwer, 1994.
[6] M. Grabisch and J. M. Nicolas, "Classification by fuzzy integral: Performance and Tests," *Fuzzy Sets Syst.*, vol. 65, pp. 255–271, 1994.
[7] M. Grabisch, "A new algorithm for identifying fuzzy measures and its application to pattern recognition," in *Proc. Joint Conf. 4th IEEE Int. Conf. Fuzzy Syst. 2nd Int. Fuzzy Eng. Symp.*, 1995, pp. 145–150.
[8] R. J. Hathaway, J. C. Bezdek, and W. Pedrycz, "A parametric model for fusing heterogeneous fuzzy data," *IEEE Trans. Fuzzy Syst.*, vol. 4, no. 3, pp. 270–281, Jun. 1996.

[9] F. Herrera, M. Lozano, and J. L. Verdegay, "Tackling real-coded genetic algorithms: Operators and tools for the behaviour analysis," *Artif. Intell. Rev.*, vol. 12, pp. 265–319, 1998.

[10] H. Ishibuchi, R. Fujioka, and H. Tanaka, "Neural network that learn from fuzzy if-then rules," *IEEE Trans. Fuzzy Syst.*, vol. 1, no. 2, pp. 85–97, Feb. 1993.

[11] H. Ishibuchi, T. Shirai, K. Nagasaka, and T. Miyoshi, "Neuro-fuzzy ID3: A method of inducing fuzzy decision trees with linear programming for maximizing entropy and an algebraic method for incremental learning," *Fuzzy Sets Syst.*, vol. 81, pp. 157–167, 1996.

[12] H. Ishibuchi, T. Yamamoto, and T. Nakashima, "Fuzzy data mining: Effect of fuzzy discretization," in *Proc. IEEE Int. Conf. Data Mining*, 2001, pp. 241–248.

[13] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic—Theory and Applications*. Upper Saddle River, NJ: Prentice-Hall, 1995.

[14] M. Mares, *Computation over Fuzzy Quantties*. Boca Raton, FL: CRC, 1994.

[15] R. S. Michalski, I. Mozetic, and J. R. Hong, "The multipurpose incremental learning system," in *Proc. 5th National Conf. Artif. Intell.*, 1986, pp. 1041–1045.

[16] L. Mikenina and H.-J. Zimmermann, "Improved feature selection and classification by the 2-additive fuzzy measure," *Fuzzy Sets Syst.*, vol. 107, pp. 197–218, 1999.

[17] T. Murofushi, M. Sugeno, and M. Machida, "Non monotonic fuzzy measures and the choquet integral," *Fuzzy Sets Syst.*, vol. 64, pp. 73–86, 1994.

[18] P. A. Nava, "Implementation of neuro-fuzzy systems through interval mathematics," in *Proc. IEEE ISIC/CIRA/ISAS Joint Conf.*, 1998, pp. 365–369.

[19] E. Pap, *Null-Additive Set Functions*. Boston, MA: Kluwer, 1995.

[20] V. Petridis and V. G. Kaburlasos, "Modeling of systems using heterogeneous data," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 1999, pp. V308–V313.

[21] J. R. Quinlan, "Induction of decision," *Mach. Learn.*, vol. 1, pp. 81–106, 1986.

[22] C.-S. Shieh and C.-T. Lin, "A vector neural network for emitter identification," *IEEE Trans. Antennas Propag.*, vol. 50, no. 8, pp. 1120–1127, Aug. 2002.

[23] E. C. C. Tsang, D. S. Yeung, and X. Z. Wang, "OFFSS: Optimal fuzzy-valued feature subset selection," *IEEE Trans. Fuzzy Syst.*, vol. 11, no. 2, pp. 202–213, Apr. 2003.

[24] Z. Wang, "A new genetic algorithm for nonlinear multiregressions based on generalized Choquet integrals," in *Proc. 12th IEEE Int. Conf. Fuzzy Syst.*, 2003, vol. 2, pp. 819–821.

[25] Z. Wang and G. J. Klir, *Fuzzy Measure Theory*. New York: Plenum, 1992.

[26] Z. Wang, G. J. Klir, and W. Wang, "Monotone set functions defined by choquet integral," *Fuzzy Sets Syst.*, vol. 81, pp. 241–250, 1996.

[27] Z. Wang, R. Yang, and K. S. Leung, "On the Choquet Integral with Fuzzy-Valued Integrand," in *Proc. 11th World Congr. Int. Fuzzy Syst. Assoc.*, 2005, pp. 433–437.

[28] Z. Wang, R. Yang, P. A. Heng, and K. S. Leung, "Real-valued choquet integrals with fuzzy-valued integrand," *Fuzzy Sets Syst.*, vol. 157, no. 1, pp. 256–269.

[29] K. Xu, Z. Wang, P.-A. Heng, and K.-S. Leung, "Classification by Nonlinear Integral Projections," *IEEE Trans. Fuzzy Syst.*, vol. 11, no. 2, pp. 187–201, Apr. 2003.

[30] R. Yang, Z. Wang, P. A. Heng, and K. S. Leung, "Fuzzy numbers and fuzzification of choquet integrals," *Fuzzy Sets Syst.*, vol. 153, no. 1, pp. 96–113.

[31] Y. Yuan and M. J. Shaw, "Induction of fuzzy decision tree," *Fuzzy Sets Syst.*, vol. 69, no. 2, pp. 125–139, 1995.
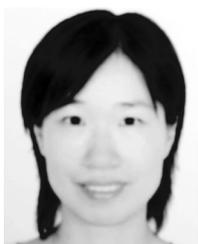
Rong Yang received the B.Sc. (Eng.) degree in electrical engineering, the M.Phil. degree in electronic and electrical engineering, and the Ph.D. degree in computer science, from the Southeast University (P.R. China) in 1998, Hong Kong University of Science and Technology in 2000, and The Chinese University of Hong Kong in 2005, respectively.

Currently, she is a Assistant Professor in the Department of Automatic Science, the College of Engineering and Technology, Shen Zhen University, Nan Shan, Shen Zhen, P.R. China. Her research interests include fuzzy theory, nonlinear integrals, nonlinear optimization, pattern recognization, soft computing techniques, and data mining.

Zhenyuan Wang graduated from Fudan University, China, and received the Ph.D. from the State University of New York at Binghamton.

He was a Visiting Professor at the State University of New York at Binghamton, New Mexico State University, and the University of Texas at El Paso, and a Research Fellow at the Chinese University of Hong Kong. Now, he is a Professor in the Department of Mathematics, University of Nebraska at Omaha. He is the author or coauthor of one book and more than 100 research papers. His research interests include fuzzy measure theory, nonlinear integrals, nonlinear optimization, soft computing techniques, and data mining.

Dr. Wang serves as an Associate Editor of the *Journal of Intelligent and Fuzzy Systems* and a member of the Editorial Board of *Fuzzy Sets and Systems* and *Fuzzy Mathematics*.

Pheng-Ann Heng (S'90–M'92) received the B.Sc. degree from the National University of Singapore in 1985, and the M.Sc. degree in computer science, the M.A. degree in applied mathematics, and the Ph.D. degree in computer science, all from Indiana University, Bloomington, in 1987, 1988, and 1992, respectively.

Currently, he is a Professor in the Department of Computer Science and Engineering, The Chinese University of Hong Kong (CUHK), Shatin, China. In 1999, he set up the Virtual Reality, Visualization and Imaging Research Centre at CUHK and serves as the Director of the Centre. He is also the Director of the CUHK Strategic Research Area in Computer Assisted Medicine, established jointly by the Faculty of Engineering and the Faculty of Medicine in 2000. His research interests include virtual reality applications in medicine, visualization, 3-D medical imaging, user interface, rendering and modelling, interactive graphics, and animation.

Kwong-Sak Leung (M'77–SM'89) received the B.Sc. (Eng.) and Ph.D. degrees from the University of London, Queen Mary College, London, U.K., in 1977 and 1980, respectively.

He worked as a Senior Engineer on contract R&D at ERA Technology and later joined the Central Electricity Generating Board to work on nuclear power station simulators in England. He joined the Computer Science and Engineering Department at the Chinese University of Hong Kong in 1985, where he is currently Professor of Computer Science and Engineering. His research interests are in soft computing including evolutionary computation, parallel computation, probabilistic search, information fusion and data mining, fuzzy data and knowledge engineering. He has authored and co-authored over 200 papers and 2 books in fuzzy logic and evolutionary computation.

Dr. Leung has been a Chair and a member of many program and organizing committees of international conferences. He is on the Editorial Board of *Fuzzy Sets and Systems* and is an Associate Editor of the *International Journal of Intelligent Automation* and *Soft Computing*. He is a s Chartered Engineer, a member of IEE and ACM, a Fellow of HKIE, and a Distinguished Fellow of HKCS in Hong Kong.