Pattern Classification Using Fuzzy Relational Calculus

Kumar S. Ray and Tapan K. Dinda

***Abstract*—Our aim is to design a pattern classifier using fuzzy relational calculus (FRC) which was initially proposed by Pedrycz . In the course of doing this, we first consider a particular interpretation of the multidimensional fuzzy implication (MFI) to represent our knowledge about the training data set. Subsequently, we introduce the notion of a fuzzy pattern vector to represent a population of training patterns in the pattern space and to denote the antecedent part of the said particular interpretation of the MFI. We introduce a new approach to the computation of the derivative of the fuzzy max-function and min-function using the concept of a generalized function. During the construction of the classifier based on FRC, we use fuzzy linguistic statements (or fuzzy membership function to represent the linguistic statement) to represent the values of features (e.g., feature F_1 is small and F_2 is big) for a population of patterns. Note that the construction of the classifier essentially depends on the estimate of a fuzzy relation $\Re$ between the input (fuzzy set) and output (fuzzy set) of the classifier. Once the classifier is constructed, the nonfuzzy features of a pattern can be classified. At the time of classification of the nonfuzzy features of the testpattens, we use the concept of fuzzy masking to fuzzify the nonfuzzy feaure values of the testpattens. The performance of the proposed scheme is tested on synthetic data. Finally, we use the proposed scheme for the vowel classification problem of an Indian language.**

***Index Terms*—Fuzzy pattern vector, fuzzy relational calculus (FRC), generalized function, multidimensional fuzzy implication (MFI), pattern classification.**

I. Introduction

In real-world pattern classification problems, fuzziness is connected with diverse facets of cognitive activity within the human being. The sources of fuzziness are related to labels expressed in pattern space, as well as, labels of classes taken into account in classification procedures. Although a lot of scientific developments have already been made in the area of pattern classification, existing techniques of pattern classification remain inferior to the human classification processes which perform extremely complex tasks. Hence, we attempt to develop a plausible tool using fuzzy relational calculus (FRC) for modeling and mimicking the cognitive process of human reasoning for pattern classification. The FRC approach to pattern classification can take care of uncertainties in feature values of patterns under different conditions like measurement error, noise, etc. Though there are several existing approaches to designing a classifier using the concept of fuzzy set/fuzzy logic [35]–[59], we have selected the concept proposed by W. Pedrycz [32] and suitably modified it to incorporate our new concept of the computation of the derivative of the fuzzy max-function and min-function. To represent the knowledge about the training data set, we consider a particular interpretation of multidimensional fuzzy implication (MFI) [26]. We consider a notion of fuzzy pattern vector, which represents the antecedent part of the said particular interpretation of the MFI to meaningfully carry out the task of pattern classification using FRC. During the construction of the classifier based on FRC, we use fuzzy linguistic statements (or fuzzy membership functions to represent the linguistic statement) to represent the values of features (e.g., feature F_1 is small and F_2 is big) for a population of patterns represented by the above fuzzy pattern vector. Note that the construction of the classifier essentially depends on the estimation of a fuzzy relation $\Re$ between the antecedent part and consequent part of the rules. Once the classifier is constructed, the nonfuzzy features of a pattern can be classified. At the time of classification of the nonfuzzy features of the testpattens, we use the concept of fuzzy masking to fuzzify the nonfuzzy feature values of the testpattens. The performance of the proposed scheme is tested on synthetic data. Finally, we use the proposed scheme for the vowel classification problem of an Indian language.

II. Statement of the Problem[1]

For the present problem, let us consider the conventional interpretation of a MFI [see App. B, Eq. (56a)] as given in

$$\left.\begin{array}{rl} \text{a)} & \text{if } x \text{ is A and } y \text{ is B then } z \text{ is C} \\ \text{or b)} & \text{if } x \text{ is A then } y \text{ is B then } z \text{ is C} \end{array}\right\} \quad (1)$$

and the notion of a fuzzy pattern vector (see App. B) which represents the antecedent clauses of (a) of (1) and locates a population of patterns P in the quantized pattern space.[2] Assume that the quantized pattern space consists of "c" universes $U_1, U_2, \ldots, U_c$ in the form $U = U_1 \times U_2 \times \cdots \times U_c$, where each U_i represents the universe on the ith feature axis $F_i, i = 1, 2, \ldots, c$.

Assume that D is a fuzzy relation [formed by the antecedent clauses of a) of (1)], which is a fuzzy set in quantized product space[2] U, namely $\mu_D : U \rightarrow [0, 1]$. Also, assume that there exists a set C_{class} of finite number of classes $c_1, c_2, \ldots, c_n$, i.e., $C_{\text{class}} = \{c_1, c_2, \ldots, c_n\}$, by which the finite range of the pattern space is covered. The consequent clause of a) of (1) is a

Manuscript received November 7, 1998; revised September 29, 1999 and December 27, 2001. This paper was recommended by Associate Editor W. Pedrycz.

K. S. Ray is with the Electronics and Communication Sciences Unit, Indian Statistical Institute, Calcutta-700 035, India (e-mail: ksray@isical.ac.in).

T. K. Dinda is with the Alumnis Software Ltd., INFINITY, and with the University of Calcutta, Calcutta, India (e-mail: tapan@alumnux.com).

Digital Object Identifier 10.1109/TSMCB.2002.804361

[1] For further clarity of the section, the see Appendix B.

[2] See Remark 8 of Appendix B.

fuzzy set $C = \sum_{j=1}^{n} \mu_c(c_j)/c_j$, where $\mu_c(C_j)$ denotes the degree of belongingness of the population of patterns P to the class c_j, for $j = 1, 2, \ldots, n$ (see Example 2 of Appendix B).

Therefore, by considering the conventional interpretation of a MFI, the fuzzy set D formed by the antecedent clauses of a) of (1) is associated with the fuzzy set C which represents the consequent clause of a) of (1). Hence, there exists a relation between D and C. More precisely, D and C are related via a certain relation $\Re$ (i.e., $D\Re C$), which is presently unknown and has to be estimated, based on the training data set, for the design of the classifier. Now, for the testing of the classifier, we specify how C is derived from given D and estimated $\Re$. We may consider the fuzzy relational equation, namely, a direct equation

$$C = Do\Re \tag{2}$$

where $0 \equiv \max - t$ composition operator, where t is a T-norm operator.

Equation (2) can be rewritten, in terms of the membership function, in the following form:

$$\mu_C(C_j) = \bigvee_{u \epsilon U} [\mu_D(u) t \mu_R(u, c_j)] \quad \text{for } j = 1, 2, \ldots, n. \tag{3}$$

This explicit form of (3) is needed for actual design study of the classifier.

Let us asume that the training set consists of ordered pairs

$$(P_1, C_1), (P_2, C_2), \ldots, (P_k, C_k)$$

and the classifier relation is supposed to specify a system of equations

$$C_l = D_l o \Re_l, \quad l = 1, 2, 3, \ldots, k \tag{4}$$

then the fuzzy relation which satisfies (4) is given by

$$\hat{\Re} = \bigcap_{l=1}^{k} \Re_l. \tag{5}$$

But the above mentioned system of equations in (4) may not have a solution [32]. Hence, in this paper we look for an approximate solution of the system of fuzzy relation equations in (4).

The advantages which we obtain from FRC approach to pattern classification are as follows

- We obtain the local description of the pattern space in terms of few quantized zones [61]. Depending upon the need or the problem, we may increase or decrease the granularity of our description of pattern space with smaller or bigger quantized zones.
- For estimating the relation $\Re$ of a classifier, we do not have to select the representative data set from the given set of data (patterns). Instead, we use the gross property of few populations of the given data (patterns) spread over the pattern space by using few fuzzy pattern vectors which are formed by the different combinations of the primary fuzzy terms defined over the universe of the individual feature axis (see the Appendix B) and which describe the overall distribution of patterns in pattern space.
- We obtain multiple classification which is very natural in the case of overlapped classes of patterns.

III. Existing Method to Solve Fuzzy Relation Equation

The numerical solution of fuzzy relational equation has been proposed by several researchers [1]–[8], [10], [13], [17], [21]. In this section, we briefly review the method proposed by Pedrycz [2]. We focus our attention on max-t composition operator of fuzzy relational equations, which are defined on finite spaces

$$C = Do\Re \tag{6}$$

where $o \equiv \max - t$ composition operator, D and C are the fuzzy sets defined on the universe of discourses $U = \{u_1, u_2, \ldots, u_m\}$ and $C_{\text{class}} = \{c_1, c_2, \ldots, c_n\}$, respectively, and $\Re$ is the fuzzy relation on $U \times C_{\text{class}}$. Let $r_{ij} = (u_i, c_j/u_i \epsilon U, c_j \epsilon C_{\text{class}}), i = 1, 2, \ldots, m, j = 1, 2, \ldots, n,$; then, the fuzzy sets D and C and fuzzy relation $\Re$ are as follows:

$$\left.\begin{array}{r} D = [\mu_D(u_i)]_{1\times m} \epsilon F(U) \\ C = [\mu_C(c_j)]_{1\times n} \epsilon F(C_{\text{class}}) \\ \Re = [\mu_\Re(r_{ij})]_{m\times n} \epsilon F(U \times C_{\text{class}}) \end{array}\right\}. \tag{7}$$

If the universe U of the quantized pattern space consists of 'c' features, say $F_i, i = 1, 2, \ldots, c$, the D is a fuzzy set defined on the quantized product spaces of $U_1, U_2, \ldots, U_c$, that is $U = U_1 \times U_2 \times \cdots \times U_c$, where $U_i = \{u_1^i, u_2^i, \ldots, u_{m_i}^i\}$ is the universe of the ith feature axis F_i with card $(U_i) = m_i$. Let D^i be the fuzzy set on U_i, i.e., $D^i = [\mu_{D^i}(u_j^i)]_{1\times m_i} \epsilon F(U_i)$ for $i = 1, 2, \ldots, c$; then, card $(U) = m = \prod_{i=1}^{c} m_i$ and u_i is the $c$tuple each of type $u_i = (u_{i_1}^1, u_{i_2}^2, \ldots, u_{i_c}^c / u_{i_p}^p \epsilon U_p, p = 1, 2, \ldots, c)$, and corresponding membership value belonging to D is determined as (8) shown at the bottom of the page, where $i = \sum_{p=1}^{c-1} (\Pi_{k>p}^{c} m_k)(i_p - 1) + i_c$, for each $i_p = 1, 2, \ldots, m_p, p = 1, 2, \ldots, c$. Equation (6) can be put in the following form:

$$\mu_c(c_j) = \bigvee_{i=1}^{m} \{\mu_D(u_i) t \mu_\Re(r_{ij}\}, \quad \text{for } j = 1, 2, \ldots, n \tag{9}$$

where t is the t-norm operator.

Thus, from (8) and (9), where 't' of (9) is one of the operators in $\{$prod, min$\}$, we get following four types of problems:

Type I: by using i) of (8) and $t \equiv \text{prod}$ of (9);
Type II: by using ii) of (8) and $t \equiv \text{prod}$ of (9);
Type III: by using i) of (8) and $t \equiv \min$ of . (9);
Type IV: by using ii) of (8) and $t \equiv \min$ of (9).

$$\left.\begin{array}{l} \text{i) } \mu_D(u_i) = \bigwedge_{p=1}^{c} \left\{\mu_{D^p}\left(u_{ip}^p\right)\right\} = \mu_{D^1}\left(u_{i_1}^1\right) \wedge \mu_{D^2}\left(u_{i_2}^2\right) \ldots \wedge \mu_{D^c}\left(u_{i_c}^c\right) \\ \text{ii) } \mu_D(U_i) = \prod_{p=1}^{c} \left\{\mu_{D^p}\left(U_{ip}^p\right)\right\} = \mu_{D^1}\left(u_{i_1}^1\right) \cdot \mu_{D^2}\left(u_{i_2}^2\right) \ldots \mu_{D^c}\left(u_{i_c}^c\right) \end{array}\right\} \tag{8}$$

Let E be the sum of the square of the error over $p = 1, 2, \ldots, n$ and is defined by

$$E = \sum_{p=1}^{n} \{\mu_C(c_p) - \mu_{\tilde{C}}(c_p)\}^2 \tag{10}$$

where C is the calculated fuzzy set using (9), and $\tilde{C}$ is the desired fuzzy set.

Now, the basic problem is to estimate $\Re = [\mu_\Re(r_{ij})]_{m \times n}$ via some given D and C which minimize E defined in (10) and satisfying $\{\mu_\Re(r_{ij}) \wedge (1 - \mu_\Re(r_{ij}))\} \geq 0, \forall i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.

A general method to solve an optimization problem, defined above, is to solve a set of equations, which form the necessary conditions for a minimum of the square of the error defined in (10). Thus, we have $[(\partial E)/(\partial \mu_\Re(r_{ij}))]_{m \times n} = [0]_{m \times n}$. Now, we discuss the applicability of Newton's method and its simplification.

The Newton's iterative scheme for finding the solution of $\Re = [\mu_\Re(r_{ij})]_{m \times n}$ is

$$\mu_\Re(r_{ij})^{(s+1)} = \mu_\Re(r_{ij})^{(s)} - \alpha_s \cdot \left.\frac{\partial E}{\partial \mu_\Re(r_{ij})}\right| \Re = \Re^{(s)} \tag{11}$$

where $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n \cdot \alpha_s$ is the convergent factor and also is an nonincreasing gain factor depending on the number of iteration. It can be described as $\alpha_s = 1/(2.0 + s^k) \cdot \geq 0$ is chosen empirically in order to achieve good convergent properties and avoid significant oscillations in the iteration procedure [2].

Now

$$\frac{\partial E}{\partial \mu_\Re(r_{ij})} = 2\{\mu_C(C_j) - \mu_{\tilde{C}}(C_j)\} P_{ij} \tag{12}$$

$$\text{where } P_{ij} = \frac{\partial \mu_C(C_j)}{\partial \mu_\Re(r_{ij})}, \quad \text{i.e.,} \tag{13}$$

$$\begin{aligned} P_{ij} &= \frac{\partial}{\partial \mu_\Re(r_{ij})} \left[\bigvee_{p=1}^{m} \{\mu_D(u_p) t \mu_\Re(r_{pj})\} \right] \\ &= \frac{\partial}{\partial \mu_\Re(r_{ij})} \left[\bigvee_{p \neq i}^{m} \{\mu_D(u_p) t \mu_\Re(r_{pj})\} \right. \\ &\qquad \left. \times \bigvee \{\mu_D(u_i) t \mu_\Re(r_{ij})\} \right] \end{aligned} \tag{14}$$

for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.

If we consider t-norm operator as "prod," then the (9) is written as

$$\mu_C(c_j) = \bigvee_{i=1}^{m} \{\mu_D(u_i) \cdot \mu_\Re(r_{ij})\}, \quad \text{for } j = 1, 2, \ldots, n \tag{15}$$

and in this case P_{ij} in (14) is determined as (16) shown at the bottom of the page, for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.

Again, if we cosider t-norm operator as "$\min$," then (9) is written as

$$\mu_C(c_j) = \bigvee_{i=1}^{m} \{\mu_D(u_i) \wedge \mu_\Re(r_{ij})\}, \quad \text{for } j = 1, 2, \ldots, n \tag{17}$$

and in this case, P_{ij} is determined as

$$P_{ij} = \begin{cases} 1, & \text{if } \bigvee_{p \neq i}^{m} \{\mu_D(u_p) \wedge \mu_\Re(r_{pj})\} \leq \mu_D(u_i) \wedge \mu_\Re(r_{ij}) \\ & \text{and } \mu_D(u_i) \geq \mu_\Re(r_{ij}) \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.

Here, the derivative of the max-function and min-function in the (14), (16), and (18), respectively are as follows:

$$\frac{\partial}{\partial w} \left(w \bigvee a \right) = \begin{cases} 1, & \text{if } w > a \\ 0, & \text{if } w < a \end{cases} \tag{19}$$

where $a = \bigvee_{p \neq i}^{m} \{\mu_D(u_p) t \mu_\Re(r_{pj})\}$ and $w = \mu_D(u_i) t \mu_\Re(r_{ij})$ and

$$\frac{\partial}{\partial z}(z \wedge b) = \begin{cases} 1, & \text{if } z < b \\ 0, & \text{if } z > b \end{cases} \tag{20}$$

where $b = \mu_D(u_i)$ and $z = \mu_\Re(r_{ij})$, which are piecewise differentiable and is undefined at $w = a$ for max-function in (19) and $z = b$ for min-function in (20). Thus, we get some problems in our numerical computation [7] which may be overcome by defining the derivatives at $w = a$ and $z = b$, respectively as follows

$$\frac{\partial}{\partial w}(w \vee a) = \begin{cases} 1, & \text{if } w \geq a \\ 0, & \text{if } w < a \end{cases} \tag{21}$$

and

$$\frac{\partial}{\partial z}(z \wedge b) = \begin{cases} 1, & \text{if } z \leq b \\ 0, & \text{if } z > b. \end{cases} \tag{22}$$

Both formulas for the computation of the derivatives of the max and min functions, as mentioned above, return either 0 or 1 value of the derivatives. Such two-valued results of the derivatives have some inherent difficulties, in connection to the convergence of the solution as mentioned in [7]. To overcome such difficulties there are some propositions in [7]. In the following section, we will provide an alternative approach based on generalized functions (see Appendix A).

The above method for solving fuzzy relational equations can be extended to simultaneous fuzzy relational equations [2] as given below.

The simultaneous fuzzy relational equations for given total k number of data in the training set are as follows:

$$C_l = D_l o \Re, \quad l = 1, 2, \ldots, k, \tag{23}$$

$$P_{ij} = \begin{cases} \mu_D(u_i), & \text{if } \bigvee_{p \neq i}^{m} \{\mu_D(u_p) \cdot \mu_\Re(r_{pj})\} \leq \mu_D(u_i) \cdot \mu_\Re(r_{ij}) \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

and their membership functions are as follows

$$\mu_{C_l}(c_j) = \bigvee_{i=1}^{m} \{\mu_{D_l}(u_i) t \mu_{\Re}(r_{ij})\}, \quad \text{for } j = 1, 2, \ldots, n \quad (24)$$

where $l = 1, 2, \ldots, k$, and

$$\left.\begin{array}{r} D_l = [\mu_{D_l}(u_i)]_{1\times m} \epsilon F(U) \\ C_l = [\mu_{C_l}(c_j)]_{1\times n} \epsilon F(C_{\text{class}}) \\ \Re = [\mu_{\Re}(r_{ij})]_{m\times n} \epsilon F(U \times C_{\text{class}}) \end{array}\right\}. \quad (25)$$

In this case, the error E is taken by summing over all the data set. Thus, (10) is modified as follows:

$$E = \sum_{l=1}^{k} \sum_{p=1}^{n} \{\mu_{C_l}(c_p) - \mu_{\tilde{C}_l}(c_p)\}^2 \quad (26)$$

satisfying $\{\mu_{\Re}(r_{ij}) \wedge (1 - \mu_{\Re}(r_{ij}))\} \geq 0, i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$, where C_l is the calculated fuzzy set using (24), and $\tilde{C}_l$ is the desired fuzzy set. The iterative scheme of (11) for finding the relation $\Re$ remains the same. Only the expression α_s in (11) could be modified as $\alpha_s = 1/(2 \times k + s^k)$, which depends on the number of data k.

IV. Modified Approach to Solve Fuzzy Relational Equation

We modify the above said approach to solve the fuzzy relational equation (FRE) by incorporating a rigorous treatment on the computation of the derivative of max-function and min-function indicated in the (21) and (22), respectively.

A. Derivative of Max-Function

Let the maximum value of $h_i, i = 1, 2, \ldots, s$ be determined by a function called max-function and defined by

$$h_{\max} = \bigvee_{i=1}^{s} h_i. \quad (27)$$

Now, our intention is to calculate the derivative of max-function defined as above with respect to one of its variables. Hence, we transfer the said max-function of (27) into the following functional form

$$G(h_1, h_2, h_3, \ldots, h_s, h_{\max}) \equiv \sum_{i=1}^{s} \{H(h_{\max} - h_i) - 1\} + G_e = 0 \quad (28)$$

where $H(h_{\max} - h_i)$ is the *Heaviside function* defined by

$$H(h_{\max} - h_i) = \begin{cases} 1, & \text{if } h_{\max} > h_i \\ 0, & \text{otherwise} \end{cases} \quad (29)$$

and G_e is the number of $h_i's, i = 1, 2, \ldots, s$ that are equal to $h_{\max}$. Also, it is a constant and independent of $h_i, i = 1, 2, \ldots, s$ and $h_{\max}$.

Now, by using implicit function theorem, we write

$$\frac{\partial h_{\max}}{\partial h_r} = \frac{\partial G}{\partial h_r} \bigg/ \frac{\partial G}{\partial h_{\max}} \quad (30)$$

where $r \epsilon \{1, 2, \ldots, s\}$.

We calculate the partial derivatives $(\partial G)/(\partial h_r)$ and $(\partial G)/(\partial h_{\max})$, using the derivative of *Heaviside function* in (55) of Appendix A, as follows:

$$\frac{\partial G}{\partial h_r} = -\delta(h_{\max} - h_r) \quad (31)$$

$$\frac{\partial G}{\partial h_{\max}} = \sum_{i=1}^{s} \delta(h_{\max} - h_i) \quad (32)$$

where $\delta(*)$ is the *Dirac delta* function.

Using Eqs. (31) and (32) in (30), we get

$$\frac{\partial h_{\max}}{\partial h_r} = \frac{\delta(h_{\max} - h_r)}{\sum_{i=1}^{s} \delta(h_{\max} - h_i)} = \begin{cases} \frac{1}{N_{\max}}, & \text{if } h_{\max} = h_r \\ 0, & \text{otherwise} \end{cases} \quad (33)$$

where $N_{\max} \triangleq$ number of terms h_i, satisfying the condition $h_{\max} = h_i, i = 1, 2, \ldots, s$, i.e., $N_{\max} = \sum_{i=1}^{s} \delta(h_{\max} - h_i)$, which never vanishes because at least one of $h_i', si = 1, 2, \ldots, s$ must be equal to $h_{\max}$. So $(\partial h_{\max})/(\partial h_r)$ in (33) always exists everywhere.

B. Derivative of Min-Function

Let the minimum value of $h_i, i = 1, 2, \ldots, s$ can be determined by a function called min-function and defined by

$$h_{\min} = \bigwedge_{i=1}^{s} h_i. \quad (34)$$

Now, our intention is to calculate the derivative of min-function defined as above with respect to one of its variables. Hence, we transfer the said max-function of (34) in the following functional form:

$$L(h_1, h_2, h_3, \ldots, h_s, h_{\min}) \equiv \sum_{i=1}^{s} \{H(h_i - h_{\min}) - 1\} + L_e = 0 \quad (35)$$

where L_e is the number of $h_i's, i = 1, 2, \ldots, s$ that are equal to $h_{\min}$. Also it is a constant independent of $h_i, i = 1, 2, \ldots, s$ and $h_{\min}$.

Now, by using the implicit function theorem, we write

$$\frac{\partial h_{\min}}{\partial h_r} = -\frac{\partial L}{\partial h_r} \bigg/ \frac{\partial L}{\partial h_{\min}} \quad (36)$$

where $r \epsilon \{1, 2, \ldots, s\}$.

We calculate the partial derivatives $(\partial L)/(\partial h_r)$ and $(\partial L)/(\partial h_{\min})$, using the derivative of *Heaviside function* in (55) of Appendix A, as follows:

$$\frac{\partial L}{\partial h_r} = \delta(h_r - h_{\min}) \quad (37)$$

$$\frac{\partial L}{\partial h_{\min}} = -\sum_{i=1}^{s} \delta(h_i - h_{\min}). \quad (38)$$

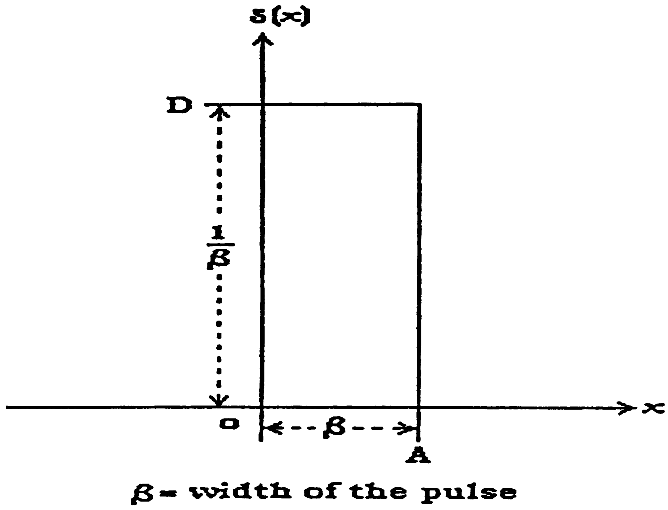


Fig. 1. Approximation of *Delta function* $\delta(x)$.

Using (37) and (38) in (36), we get

$$\frac{\partial h_{\min}}{\partial h_r} = \frac{\delta(h_r - h_{\min})}{\sum_{i=1}^{s}\delta(h_i - h_{\min})} = \begin{cases} \frac{1}{N_{\min}}, & \text{if } h_{\min} = h_r \\ 0, & \text{otherwise} \end{cases} \tag{39}$$

where $N_{\min} \stackrel{\triangle}{=}$ number of terms $h_i, i = 1, 2, \ldots, s$ satisfying the condition $h_{\min} = h_i$, i.e., $N_{\min} = \sum_{i=1}^{s}\delta(h_i - h_{\min})$, which never vanishes because at least one of $h_i's, i = 1, 2, \ldots, s$ must be equal to $h_{\min}$. So $(\partial h_{\min})/(\partial h_r)$ in (39) always exists everywhere.

Thus, from the above discussion, we understand that both the derivative of max and min functions depend on the derivative of the *Heaviside function* which is discussed, for general readability of the paper, in the Appendix A.

C. Applicable Form of the Computation of Derivative of Max and Min Functions

For the implementation of the expression of the derivative of fuzzy max and min functions, we approximate the *Delta function* using a finite pulse shown in Fig. 1. The motivation behind the approximation of the *Delta function* by a finite pulse is to incorporate the notion of uncertainties built in the given data, which are all attached with fuzzy membership functions, indicating their (data) degree of possibilities to take part in any decision making process. Thus, if we approximate the *Delta function* by a finite pulse with width β, that means we try to take care of the possibilities of all the data that fall within the range of β in our computation of the derivative of a fuzzy max and min functions. Using these approximations, we formulate the approximate derivative of the max and min functions, respectively, as follows:

$$\frac{\partial h_{\max}}{\partial h_r} \approx \begin{cases} \frac{1}{N_{\max}}, & \text{if } h_{\max} - h_r \le \beta \\ 0, & \text{otherwise} \end{cases} \tag{40}$$

where $N_{\max} \stackrel{\triangle}{=}$ number of terms h_i, satisfying the condition $h_{\max} - h_i \le \beta$, for $i = 1, 2, \ldots, s$, and the parameter β controls the width of the pulse

$$\frac{\partial h_{\min}}{\partial h_r} \approx \begin{cases} \frac{1}{N_{\min}}, & \text{if } h_r - h_{\min} \le \beta \\ 0, & \text{otherwise} \end{cases} \tag{41}$$

where $N_{\min} \stackrel{\triangle}{=}$ number of terms h_i, satisfying the condition $h_i - h_{\min} \le \beta$, for $i = 1, 2, \ldots, s$.

Now, the expression in (13) can be written as

$$P_{ij} = \frac{\partial \mu_C(c_j)}{\partial\{\mu_D(u_i) t \mu_{\Re}(r_{ij}\}} \cdot \frac{\partial\{\mu_D(u_i) t \mu_{\Re}(r_{ij})\}}{\partial \mu_{\Re}(r_{ij})}. \tag{42}$$

Comparing (27) with (15) we have $h_{\max} = \mu_C(c_j)$ and $h_i = \mu_D(u_i) \cdot \mu_{\Re}(r_{ij}), i = 1, 2, \ldots, m$. Using (40) in the above (42) where $t \equiv$ ‘prod' we get the derivative, P_{ij} of (15) as

$$P_{ij} = \begin{cases} \frac{\mu_D(u_i)}{N_{\max}}, & \text{if } \mu_C(c_j) - \{\mu_D(u_i) \cdot \mu_{\Re}(r_{ij})\} \le \beta \\ 0, & \text{otherwise} \end{cases} \tag{43}$$

where $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. These results are used only for the problems of Types and II of Section III.

Comparing (34) with (17), we have $h_{\min} = \mu_D(u_i) \wedge \mu_{\Re}(r_{ij})$. Now, there is only one variable in $h_{\min}$ as given above so $N_{\min} = 1$ only when $\mu_D(u_i) \ge \mu_{\Re}(r_{ij})$. Therefore, the derivative

$$\frac{\partial h_{\min}}{\partial \mu_{\Re}(r_{ij})} = \begin{cases} 1, & \text{if } \mu_D(u_i) \ge \mu_{\Re}(r_{ij}) \\ 0, & \text{otherwise.} \end{cases} \tag{44}$$

Using (44) in (42) where $t \equiv$ ‘min', we get the derivative, P_{ij} of (17) as

$$P_{ij} = \begin{cases} \frac{1}{N_{\max}} & \text{if } \mu_C(c_j) - \{\mu_D(u_i) \wedge \mu_{\Re}(r_{ij})\} \le \beta \\ & \text{and } \mu_D(u_i) \ge \mu_{\Re}(r_{ij}) \\ 0 & \text{otherwise.} \end{cases} \tag{45}$$

where $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. These results are used only for the problems of Types III and IV of Section III.

D. Algorithm for the Estimation of $\Re$

This algorithm gives the step-by-step calculation of $\Re$ using the modified computational approach.

Step 1) Start with an initial trial values of $\Re^{(o)} = [\mu_{\Re}(r_{ij})]_{m \times n}$ such that $\{\mu_{\Re}(r_{ij}) \wedge (1 - \mu_{\Re}(r_{ij}))\} \ge 0$.

Step 2) Set the width of the pulse β, convergent factor κ, the error threshold ϵ, and maximum number of iterations $s_{\max}$. Set the initial iteration number $s = 0$.

Step 3) Set new iteration number $s = s + 1$.

Step 4) Using the given fuzzy data D_l and $\tilde{C}_l$, evaluate C_l by (23) and E by (26).

Step 5) Evaluate $(\partial E)/(\partial \mu_{\Re}(r_{ij}))$, in (12) using either (43) (for the problems of Types III and IV) for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$ and α_s.

Step 6) Update the values of $\mu_{\Re}(r_{ij})$ using the Newton's iterative scheme [see (11)],

$$\mu_{\Re}(r_{ij})^{(s+1)} = \mu_{\Re}(r_{ij})^{(s)} - \alpha \cdot \left.\frac{\partial E}{\partial \mu_{\Re}(r_{ij})}\right| \Re = \Re^{(s)}$$

where $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.

TABLE I
FUZZY SETS $D_l^1, D_l^2, \tilde{C}_l, l = 1, 2, \ldots, 8$ FOR THE FUZZY SYSTEM

1	Membership values of fuzzy set D_l^1			Membership values of fuzzy set D_l^2				Membership values of fuzzy set $\tilde{C}_l$			
1	0.00	0.27	0.80	0.97	0.11	0.00	0.00	0.00	0.35	0.80	0.11
2	1.00	0.23	0.00	0.00	0.35	1.00	0.00	0.08	0.66	0.30	0.18
3	0.00	0.05	1.00	0.08	0.70	0.30	0.00	0.00	0.03	0.15	0.70
4	0.38	0.92	0.00	0.00	0.02	0.15	0.95	0.01	0.12	0.38	0.71
5	0.37	0.48	0.00	0.00	0.10	0.30	0.90	0.02	0.20	0.37	0.37
6	0.00	0.90	0.02	0.00	0.10	0.45	0.02	0.00	0.12	0.30	0.36
7	0.20	1.00	0.08	0.00	0.12	0.50	0.01	0.02	0.13	0.33	0.40
8	0.75	0.12	0.00	0.00	0.10	0.15	0.80	0.01	0.25	0.75	0.60

TABLE II
FUZZY SETS $D_l, l = 1, 2, \ldots, 8$ FOR TYPE I

1	Membership values of fuzzy set D_l											
1	0.00	0.00	0.00	0.00	0.27	0.11	0.00	0.00	0.80	0.11	0.00	0.00
2	0.00	0.35	1.00	0.00	0.00	0.23	0.23	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.05	0.05	0.05	0.00	0.08	0.70	0.30	0.00
4	0.00	0.02	0.15	0.38	0.00	0.02	0.15	0.92	0.00	0.00	0.00	0.00
5	0.00	0.10	0.30	0.37	0.00	0.10	0.30	0.48	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.10	0.45	0.02	0.00	0.02	0.02	0.02
7	0.00	0.12	0.20	0.01	0.00	0.12	0.50	0.01	0.00	0.08	0.08	0.01
8	0.00	0.10	0.15	0.75	0.00	0.10	0.12	0.12	0.00	0.00	0.00	0.00

TABLE III
FUZZY SETS $D_l, l = 1, 2, \ldots, 8$ FOR TYPE II

1	Membership values of fuzzy set D_l											
1	0.00	0.00	0.00	0.00	0.26	0.03	0.00	0.00	0.77	0.09	0.00	0.00
2	0.00	0.00	0.35	1.00	0.00	0.08	0.23	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.03	0.01	0.00	0.08	0.70	0.30	0.00
4	0.00	0.01	0.06	0.36	0.00	0.02	0.14	0.87	0.00	0.00	0.00	0.00
5	0.00	0.04	0.11	0.33	0.00	0.05	0.14	0.43	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.09	0.40	0.02	0.00	0.00	0.01	0.00
7	0.00	0.02	0.10	0.00	0.00	0.12	0.50	0.01	0.00	0.01	0.04	0.00
8	0.00	0.07	0.11	0.60	0.00	0.01	0.02	0.10	0.00	0.00	0.00	0.00

Step 7) Now test whether $\{\mu_{\Re}(r_{ij}) \wedge (1 - \mu_{\Re}(r_{ij}))\} \geq 0$, or not for all $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. If not, then construct a set of index pairs

$$NF = \{(i,j)/\{\mu_{\Re}(r_{ij}) \wedge (1 - \mu_{\Re}(r_{ij}))\} < 0, \quad i = 1, 2, \ldots, m \quad \text{and} \quad j = 1, 2, \ldots, n\}.$$

Set $\{\mu_{\Re}(r_{ij}) \wedge (1 - \mu_{\Re}(r_{ij}))\} = 0, \quad \forall (i,j) \epsilon NF.$

Step 8) Repeat from Step 3 until $E < \epsilon$ and / or $s = s_{\max}$.

E. Illustration of the Modified Approach to the Estimation of $\Re$

We illustrate the modified method based on the data set (see Table I) given by Pedrycz [2]. Here, $U = U_1 \times U_2$, $\text{card}(U_1) = m_1 = 3$, and $\text{card}(U_2) = m_2 = 4$. Therefore, $\text{card}(U) = m_1 \times m_2 = 12$. Now the membership values of $D_l \epsilon F(U)$, given by the formula $\mu_{D_l}(u_i) = (\mu_{D_l^1}(u_{i_1}^1) \wedge \mu_{D_l^2}(u_{i_2}^2))$, where $i = 4(i_1 - 1) + i_2, i_1 = 1, 2, 3$ and $i_2 = 1, 2, 3, 4$, are shown in Table II, and those of D_l obtained by the formula $\mu_{D_l}(u_i) = (\mu_{D_l^1}(u_{i_1}^1) \cdot \mu_{D_l^2}(u_{i_2}^2))$, where $i = 4(i_1 - 1) + i_2, i_1 = 1, 2, 3$, and $i_2 = 1, 2, 3, 4$ are shown in Table III. We start with an initial trial values $\mu_{\Re}(r_{ij}) = 0, i = 1, 2, \ldots, 12$ and $j = 1, 2, 3, 4$. The value $\kappa = 10^{-4}$ is chosen to ensure good convergence properties. The width of the pulse is $\beta = 0.05$. The error threshold is

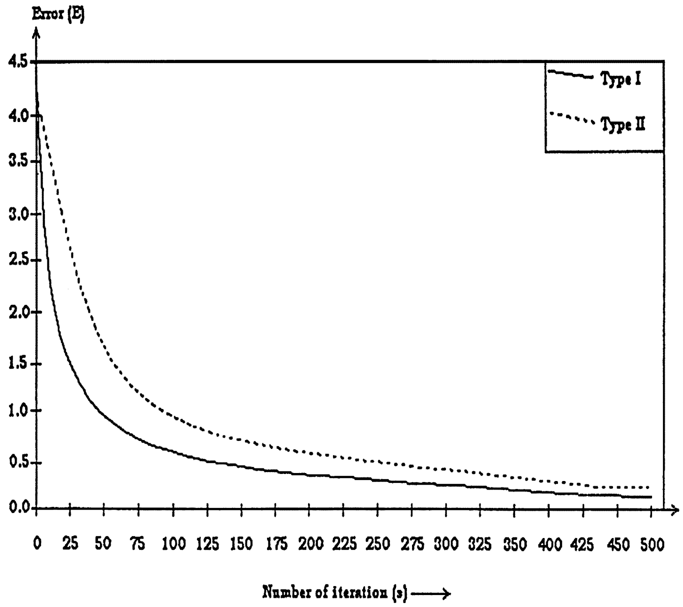


Fig. 2. Squared error (E) against each 25 iteration (s).

TABLE IV
SOLUTIONS OF RELATION $\Re$ OF PROBLEMS TYPE I AND TYPE II

0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.02	0.03	0.03	0.03	0.02	0.04	0.03	0.02
0.08	0.67	0.30	0.18	0.08	0.67	0.30	0.14
0.01	0.33	1.00	0.80	0.02	0.43	1.00	1.00
0.00	0.02	0.04	0.01	0.00	0.03	0.03	0.01
0.01	0.05	0.05	0.04	0.01	0.02	0.03	0.03
0.01	0.21	0.67	0.80	0.02	0.26	0.70	0.83
0.01	0.09	0.19	0.78	0.02	0.12	0.43	0.82
0.00	0.44	1.00	0.07	0.00	0.45	1.00	0.14
0.00	0.02	0.04	1.00	0.00	0.01	0.03	1.00
0.00	0.01	0.01	0.05	0.00	0.00	0.01	0.05
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Type I				**Type II**		

$\epsilon = 10^{-3}$, and the maximum number of iterations is $s_{\max} = 500$. The values of the error, calculated every 25 steps of iterations, are displayed in Fig. 2. The solutions of $\Re$s, of the problems of Types I II of Section III, are shown in Table IV.

V. DESIGN OF THE CLASSIFIER BASED ON FUZZY RELATIONAL CALCULUS (FRC)

In classifier design (see Fig. 3), two phases exist, namely, the learning phase (training phase), where we estimate the fuzzy relation $\Re$ based on the algorithm of section IV-D, and the testing phase (classification phase), where we test the performance of the classifier using (3) which inovlves the expression $\Re$.

At the beginning of the training phase, we discretize (quantize) the individual feature axis and the entire pattern space in the following way.

Determine the lower and upper bounds of the data of ith feature value. Let f_j^i be the jth data of the ith feature F_i, and let d_i be the length of segmentation along ith feature axis.

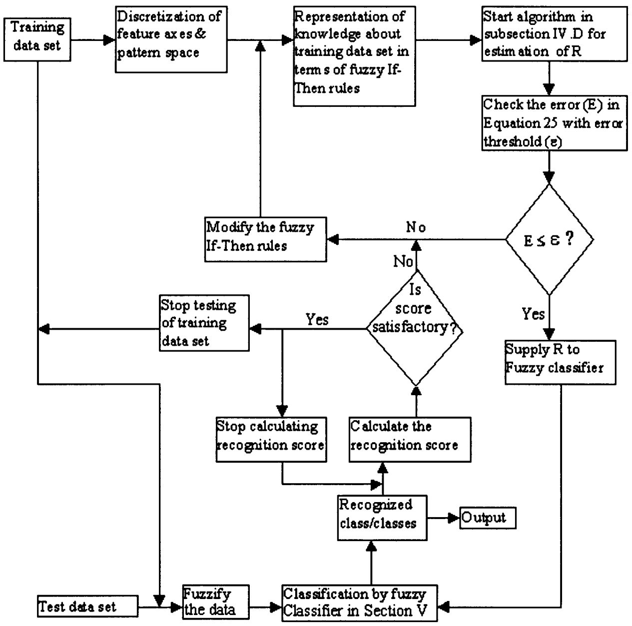


Fig. 3. Classifier based on fuzzy relational calculus.

Minimum of the data $f_j^i, j = 1, 2, \ldots$ of the ith feature is $f_{\min}^i = \min_j(f_j^i)$. Let $r_{\min}^i$ be the remainder when $f_{\min}^i$ is divided by d_i. Therefore, the lower bound of the ith feature axis is

$$LB_i = \begin{cases} f_{\min}^i, & \text{if } r_{\min}^i = 0 \\ f_{\min}^i - r_{\min}^i, & \text{otherwise.} \end{cases} \quad (46)$$

This LB_i is taken as the ith coordinate of the origin.

Again, maximum of the data $f_j^i, j = 1, 2, \ldots$ of the ith feature is $f_{\max}^i = \max_j(f_j^i)$. Let $r_{\max}^i$ be the remainder when $f_{\max}^i$ is divided by d_i. Therefore, the upper bound of the ith feature axis is

$$UB_i = \begin{cases} f_{\max}^i, & \text{if } r_{\max}^i = 0 \\ f_{\max}^i + (d_i - r_{\max}^i), & \text{otherwise.} \end{cases} \quad (47)$$

Let U_i be the universe of discourse on the ith feature axis F_i; then, U_i has $m_i = (UB_i - LB_i)/d_i$ generic elements and these are $u_j^i, j = 1, 2, \ldots, m_i$, which we define as follows

$$u_j^i = \begin{cases} [LB_i + (j-1)\cdot d_i, LB_i + j \cdot d_i)] \\ \quad \text{for } j = 1, 2, \ldots, (m_i - 1) \\ [LB_i + (m_i - 1)\cdot d_i, UB_i] \\ \quad \text{for } j = m_i. \end{cases} \quad (48)$$

Let the universe on the ith feature axis $U_i = \{u_1^i, u_2^i, \ldots, u_{m_i}^i\}$. Let the Cartesian product space of the universe $U_i, i = 1, 2, \ldots, c$ be U, i.e., $U = U_1 \times U_2 \times \cdots \times U_c$ having elements each of type $u_i = (u_{i_1}^1, u_{i_2}^2, \ldots, u_{i_c}^c / u_{i_p}^p \epsilon U_p, p = 1, 2, \ldots, c)$, where $i = \sum_{p=1}^{c-1} (\Pi_{l>p}^c m_l)(i_p - 1) + i_c$ for each $i_p = 1, 2, \ldots, m_p, p = 1, 2, \ldots, c$.

Now, we define k_i fuzzy sets on U_i, say, $D_j^i, j = 1, 2, \ldots, k_i$ which are in shown in Table V. So there are $k = \Pi_{j=1}^c k_j$ fuzzy **If-Then** rules as follows:

R_l: If F_1 is $D_{j_1}^1$ and F_2 is $D_{j_2}^2$ and $\ldots\ldots F_c$ is $D_{j_c}^c$, then C is $C_l \epsilon F(C_{\text{class}})$, where $l = \sum_{p=1}^{c-1} (\Pi_{q>p}^c k_q)(j_p - 1) + j_c$ for each $j_p = 1, 2, \ldots, k_p, p = 1, 2, \ldots, c$, and C_{class} is the universe of discourse constructed by all the classes in the pattern space, i.e., $C_{\text{class}} = \{c_1, c_2, \ldots, c_n\}$.

TABLE V
FUZZY SETS IN $F(U_i)$

$(*)$	u_1^i	u_2^i	....	$u_{m_i}^i$
D_1^i	$\mu_{D_1^i}(u_1^i)$	$\mu_{D_1^i}(u_2^i)$	....	$\mu_{D_1^i}(u_{m_i}^i)$
D_2^i	$\mu_{D_2^i}(u_1^i)$	$\mu_{D_2^i}(u_2^i)$	....	$\mu_{D_2^i}(u_{m_i}^i)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$D_{k_i}^i$	$\mu_{D_{k_i}^i}(u_1^i)$	$\mu_{D_{k_i}^i}(u_2^i)$	....	$\mu_{D_{k_i}^i}(u_{m_i}^i)$

TABLE VI
FUZZY SETS IN $F(C_{\text{class}})$ FOR $\circ\epsilon$\{'min', 'prod'\}

$(\circ)$	D_1^2	D_2^2	....	$D_{k_2}^2$
D_1^1	C_{11}	C_{12}	....	C_{1k_2}
D_2^1	C_{21}	C_{22}	....	C_{2k_2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$D_{k_1}^1$	$C_{k_1 1}$	$C_{k_1 2}$	....	$C_{k_1 k_2}$

If D_l is the fuzzy set which is a fuzzy pattern vector (see Definition 7) formed by the antecedent clauses of the rule R_l, i.e., $D_l \epsilon F(U)$, then the membership value of the belongingness of u_i in D_l is determined by (8). According to the fuzzy implication method, we write $\Re : F(U_1) \times F(U_2) \times \cdots \times F(U_c) \to F(C_{\text{class}})$.

The membership value of the class $c_p \epsilon C_{\text{class}}$ when $D_{j_1}^1$ is on $U_1, D_{j_2}^2$ is on U_2, etc., is taken in the following way:

$$\mu_{C_l}(C_p) = \bigvee_{u_i \epsilon c_p \cap FZ(D_{j_1}^1, D_{j_2}^2, \ldots, D_{j_c}^c)} (\mu_{D_l}(u_i)) \quad (49)$$

where $FZ(D_{j_1}^1, D_{j_2}^2, \ldots, D_{j_c}^c)$ is the zone which represents the tip of the fuzzy pattern vector (see Fig. 7 in Appendix B) and is constructed by the fuzzy sets $D_{j_1}^1, D_{j_2}^2, \ldots, D_{j_c}^c$ of the rule R_l, where $j_p = 1, 2, \ldots, k_p$ for each $p = 1, 2, \ldots, c$.

For two-dimensional (2–D) pattern space, we may construct the rules in the compact form as shown in Table VI.

$R_{i_1 i_2}$: If F_1 is $D_{i_1}^1$ and F_2 is $D_{i_2}^2$, then C is $C_{i_1 i_2} \epsilon F(C_{\text{class}})$ and the membership value of each class $c_p, p \epsilon \{1, 2, \ldots, n\}$ of the fuzzy set $C_{i_1 i_2}$, where $i_l = 1, 2, \ldots, k_l, l = 1, 2$ will be determined by(49). Based on the generated fuzzy rules as stated above, we estimate the fuzzy relation $\Re$ at the end of training phase using the algorithm of the section IV-D. In the course of estimating $\Re$, if the error given by (26) does not reach the desired threshold, even after a sufficient number of iterations, we may have to modify the initial fuzzy **if-then** rules to represent our knowledge about the training data set. On the other hand, after reaching the error threshold, we cross-verify the quality of the estimated $\Re$ by checking the classification score of the training data set (based on which the fuzzy **if-then** rules were initially generated for estimating $\Re$). If the classification score of the training data set (which are fuzzified by fuzzy masking at the time of testing) does not reach the satisfactory threshold (say 80% recognition score is set as threshold), we may have to modify the initial fuzzy **if-then** rules to represent our knowledge about the training data set. After satisfactory estimation

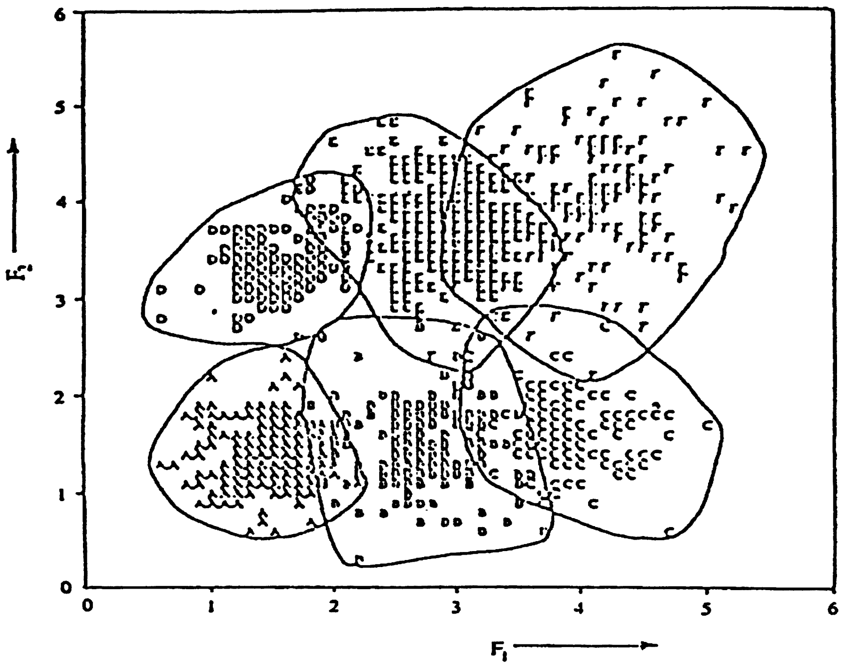


Fig. 4. First synthetic data.

of $\Re$, we switch over to the testing phase (classification phase), where we consider the classification of data which does not belong to the training data set.

At the testing phase (classification phase), we use (3), as stated in Section II. The features of the selected patterns are fuzzified using the concept of the fuzzy masking. The classification results obtained from (3) produces a fuzzy set $C = \sum_{i=1}^{n} \mu C(c_i)/c_i$, which represents the degree of occurrence of each testpatten at different classes in the quantized pattern space. We, thus, get a fuzzy classification of a testpatten. To calculate the recognition score from the above result, we have to go through a certain decision process. In the first stage of our decision process, we increase the level of confidence by prescribing a α-cut of the fuzzy set C, i.e.,

$$C_\alpha = \{c_i/\mu_C(c_i) \geq \alpha; \quad c_i \epsilon C_{\text{class}}\}.$$

If $C_\alpha = 0$ (empty set), then the given testpatten is not recognized by the present classifier. Otherwise;

$$hgt(C) = \bigvee_{c_i \epsilon c_\alpha} \mu_C(c_i).$$

Now, we get the set of recognized classes as

$$\text{Class}_{\text{recognize}} = \{c_i/hgt(C) - \mu_C(c_i) \leq \theta, c_i \epsilon C_\alpha\}$$

where θ is a small threshold prescribed by the designer to capture the relative change in membership values among the elements of the recognized classes $\text{Class}_{\text{recognize}}$.

i) In case $\text{Class}_{\text{recognize}}$ is a singleton set, then the given testpatten is recognized uniquely. ,
ii) Otherwise, multiple classifications of the given testpatten occur.

The notion of multiple classification is very natural in the case of testpattens occurring at overlapped classes. Such choice of multiple classifications sometimes stands as a kind of grace, to take care of all uncertainties (e.g., uncertainties in the representation of knowledge about training patterns, uncertainties in the process of fuzzification, through fuzzy masking, of the testpattens etc.) in our classification process.

VI. Effectiveness of the Proposed Method

To test the effectiveness of our design, as stated in Section V, we consider the classification of two synthetic data as shown in Figs. 4 and 5. At the time of writing fuzzy **If-Then** rules for the classifier, we may consider complete cover of the pattern space (see Appendix B), but as the consideration of complete cover of the pattern space does not bring any significant change in classification score, for practical purposes, without loss of generality, we consider partial cover of the pattern space.

A. Classification of First Synthetic Data

For the data shown in Fig. 4, we choose length of segmentations $d_1 = 0.5 = d_2$. Therefore, we get $LB_1 = 0 = LB_2$ by (46) and $UB_1 = 6 = UB_2$ by (47). Thus, $m_1 = (UB_1 - LB_1)/d_1 = 12$ and $m_2 = (UB_2 - LB_2)/d_2 = 12$.

1) For the Problem of Type I of Section III: We define $k_1 = 4$ fuzzy sets on U_1 and $k_2 = 3$ fuzzy sets on U_2 which are shown in Tables VII and VIII respectively and $k = k_1 \times k_2 = 12$ fuzzy **If-Then** rules and their consequent parts are shown in Table IX.

Now we start with initial trial values of $\mu_\Re(r_{ij}) = 0, i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n, k = 10^{-3}, \beta = 0.05, \epsilon = 10^{-2}, S_{\max} = 500$ and terminate the iteration scheme at $S_{\max}$. The classification scores are shown in Table X.

2) For the Problem of Type II of Section III: We define $k_1 = 4$ fuzzy sets on U_1 and $k_2 = 4$ fuzzy sets on U_2, so we can find $k = k_1 \times k_2 = 16$ fuzzy **If-Then** rules.

Now we start with initial trial values of $\mu_\Re(r_{ij}) = 0, i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n, \kappa = 10^{-3}, \beta = 0.05, \epsilon =$

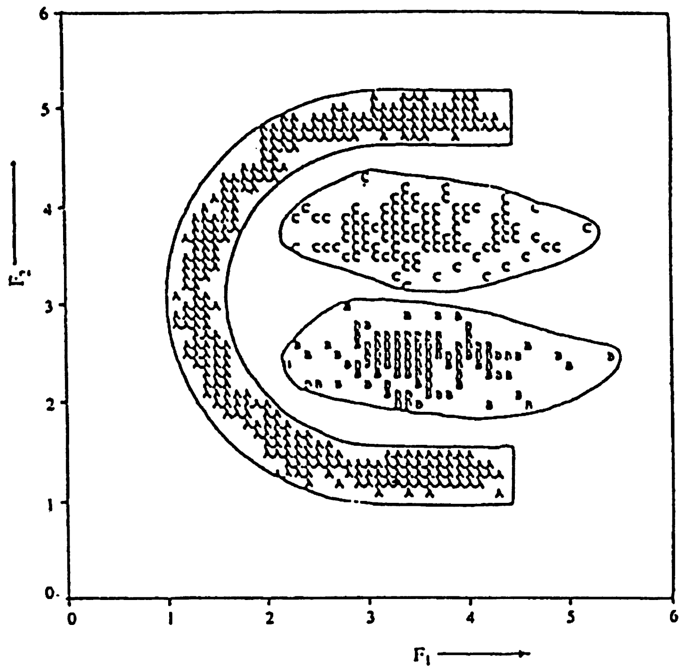


Fig. 5. Second synthetic data.

TABLE VII
FUZZY SETS IN $F(U_1)$ FOR THE FIRST SYNTHETIC DATA FOR THE PROBLEM OF TYPE I

(*)	u_1^1	u_2^1	u_3^1	u_4^1	u_5^1	u_6^1	u_7^1	u_8^1	u_9^1	u_{10}^1	u_{11}^1	u_{12}^1
D_1^1	0.3	0.7	1.0	0.7	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0
D_2^1	0.0	0.0	0.1	0.3	0.7	1.0	0.7	0.3	0.1	0.0	0.0	0.0
D_3^1	0.0	0.0	0.0	0.0	0.1	0.3	0.7	1.0	0.7	0.3	0.1	0.0
D_4^1	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.3	0.7	1.0	0.7	0.3

TABLE VIII
FUZZY SETS IN $F(U_2)$ FOR THE FIRST SYNTHETIC DATA FOR THE PROBLEM OF TYPE I

(*)	u_1^2	u_2^2	u_3^2	u_4^2	u_5^2	u_6^2	u_7^2	u_8^2	u_9^2	u_{10}^2	u_{11}^2	u_{12}^2
D_1^2	0.3	0.7	1.0	0.7	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0
D_2^2	0.0	0.0	0.0	0.1	0.3	0.7	1.0	0.7	0.3	0.1	0.0	0.0
D_3^2	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.3	0.7	1.0	0.7	0.3

$10^{-2}, S_{\max} = 500$ and terminate the iteration scheme at $S_{\max}$. The classification scores are shown in Table XI.

B. Classification of Second Synthetic Data

For the data shown in Fig. 5, we choose length of segmentations $d_1 = 0.5 = d_2$. Therefore, we get, $LB_1 = 0 = LB_2$ by (46) and $UB_1 = 6 = UB_2$ by (47). Thus, $m_1 = (UB_1 - LB_1)/d_1 = 12$ and $m_2 = (UB_2 - LB_2)/d_2 = 12$. By (48), we get

$$U_1 = \{u_1^1, u_2^2, \ldots, u_{12}^1\}$$
$$U_2 = \{u_1^2, u_2^2, \ldots, u_{12}^2\}.$$

For both of the problems of Types I and II of Section III, we define $\kappa_1 = 3$ fuzzy sets on U_1 and $\kappa_2 = 4$ fuzzy sets on U_2 so we can find $\kappa = \kappa_1 \times \kappa_2 = 12$ fuzzy **If-Then** rules.

Now, for both the problems, we start with initial trial values of $\mu_{\Re}(r_{ij}) = 0, i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n, \kappa =$

TABLE IX
FUZZY SETS IN $F(C_{\text{class}})$ FOR THE FIRST SYNTHETIC DATA FOR THE PROBLEM OF TYPE I

Rule	Antecedent part			Consequent part					
l	F_1	and	F_2	c_1	c_2	c_3	c_4	c_5	c_6
R_1	D_1^1	and	D_1^2	1.00	0.70	0.00	0.10	0.10	0.00
R_2	D_1^1	and	D_2^2	0.30	0.30	0.00	1.00	0.70	0.10
R_3	D_1^1	and	D_3^2	0.00	0.00	0.00	0.70	0.70	0.10
R_4	D_2^1	and	D_1^2	0.70	1.00	0.70	0.10	0.30	0.30
R_5	D_2^1	and	D_2^2	0.30	0.70	0.70	0.70	1.00	1.00
R_6	D_2^1	and	D_3^2	0.00	0.00	0.00	0.70	1.00	0.70
R_7	D_3^1	and	D_1^2	0.10	1.00	1.00	0.10	0.30	0.30
R_8	D_3^1	and	D_2^2	0.10	0.70	0.70	0.10	1.00	1.00
R_9	D_3^1	and	D_3^2	0.00	0.00	0.00	0.10	0.70	1.00
R_{10}	D_4^1	and	D_1^2	0.00	0.30	1.00	0.00	0.10	0.30
R_{11}	D_4^1	and	D_2^2	0.00	0.10	0.70	0.00	0.30	1.00
R_{12}	D_4^1	and	D_3^2	0.00	0.00	0.00	0.00	0.30	1.00

TABLE X
CLASSIFICATION SCORES OF FIRST SYNTHETIC DATA FOR THE PROBLEM OF TYPE I

	To						Number of	Recognition
From	A	B	C	D	E	F	data	score(%)
A	109	2	0	0	0	0	116	93.97
B	7	105	16	0	1	1	117	89.74
C	0	71	71	0	1	1	79	89.87
D	0	0	0	77	10	9	89	86.52
E	0	28	0	1	135	98	151	89.40
F	0	2	0	0	64	111	115	96.52
Total	109	105	71	77	135	111		91.15

TABLE XI
CLASSIFICATION SCORES OF FIRST SYNTHETIC DATA FOR THE PROBLEM OF TYPE II

	To						Number of	Recognition
From	A	B	C	D	E	F	data	score(%)
A	114	6	0	5	4	4	116	93.28
B	10	110	18	2	3	3	117	94.02
C	0	73	73	0	1	1	79	92.41
D	0	0	0	79	12	11	89	88.76
E	0	28	0	3	137	117	151	90.73
F	0	3	1	0	65	113	115	98.26
Total	114	110	73	79	137	113		93.85

$10^{-3}, \beta = 0.05, \epsilon = 10^{-2}, S_{\max} = 500$ and terminate the iteration scheme at $S_{\max}$. The classification scores are shown in Table XII.

VII. APPLICATIONS

After achieving satisfactory results on a synthetic set of data, we apply the proposed design for the vowel classification problem of an Indian language, namely Telugu [24]. In the following subsections, we discuss the classification results.

For the data shown in Fig. 6, we choose length of segmentations $d_1 = 50$ and $d_2 = 100$. Therefore, we get, $LB_1 = 200$ and $LB_2 = 600$ by (46) and $UB_1 = 850$ and $UB_2 = 2600$ by (47). Thus, $m_1 = (UB_1 - LB_1)/d_1 = 13$ and $m_2 = (UB_2 - LB_2)/d_2 = 20$. By (48), we get

$$U_1 = \{u_1^1, u_2^1, \ldots, u_{13}^1\}$$
$$U_2 = \{u_1^2, u_2^2, \ldots, u_{20}^2\}.$$

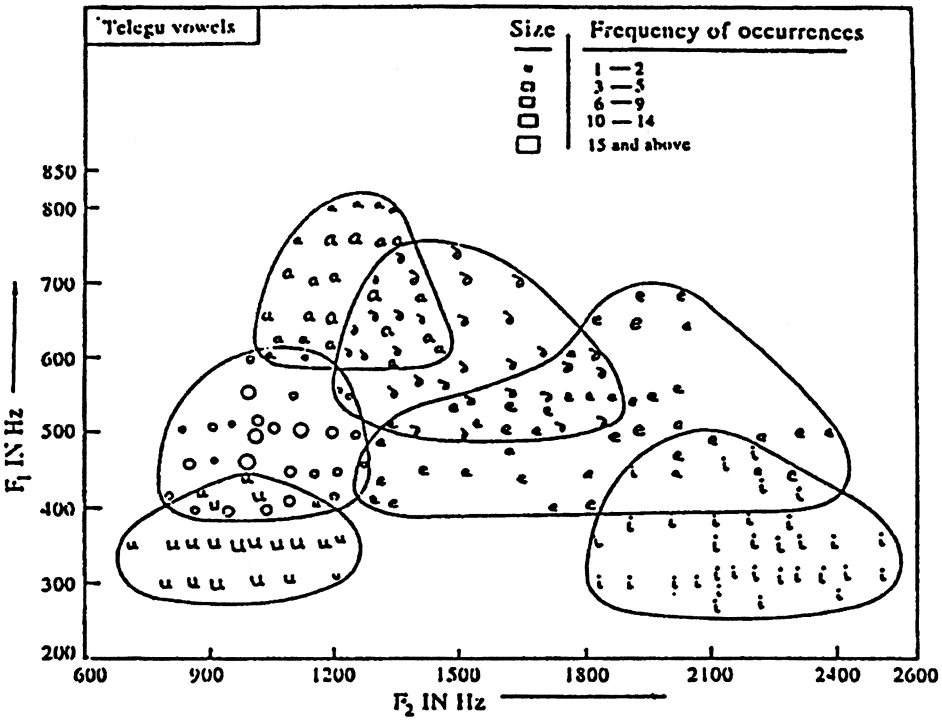


Fig. 6. Telegu vowels.

TABLE XII
CLASSIFICATION SCORES OF SECOND SYNTHETIC DATA

From	To A	To B	To C	Number of data	Recognition score(%)
A	287	13	6	287	100.00
B	8	96	0	97	98.97
C	12	0	99	100	99.00
Total	287	96	99		99.59

Problem of Type I

From	To A	To B	To C	Number of data	Recognition score(%)
A	283	43	38	287	98.61
B	2	96	0	97	98.97
C	3	0	99	100	99.00
Total	283	96	99		98.76

Problem of Type II

TABLE XIII
CLASSIFICATION SCORES OF TELUGU VOWEL FOR THE PROBLEM OF TYPE I

From	To a	e	i	o	u	δ	Number of data	Recognition score(%)
a	82	4	0	29	0	74	83	98.80
e	7	193	52	21	7	82	200	96.50
i	0	48	119	0	0	0	133	89.47
o	15	26	0	114	72	30	116	98.28
u	0	10	0	48	107	2	112	95.54
δ	41	32	0	19	1	65	66	98.48
Total	82	193	119	114	107	65		95.77

For both the problems of Types I and II of Section III, we define $\kappa_1 = 5$ fuzzy sets on U_1 and $\kappa_2 = 7$ fuzzy sets on U_2, so we can find $\kappa = \kappa_1 \times \kappa_2 = 35$ fuzzy **If-Then** rules.

Now, for both the problems, we start with initial trial values of $\mu_{\Re}(r_{ij}) = 0, 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n, \kappa = 10^{-3}, \beta =$

TABLE XIV
CLASSIFICATION SCORES OF TELUGU VOWEL FOR THE PROBLEM OF TYPE II

From	To a	e	i	o	u	∂	Number of data	Recognition score(%)
a	82	4	0	29	0	74	83	98.80
e	7	196	52	21	7	83	200	98.00
i	0	50	126	0	0	0	133	94.74
o	16	26	0	114	72	30	116	98.28
u	0	10	0	48	108	2	112	96.43
∂	40	32	0	19	1	65	66	98.48
Total	82	196	126	114	108	65		97.32

TABLE XV
COMPARATIVE STUDY

Different types of Classifiers	Recognition Score (%) First synthetic data (Figure 4)	Second synthetic data (Figure 5)	Vowel data (Figure 6)
Bayesian	78.0	80.9	80.3
Fuzzy c-means	80.3	82.9	80.5
Conventional multilayer perception (MLP)	88.15	78.0	90.0
Present method with max-min operator	91.15	99.59	95.77
Present method with max-product operator	93.85	98.76	97.32

$0.05, \epsilon = 10^{-2}, S_{\max} = 1000$, and terminate the iteration scheme at $S_{\max}$. The classification scores are shown in Tables XIII and XIV.

VIII. COMPARATIVE STUDY

In Table XV we have compared the performance (in terms of recognition score) of the present classifier with those of some existing ones. The results shown in Table XV indicate that the performance of the present design of the classifier is comparable with those of some existing ones.

IX. Conclusion

In this paper, we consider a particular interpretation [i.e., (a) of (1)] of MFI and introduce a notion of fuzzy pattern vector which represents the antecedent part of the interpretation a) of (1). The advantage of considering such notion is two-fold. First, we can describe a population of training patterns by linguistic features. Second, the notion of fuzzy pattern vector helps us formulate the consequent part of a) of (1) (see Example 2 of Appendix B). We develop a new approach to the computation of the derivative of the fuzzy max/min function. A detail design of pattern classifier based on FRC is developed and very promising results are obtained. We compute the performance of the present classifier with those of some existing classifiers and get satisfactory response. A neural net version of the present design to estimate the fuzzy relation $\Re$ (for classification problem) would be the scope for future work. In the present design study we have only considered the problems of Types I and II of Section III. Similar results are also obtainable for the problems of Types III and IV.

Appendix A

Good Function

Definition 1: A function $\gamma(x)$ is said to be a good function if it is infinitely many differentiable everywhere on R and if

$$|x|\overset{\lim}{\rightarrow}\infty^{|x^r \frac{d^\kappa \gamma(x)}{dx^\kappa}|=0}$$

for every integer $\kappa \geq 0$ and every integer $r \geq 0$.

The function $\gamma(x) = e^{-x^2}$ is a good function.

A good function has the following properties:

1) If $\gamma_1(x)$ and $\gamma_2(x)$ are good, then $\gamma_1(x) + \gamma_2(x)$ and $\gamma_1(x)\gamma_2(x)$ are also good.
2) If $\gamma(x)$ is good, then $(d\gamma(x)/dx)$ is also good.
3) If $\gamma(x)$ is good, then $\gamma(ax+b)$, where a and b are real constants, is also good.

Fairly Good Function

Definition 2: A function $\psi(x)$ is said to be a fairly good function if it is infinitely many differentiable everywhere on R and if there is a some fixed N such that

$$|x|\overset{\lim}{\rightarrow}\infty^{|x^{-N} \frac{d^\kappa \Psi(x)}{dx^\kappa}|=0}$$

for every integer $\kappa \geq 0$.

A simple example of a fairly good function is e^{ix}, but e^x is not a fairly good function.

A fairly good function has the following properties:

1) If $\psi_1(x)$ and $\psi_2(x)$ are fairly good, then $\psi_1(x) + \psi)2(x)$ and $\psi_1(x)\psi_2(x)$ are also fairly good.
2) If $\psi(x)$ is fairly good, then $(d\psi(x)/dx)$ is also fairly good.
3) If $\psi(x)$ is fairly good, then $\psi(ax+b)$, where a and b are real constants, is also fairly good.

Generalized Function

We first give the following definitions

Definition 3: A sequence $\{\gamma_n(x)\}$ of good functions is said to be regular if for every given good function $\gamma(x)$, such that

$$n\overset{\lim}{\rightarrow}\infty \int_{-\infty}^{\infty} \gamma_n(x)\gamma(x)\,dx$$

exists and is finite

$$n\overset{\lim}{\rightarrow}\infty \int_{-\infty}^{\infty} \gamma_n(x)\gamma(x)\,dx = \int_{-\infty}^{\infty} \gamma(x)\,dx.$$

The sequence $\{e^{-(x^2/n)}\}$ of good functions is regular.

Definition 4: Two regular sequences $\{\alpha_n(x)\}$ and $\{\beta_n(x)\}$ are said to be equivalent if and only if

$$n\overset{\lim}{\rightarrow}\infty \int_{-\infty}^{\infty} \alpha_n(x)\gamma(x)\,dx = n\overset{\lim}{\rightarrow}\infty \int_{-\infty}^{\infty} \beta_n(x)\gamma(x)\,dx.$$

The two sequences $\{e^{-(x^2/n}\}$ and $\{e^{-(x^2/2n}\}$ of good functions are regular and equivalent.

Definition 5: An equivalence class of regular sequences is a generalized function.

A conventional notation is to write g as a generalized function associated with the equivalence class of which $\{\gamma_n(x)\}$ is a typical member. Now, we write

$$n\overset{\lim}{\rightarrow}\infty \int_{-\infty}^{\infty} \gamma_n(x)\gamma(x)\,dx = \oint_{-\infty}^{\infty} g(x)\gamma(x)\,dx.$$

The $\oint$ emphasizes that a limiting pocess is involved and that the quantity on the righthand side is not an ordinary integral. Later on, when certain properties have been established, it will be found reasonable to replace $\oint$ by $\int$.

Let $\{\alpha_n(x)\}$ and $\{\beta_n(x)\}$ are two regular sequences defining the generalized function g_1 and g_2, respectively. Thus

$$n\overset{\lim}{\rightarrow}\infty \int_{-\infty}^{\infty} \alpha_n(x)\gamma(x)\,dx = \oint_{-\infty}^{\infty} g_1(x)\gamma(x)\,dx.$$
$$n\overset{\lim}{\rightarrow}\infty \int_{-\infty}^{\infty} \beta_n(x)\gamma(x)\,dx = \oint_{-\infty}^{\infty} g_2(x)\gamma(x)\,dx.$$

If the above regular sequences are equivalent, then the above limits are equal. Consequently, we have:

$$\oint_{-\infty}^{\infty} g_1(x)\gamma(x)\,dx = \oint_{-\infty}^{\infty} g_2(x)\gamma(x)\,dx \quad \text{iff } g_1(x) = g_2(x). \tag{50}$$

Proposition 1: The sequence $\{\sqrt{(n/\pi)}e^{-nx^2}\}$ is regular. Define a generalized function, denoted by $\delta(x)$ such that

$$\oint_{-\infty}^{\infty} \delta(x)\gamma(x)\,dx = \gamma(0).$$

Proof: Now, we have

$$\int_{-\infty}^{\infty} e^{-nx^2}\gamma(x)\,dx = \gamma(0)\int_{-\infty}^{\infty} e^{-nx^2}\,dx + \int_{-\infty}^{\infty} \{\gamma(x)-\gamma(0)\}e^{-nx^2}\,dx \tag{51}$$

again by substituting $z = nx^2$, we get

$$\int_{-\infty}^{\infty} e^{-nx^2}\,dx = \int_0^{\infty} \frac{e^{-z}}{\sqrt{nz}}\,dz = \frac{1}{\sqrt{n}}\int_0^{\infty} e^{-z} z^{-1/2}\,dz = \frac{1}{\sqrt{n}}\tau(1/2) = \sqrt{\frac{\pi}{n}}. \tag{52}$$

Also, by mean value theorem of integral calculus, we have

$$|\gamma(x) - \gamma(0)| = \left|\int_0^x \frac{d\gamma(x)}{dx}\,dx\right| \le M|x|. \tag{53}$$

Here, $\gamma(x)$ is a good function, so $(d\gamma(x)/dx)$ is also good and is bounded by M (say). Using (53), we get

$$\left|\int_{-\infty}^{\infty}(\gamma(x)-\gamma(0))e^{-nx^2}\,dx\right| \le 2M\int_0^{\infty} xe^{-nx^2}\,dx = \frac{M}{n}[-e^{-nx^2}]_0^{\infty} = \frac{M}{n}. \tag{54}$$

Using (52) and (54) in (51), we get

$$\int_{-\infty}^{\infty}\sqrt{\frac{n}{\pi}}e^{-nx^2}\gamma(x)\,dx = \gamma(0) + 0\left(\frac{M}{\sqrt{n}}\right)$$

and taking limit, we get

$$\lim_{n\to\infty}\int_{-\infty}^{\infty}\sqrt{\frac{n}{\pi}}e^{-nx^2}\gamma(x)\,dx = \gamma(0).$$

Therefore, the given sequence is regular by Definition 3 and defines a generalized fnction $\delta(x)$ such that

$$\lim_{n\to\infty}\int_{-\infty}^{\infty}\sqrt{\frac{n}{\pi}}e^{-nx^2}\gamma(x)\,dx = \oint_{-\infty}^{\infty}\delta(x)\gamma(x)\,dx = \gamma(0).$$

Definition 6: The sequence $\{((d\gamma_n(x))/dx)\}$ is regular and defines a generalized function, which will be denoted by $g'(x)$ and called the derivative of g.

The sequence $\{((d\gamma_n(x))/dx)\}$ is regular because

$$\int_{-\infty}^{\infty}\frac{d\gamma_n(x)}{dx}\gamma(x)\,dx = [\gamma(x)\gamma_n(x)]_{-\infty}^{\infty} - \int_{-\infty}^{\infty}\gamma_n(x)\frac{d\gamma(x)}{dx}\,dx$$

but $\gamma_n(x)\gamma(x)$, being a good function, vanishes as $x \to \pm\infty$ and $(d\gamma(x)/dx)$ is a good function so that

$$\lim_{n\to\infty}\int_{-\infty}^{\infty}\frac{d\gamma_n(x)}{dx}\gamma(x)\,dx = -\oint_{-\infty}^{\infty} g(x)\frac{d\gamma(x)}{dx}\,dx.$$

This also demonstrates that equivalent $\{\gamma_n(x)\}$ give equivalent $((d\gamma_n(x))/dx)$, so we can write

$$\oint_{-\infty}^{\infty} g'(x)\gamma(x)\,dx = -\oint_{-\infty}^{\infty} g(x)\frac{d\gamma(x)}{dx}\,dx.$$

Calculation of the Derivative of Heaviside Function $H(x)$

The *Heaviside function* is first defined insection IV-A [see (29)]. Our objective is to calculate the the derivative of this function. By using Definition 6, we have

$$\oint_{-\infty}^{\infty} H'(x)\gamma(x)\,dx = -\oint_{-\infty}^{\infty} H(x)\gamma'(x)\,dx$$

since $\gamma(x)$ is good, $\gamma(x)H(x)$ is a good function which vanishes as $x \to \pm\infty$.

Hence

$$\oint_{-\infty}^{\infty} H'(x)\gamma(x)\,dx = -\int_{0+}^{\infty}\gamma'(x)\,dx = \gamma(0+).$$

Since $\gamma(x)$ is continuous everywhere on $[-\infty,\infty]$, we have $\gamma(0+) = \gamma(0) = \gamma(0-)$.

Using Proposition 1 and (50), we get

$$\oint_{-\infty}^{\infty} H'(x)\gamma(x)\,dx = \gamma(0) = \oint_{-\infty}^{\infty}\delta(x)\gamma(x)\,dx \Rightarrow H'(x) = \delta(x). \tag{55}$$

We use this result in sections IV-A [see (31) and (32)] and IV-B [see (37) and (38)].

APPENDIX B

Multidimensional Fuzzy Implication and the Notion of Fuzzy Pattern Vector

For simplicity of discussion and/or demonstration, let us restrict ourselves to the problem of pattern classification on R^2. Without lack of any generality, all the discussions and/or demonstrations are valid for the problem of pattern classification on $R^c, c > 2$. Let us now give a brief discussion on the correspondence between the conventional approach and the FRC approach to pattern classification.

In the conventional approach, the position of each pattern (say P) on the finite range of pattern space is represented by a pattern vector $\vec{F_c}$, and we always try to discriminate among patterns by classifying the pattern vector $\vec{F_c}$ (see Fig. 7).

Therefore, from the given data set (i.e., the training data set), we always know where the patterns are located and then try to separate them by some appropriate decision function. Subsequently, we use the said decision function to classify the test-patten vectors.

If we try to mimick the cognitive process of human reasoning for pattern classification; however, then, the first problem is to represent (from a given set of imprecise observations stated in terms of fuzzy **if-then** rules) the patterns on the pattern space in an appropriate fashion so that we can develop a suitable inferencing technique for classification.

Since a multidimensional fuzzy implication (MFI) [26] such as "if (x is A, y is B) then z is C" where A, B, C are fuzzy sets, is not merely a collection of 1–D implications, a conventional interpretation is usually taken in the multidimensional case. For

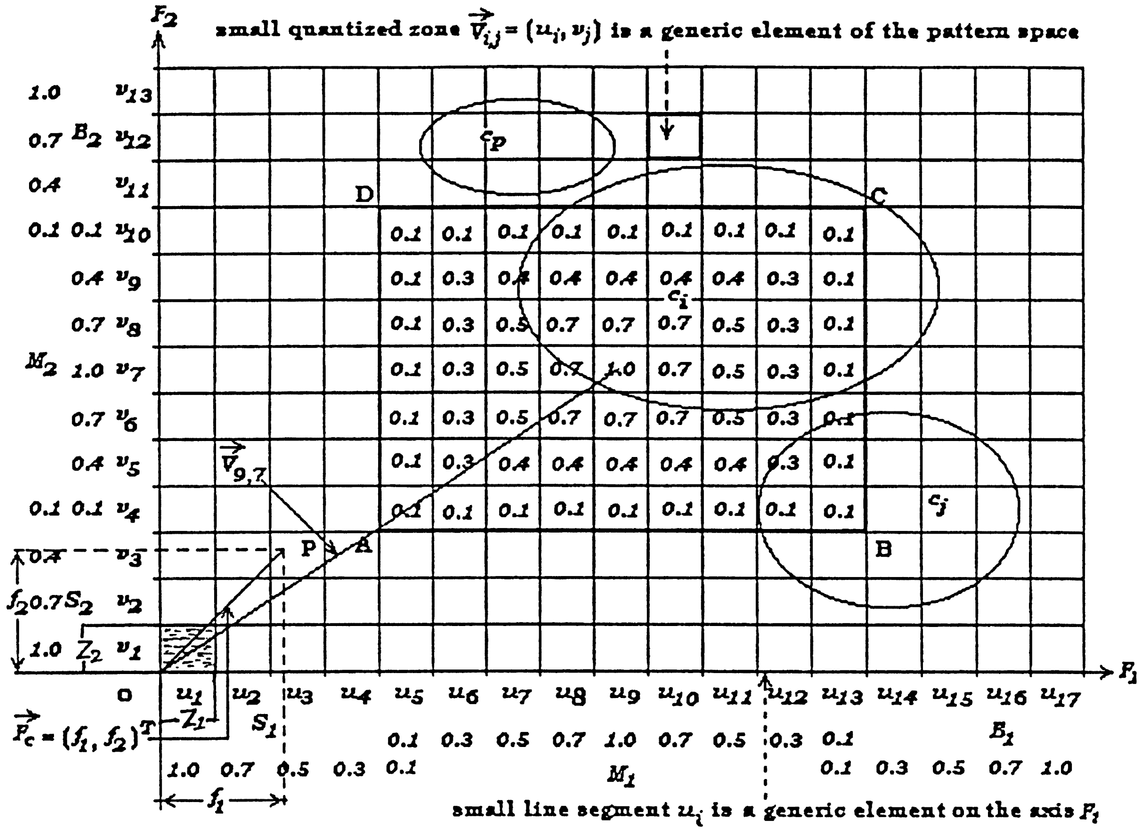


Fig. 7. Representation of a fuzzy Pattern vector. Key: *The area ABCD represents the fuzzy pattern vector* $\vec{F}_f = (M_1, M_2)^T$, where M_1 and M_2 are the fuzzy sets on F_1 and F_2 axes. The membership values of small quantized zones are determined by the relation $\mu_{\vec{F}_f}(\vec{V}_{ij}) = \min(\mu_{M_1}(u_i), \mu_{M_2}(v_j))$, where $\vec{V}_{ij} = (u_i, v_j)\forall_i = 1, 2, \ldots, 17; j = 1, 2, \ldots, 13$. Instead of "min" operator, we may use algebraic product, etc., depending upon the way we want to write the relation formed by the antecedent clauses of a one dimensional fuzzy implication.

example, according to the conventional interpretation, the above 2–D implication is translated into,

$$\left.\begin{array}{rlll} \text{a)} & \text{if } x \text{ is } A & \text{and } y \text{ is } B, & \text{then } z \text{ is } C \\ \text{or b)} & \text{if } x \text{ is } A & \text{and } y \text{ is } B, & \text{then } z \text{ is } C \end{array}\right\}. \qquad (56)$$

Thus, by interpretation a) of (56), at every observation (rule), features are given in the form of fuzzy sets which represent the antecedent clauses of the fuzzy **if-then** rule and which are defined over the universe of the feature axes. Therefore, in such a case, we cannot represent an individual pattern by a vector. Instead, from a given observation (rule), we can represent a population of patterns P in an area (in case of R^2 but a region in general) on the pattern space by a fuzzy pattern vector (see Fig. 7).

Now we consider the following remarks.

Remark 1: We quantize the universe of individual features axis by a small line segments [30], as shown in Fig. 7. Thus, we make the universe of each feature axis finite.

Remark 2: Over the quantized universe of the individual feature axis, we define the primary fuzzy terms $Z_i, S_i, M_i, B_i, \forall i$, where $Z_i \equiv \mathbf{zero}, S_i \equiv \mathbf{small}, M_i \equiv \mathbf{medium}$, and $B_i \equiv \mathbf{big}$ (see Fig. 7) [30].

Remark 3: Primary fuzzy terms Z_i and S_i are completely overlapped, as shown in Fig. 7. Also, Z_i and S_i may be partially overlapped without any lack of generality.

Remark 4: For the present treatment, we assume the primary fuzzy term $Z_i \forall_i$ as a fuzzy singleton. It may not be a fuzzy singleton in general.

Remark 5: As we have quantized our pattern space on R^2 by small square grids (see Fig. 7), a fuzzy point on the quantized pattern space is represented from an area ABCD which contains a fuzzy relation which is a fuzzy set in the quantized product space $U_1 \times U_2$.

Remark 6: The fuzzy point as stated in Remark 5 is linguistically described as (M_1, M_2) on the quantized product space.

Remark 7: In Fig. 7, the pair (F_1 is Z_1, F_2 is Z_2) is the initial point in the qunatized product space. This initial point is a fuzzy point which is a fuzzy singleton in the quantized product space. This is basically a single point fuzzy relation having membership value 1 in the quantized product space.

Remark 8: In the present text, we identify quantized pattern space and quantized product space $(U_1 \times U_2)$.

Let us now consider the following definition of a fuzzy vector between the initial point and a fuzzy point on the quantized product space.

Definition 7: Let $\vec{F}_f$ be a fuzzy vector having "c" components, each of which is a fuzzy set D^i defined over the universe U_i of the feature axis F_i. The fuzzy vector $\vec{F}_f$ is a fuzzy set in the quantized product space $U_1 \times U_2 \times \cdots \times U_c$. Each element of the fuzzy set is a vector having the same initial point but different terminal points which are the elements of the fuzzy point which is a fuzzy set (see Remark 5). Each terminal point of each vector in the set carries one membership value indicating its (vector's) degree of belongingness to the set $\vec{F}_f$. A fuzzy vector $\vec{F}_f$ is represented as

$$\vec{F}_f = \left\{ \left(\mu_{\vec{F}_f}((\vec{V}), \vec{V}) \, \forall \vec{V} \epsilon U_1 \times U_2 \times \cdots \times U_c \right\}$$

where $\mu_{\vec{F}_f} : U_1 \times U_2 \times \cdots \times U_c \rightarrow [0, 1]$ is the membership function of $\vec{F}_f$, and $\mu_{\vec{F}_f}(\vec{V})$ is the grade of membership of $\vec{V}$ in $\vec{F}_f$.

Remark 9: Without lack of generality, we may extend Definition 7 of a fuzzy vector between two fuzzy points, but such generality is not needed for the present discussion.

The process of defuzzification of the fuzzy vector $\vec{F}_f$ is performed on selecting the elements of the fuzzy vector $\vec{F}_f$, which is a fuzzy set, having highest membership values. The defuzzified version of the fuzzy vector $\vec{F}_f$ is a set. In the case that the said defuzzified version is a singlton, then the defuzzified version of $\vec{F}_f$ becomes the crisp vector as stated in this section (see Example 1). The fuzzy set D^i as mentioned in the Definition 7 is an element of the term set as discussed earlier in this section. The universe of the components of $\vec{F}_f$, i.e., the fuzzy set D^i, may be continuous/discrete and the universe of $\vec{F}_f$ may be continuous/discrete. If the universes are discrete, we should follow the numerical definition of membership functions; otherwise, we should follow the functional definition. If the defuzzified version of $\vec{F}_f$ reduces to the crisp vector as stated earlier, the membership value at the terminal point of the vector $\vec{V}$ of $\vec{F}_f$ can alternatively be interpreted as the highest possibility of $\vec{V}$ to hold the property of the fuzzy vector $\vec{F}_f$. By the term property, we mean a particular combination of the elements of different term sets. For instance, with respect to Fig. 7, the property assumed by fuzzy vector $\vec{F}_f$ is $(M_1, M_2)^T$. Like this, we can have property $(M_1, S_2)^T, (M_1, B_2)^T$, etc.

Thus, we introduce the notion of a fuzzy vector (i.e., $\vec{F}_f$) which is an analogous representation of $\vec{F}_c$ on R^2 (see Fig. 7). When we write fuzzy **if-then** rules to represent the patterns on R^2, the fuzzy vector, as stated above, becomes a fuzzy pattern vector. The fuzzy pattern vector no longer represents a single pattern on R^2; rather it represents a population of patterns.

Example 1: Let us consider the following fuzzy pattern vector (see Fig. 7),

$$\begin{aligned} \begin{bmatrix} F_1 & \text{is} & M_1 \\ F_2 & \text{is} & M_2 \end{bmatrix} &= \vec{F}_f = \{\mu_{\vec{F}_f}(\vec{V})/\vec{F}\} \\ &= \sum_{i=1}^{17} \sum_{j=1}^{13} \mu_{\vec{F}_f}(\vec{V}_{i,j})/\vec{V}_{i,j} = 0.1/\vec{V}_{5,4} \\ &\quad + \ldots 1.0/\vec{V}_{9,7} + \cdots + 0.1/\vec{V}_{13,10} + \ldots \end{aligned}$$

where "+" and "$\sum$" are in the set theoretic sense and M_1 is a fuzzy set $\{0.1/u_5, 0.3/u_6, 0.5/u_7, 0.7/u_8, 1.0/u_9, 0.7/u_{10}, 0.5/u_{11}, 0.3/U_{12}, 0.1/u_{13}\}$ on the universe U_1, M_2 is a fuzzy set $\{0.1/v_4, 0.4/v_5, 0.7/v_6, 1.0/v_7, 0.7/v_8, 0.4/v_9, 0.1/v_{10}\}$ on the universe U_2 and each vector $\vec{V}_{ij}$ is a generic element of the fuzzy set $\vec{F}_f$, that represents one small quantized zone on the pattern space. Thus instead of a point pattern, a population of patterns is represented by $\vec{F}_f$. Here the fuzzy set $\vec{F}_f$ is defined over the universe $U_1 \times U_2$, i.e., the quantized product space. In this case, note that the defuzzified version of the fuzzy vector $\vec{F}_f$ is a singleton represented by the vector $\vec{V}_{9,7}$.

Depending upon the area of each class occupied from $\vec{F}_f$, (see Fig. 7), we determine the degree of occurrence of different classes of patterns under that $\vec{F}_f$. Such degree of occurrence is represented by a fuzzy set which is the consequent part of an MFI and which is defined in the quantized pattern space which is the universe of all pattern classes. Thus, in the same pattern space, i.e., in the quantized product $U_1 \times U_2$ (if $c = 2$), we define two types of fuzzy sets; one fuzzy set is represented by $\vec{F}_f$, and the other fuzzy set is the consequent part of a MFI. The consequent part of a MFI, which is a fuzzy set, simply indicates the relative position of a fuzzy pattern vector with respect to different classes of patterns in the quantized pattern space. Once the antecedent part, and the consequent part of a MFI are represented by two types of fuzzy sets as stated above, our next job is to attach a meaningful interpretation to the said representations.

Example 2: Let us consider the fuzzy pattern vector $\vec{F}_f$ of Fig. 7. The position of $\vec{F}_f$ on the pattern space means the area ABCD. The position ABCD of $\vec{F}_f$ is obtained when

$$\vec{F}_f = \begin{bmatrix} F_1 & \text{is} & M_1 \\ F_2 & \text{is} & M_2 \end{bmatrix}$$

and the position of $\vec{F}_f$ is changed when

$$\vec{F}_f = \begin{bmatrix} F_1 & \text{is} & M_1 \\ F_2 & \text{is} & S_2 \end{bmatrix} \text{ etc}$$

Now, if we try to compute the fuzzy set C which is the consequent part of the following MFI:

$$\text{If } \begin{pmatrix} F_1 & \text{is} & M_1 \\ F_2 & \text{is} & M_2 \end{pmatrix} \rightarrow C$$

we have to consider the relative position of $\vec{F}_f$, i.e., the area ABCD with respect to the defined cover c_i, c_j, and c_p. For simplicity of demonstration, we consider partial cover.

From Fig. 7, it is obvious that the area of class c_i is substantially occupied by ABCD. Looking at the possibility values of the small quantized zones of c_i occupied by ABCD, we can have the following four types of estimate of class-membership for the class c_i.

For class c_i

1) *optimistic estimate*: the highest membership value of the small quantized zones of the area of class c_i occupied by ABCD, for instance, 1.0 indicate by $\vec{V}_{9,7}$ of Fig. 7 (also see Example 1).
2) *pessimistic estimate*: the lowest membership value of the small quantized zones of the area of class c_i occupied by ABCD, for instance, 0.1 (see Fig. 7).
3) *expected estimate*: average of the membership values of all the small quantized zones of the area of class c_i occupied by ABCD, for instance, 0.381 (see Fig. 7).
4) *most likely estimate*: comes from the subjective quantification of human perception as mentioned in [28]. Here, in this example, the subjective quantification of belongingness of a population of patterns to a particular class (say c_i) is achieved looking at the area of the said class occupied by the fuzzy vector, for instance, 0.7 (see Fig. 7), which is the subjective quantification of human percep-

tion as stated above and which may vary since the perception from one person to another varies within certain limit; but cannot be changed abruptly. For further verification of the subjective quantification of our perception, we may consider the following simple calculation.

Let the area of the class c_i of Fig. 7 be approximated by the total number of small quantized zones covered up (partly or fully) by the contour of the class c_i. From Fig. 7, we see 46 that such zones represent the area of the class c_i. The area of the class c_i occupied by ABCD of the fuzzy vector is 32 zones. Now, if we take the ratio $(32 \div 46) = 0.696$, which is the computed value of the belongingness of a population of patterns to class c_i with respect to the fuzzy pattern vector $\vec{F}_f$ of Fig. 7, then we see that the subjective quantification of belongingness of a population of patterns, i.e., 0.7 or 0.6 or 0.8, lies close to that of the computed value.

Similarly, we can have the estimates of the class-membership for the classes c_j and c_p.

For class c_j

1) optimistic estimate $= 0.3$;
2) pessimistic estimate $= 0.1$;
3) expected estimate $= (0.3 + 0.3 + 0.1 + 0.1 + 0.1 + 0.1)/6 = 1.0/6 = 0.167.$;
4) most likely estimate $= 0.3$. Note that here the computed value of belongingness is $(6/23) = 0.26$.

For class c_p

All estimates are zero.

Thus, we get four fuzzy sets for the consequent part of a MFI (i.e., the fuzzy set C)

$$C_{\text{opt}} = \{1.0/c_i, 0.3/c_j, 0.0/c_p\}$$
$$C_{\text{pess}} = \{0.1/c_i, 0.1/c_j, 0.0/c_p\}$$
$$C_{\text{expt}} = \{0.381/c_i, 0.167/c_j, 0.0/c_p\}$$
$$C_{\text{most}} = \{0.7/c_i, 0.3/c_j, 0.0/c_p\}.$$

In this paper, for all subsequent discussions for the design study of the classifier based on FRC, we assume the optimistic estimate of the fuzzy set C (without mentioning anything like opt, pess, expt, most) which represents the consequent part of (a) of (56). Instead of considering the optimistic estimate of the consequent part of a) of (56) in our study, we may use other kind of estimates, as stated above.

REFERENCES

[1] H.-F. Wang, "Numerical analysis on fuzzy relation equations with various operators," *Fuzzy Sets Syst.*, vol. 53, pp. 155–166, 1993.

[2] W. Pedrycz, "Numerical and applications aspects of fuzzy relational equations," *Fuzzy Sets Syst.*, vol. 11, pp. 1–18, 1983.

[3] ——, "Approximate solutions of Fuzzy relational equations," *Fuzzy Sets Syst.*, vol. 28, pp. 183–201, 1988.

[4] ——, "Applications of fuzzy relational equations for methods of reasoning in presence of fuzzy data," *Fuzzy Sets Syst.*, vol. 16, pp. 163–174, 1985.

[5] ——, "Processing of relational structures: Fuzzy relational equations," *Fuzzy Sets Syst.*, vol. 40, pp. 77–106, 1991.

[6] ——, "On generalized fuzzy relational equations and their applications," *Math. Anal. Appl.*, vol. 107, pp. 520–536, 1985.

[7] N. Ikoma, W. Pedrycz, and K. Hirota, "Estimation of fuzzy relational matrix by using probabilistic descent method," *Fuzzy Sets Syst.*, vol. 57, pp. 335–349, 1993.

[8] H. Hellendoorn, "The generalized modus ponens considered as a fuzzy relationa," *Fuzzy Sets Syst.*, vol. 48, pp. 29–48, 1992.

[9] R. Fuller and H. J. Zimmerman, "Fuzzy reasoning for fuzzy mathematical programming problems," *Fuzzy Sets Syst.*, vol. 60, pp. 121–133, 1993.

[10] A. Dinola, W. Pedrycz, S. Sessa, and E. Sanchez, "Fuzzy relation equations theory as a basis of fuzzy modeling: An overview," *Fuzzy Sets Syst.*, vol. 40, pp. 415–429, 1991.

[11] M. K. Chakraborty, "Some aspects of [0, 1] fuzzy relations and a few suggestions toward its use," *Approx. Reas. Expert Syst.*, pp. 139–157, 1985.

[12] S. V. Ovchinnikov and T. Riera, "On fuzzy classifications," *Fuzzy Sets Syst.*, vol. 49, pp. 119–132, 1992.

[13] S. Gottwald, "Approximately solving fuzzy relation equations: Some mathematical results and some heuristic proposals," *Fuzzy Sets Syst.*, vol. 66, pp. 175–193, 1994.

[14] L. A. Zadeh, "Similarity relations and fuzzy orderings," *Inform. Sci.*, vol. 3, pp. 177–200, 1971.

[15] K. Hirota and W. Pedrycz, "Analysis and synthesis of fuzzy systems by the use of probabilistic sets," *Fuzzy Sets Syst.*, vol. 10, pp. 1–13, 1983.

[16] Y. Yoshinari, W. Pedrycz, and K. Hirota, "Construction of fuzzy models through clustering techniques," *Fuzzy Sets Syst.*, vol. 54, pp. 157–165, 1993.

[17] W. U. Wangming, "Fuzzy reasoning and fuzzy relational equations," *Fuzzy Sets Syst.*, vol. 20, pp. 67–78, 1986.

[18] J. Jacas and J. Recasens, "Eigenvectors and generators of a fuzzy relation," *IEEE Trans. Syst., Man, Cybern.*, vol. 22, pp. 687–694, 1992.

[19] T. Arnould and S. Tano, "A rule based method to calculate exactly widest solution sets of a max-min fuzzy relation uniquely," *Fuzzy Sets Syst.*, vol. 64, pp. 39–58, 1994.

[20] P. B. Nikolai and D. G. Naryshkin, "On identification of multidimensional fuzzy systems," *Fuzzy Sets Syst.*, vol. 35, pp. 325–331, 1990.

[21] W. Pedrycz, "Genetic algorithms for learning in fuzzy relation structures," *Fuzzy Sets Syst.*, vol. 69, pp. 37–45, 1995.

[22] A. Balnco, M. Delgado, and I. Requena, "Identification of fuzzy relational equations by fuzzy neural network," *Fuzzy Sets Syst.*, vol. 71, pp. 215–226, 1995.

[23] L. Gorzabz and A. Marin, "Extending fuzzy relations," *Fuzzy Sets Syst.*, vol. 69, pp. 157–169, 1995.

[24] K. S. Say and J. Ghoshal, "Approximate reasoning approach to pattern recognition," *Fuzzy Sets Syst.*, vol. 77, pp. 125–150, 1996.

[25] M. Mizumoto, "Extended Fuzzy reasoning," in *Approximate Reasoning in Expert Systems*, M. Gupta, Kandel, Blander, and Kiszka, Eds, Amsterdam, The Netherlands: North Holland, 1985, pp. 71–85.

[26] M. Sugeno and T. Takagi, "Multidimensional Fuzzy reasoning," *Fuzzy Sets Syst.*, vol. 9, pp. 313–325, 1983.

[27] J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*. Reading, MA: Addison-Wesley, 1974.

[28] L. A. Zadeh, "Theory of approximate reasoning," in *Machine Intelligence*, J. E. Hayes, D. Michie, and L. I. Mikulich, Eds. London, U.K.: Ellis Horwood, 1970, pp. 149–194.

[29] Y. Tsukamoto, "An approach to fuzzy reasoning method," in *Advance in Fuzzy Set Theory and Applications*, M. M. Gupta, R. K. Ragade, and R. R. Yager, Eds, Amsterdam, The Netherlands: North Holland, 1979, pp. 137–149.

[30] C. C. Lee, "Fuzzy logic control systems: Fuzzy logic controller, Part I & II," *IEEE Trans. Syst., Man, Cybern.*, vol. 20, 1990.

[31] G. Schulz, *Fuzzy Rule Based Expert Systems and Genetic Machine Learning*. Berlin, Germany: Physica Verlag, 1995.

[32] W. Pedrycz, "Fuzzy sets in pattern recognition methodology and methods," *Pattern Recognit.*, vol. 23, no. 1/2, pp. 121–146, 1990.

[33] R. E. Bellman, R. Kalaba, and L. A. Zadeh, "Abstraction and pattern classification," *J. Math. Anal. Applicat.*, vol. 13, pp. 1–7, 1966.

[34] L. A. Zadeh, "Fuzzy sets and their applications to classification and clustering," in *Classification and Clustering*, J. Van Ryzin, Ed. New York: Academic, 1977, pp. 251–299.

[35] R. Di. Mori, *Computerized Models of Speech Using Fuzzy Algorithms*. New York: Plenum, 1983.

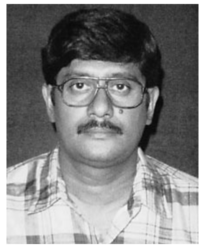
[36] R. Di. Mori and P. Laface, "Use of fuzzy algorithms for phonetic and phonetic labeling of continuous speech," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI–2, pp. 136–148, 1980.

[37] K. Hirota, "Fuzzy robot vision and fuzzy controlled robot," in *NATO ASI CIM*, I. B. Turksen, Ed, Berlin, Germany: springer, 1988.

[38] A. Seif and J. Aguilar-Martin, "Multi-group classification using fuzzy correlation," *Fuzzy Sets Syst.*, vol. 3, pp. 109–122, 1980.

[39] D. Dubois and M. C. Jaulent, "Techniques for extracting fuzzy regions," in First IFSA Congr., vol. Ii, Mallorca, Spain, July 1–6, 1985.

[40] K. Hirota, K. Iwami, and W. Pedrycz, "FCM-AD (fuzzy cluster means with additional ata) and its application to aerial images," in *Proc. II IFSA Congr.*, vol. II, Tokyo, Japan, 1987, pp. 729–732.
[41] T. L. Huntsberger, C. L. Jacobs, and R. L. Canon, "Iterative fuzzy image segmentation," *Pattern Recognit.*, vol. 18, pp. 131–138, 1985.
[42] W. J. Kickert and H. Koppleaar, "Application of fuzzy set theory to syntatic pattern recognition of handwritten capitals," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC–6, pp. 148–151, 1976.
[43] P. Siy and C. S. Chen, "Fuzzy logic for handwritten numerical character recognition," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC–4, pp. 570–575, 1974.
[44] M. Shimura, "Applications of fuzzy set theory to pattern recognition," *J. JAACE*, vol. 19, pp. 243–248, 1975.
[45] T. L. Huntsberger, Ch. Rangarajan, and S. N. Jayaramurthy, "Representation of uncertainty in computer vision using fuzzy sets," *IEEE Trans. Comput.*, vol. C–2, pp. 145–156, 1986.
[46] E. T. Lee, "Proximity measure for the the classification of geometric figures," *J. Cybern.*, vol. 2, pp. 43–59, 1972.
[47] L. A. Zadeh, K. S. Fu, K. Tanaka, and M. Shimura, Eds., *Fuzzy Sets and Their Applications to Cognitive and Decision Processes*. New York: Academic , 1975.
[48] G. Bortolan and R. Degani, "Ranking of fuzzy alternatives in electrocardiography," in *Fuzzy Information, Knowledge Representation and Decision, Analysis*, E. Sanchez and M. M. Gupta, Eds. Oxford, U.K.: Pergamon, 1983, pp. 397–402.
[49] G. Bortolan, R. Degani, K. Hirota, and W. Pedrycz, "Classification of ECG Signals—utilization of fuzzy pattern matching," in *Proc. Int. Workshop Fuzzy Syst. Applic.* , Iizuka, Japan, 1988.
[50] R. Degani and G. Bortolan, "Computerized electrocardiogram diagnosis: Fuzzy approach," in *Encyclopedia of Systems and Control*. Oxford, U.K.: Pergamon, 1987.
[51] ——, "Fuzzy numbers in computerized electrocardiography," *Fuzzy Sets Syst.*, vol. 24, pp. 345–362, 1987.
[52] W. Pedrycz and A. Gacek, "Feature selection for ECG signal classification with the aid of qualitative and quantitative criteria," in *7th Hungarian Conf. Biomed. Engg.*, Esztergom, Hungary, Sept. 16–18, 1987.
[53] W. Pedrycz, "ECG Signal classification with the aid of linguistic classifier," in *Proc. XIV Int. Conf. Med. Biomed. Eng.*, Spain, Aug. 11–16, 1985.
[54] S. Watanabe, *Pattern Recognition: Human and Mechanical*. New York: Wiley, 1985.
[55] A. Kumar, "A real-time system for pattern recognition of human sleep stages by fuzzy systems analysis," *Pattern Recognit.*, vol. 9, pp. 43–46, 1977.
[56] L. Saitta and P. Tarasso, "Fuzzy characteristics of coronary disease," *Fuzzy Sets Syst.*, vol. 5, pp. 245–258, 1981.
[57] M. A. Woodbury and J. Clive, "Clinical pure types as fuzzy partition," *J. Cybern.*, vol. 3, pp. 111–121, 1974.
[58] J. C. Bezdek and S. K. Pal, Eds., *Fuzzy Models for Pattern Recognition: Methods that Search for Structures in Data*. New York: IEEE , 1992.
[59] P. K. Simpson, "Fuzzy min-max neural network-Part 1: Classification," *IEEE Trans. Neural Networks*, vol. 13, pp. 776–786, 1992.
[60] B. Kosko, *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*. Englewood Cliffs, NJ: Prentice-Hall, 1992.
[61] B. Kosko, *Fuzzy Thinking: The New Science of Fuzzy Logic*. London, U.K.: Hyperion, 1993.



Kumar S. Ray received the B.E. degree in mechanical engineering in 1977 from Calcutta University, Calcutta, India, the M.Sc. degree in control engineering in 1980 from the University of Bradford, Bradford, U.K., and the Ph.D. degree in computer science in 1987 from Calcutta University.

He is currently a Professor of electronics and communication sciences, Indian Statistical Institute, Calcutta. He was a Visiting Faculty Member of the University of Texas at Austin under a United Nations Development Programme (UNDP) fellowship in 1990. He has 50 journal publications to his credit. He is the co-author of two edited volumes of North-Holland. His field of interest are control theory, computer vision, AI, fuzzy reasoning, neural networks, genetic algorithms, and qualitative physics.

Dr. Ray was a member of task force committee of the Government of India, Department of Electronics (DoE/MIT), for the application of AI in power plants. He is the founder member of Indian Society for Fuzzy Mathematics and Information Processing (ISFUMIP) and a member of Indian Unit for Pattern Recognition and Artificial Intelligence (IUPRAI). In 1991, he was the recipient of the K. S. Krishnan memorial award for the best system oriented paper in computer vision.

Tapan K. Dinda received the B.Sc. (Hons.) degree in mathematics in 1991 and M.Sc. degree in applied mathematics with computer programming and operation research in 1993 from Vidyasagar University, India. He is currently pursuing the Ph.D. degree from the University of Calcutta, Calcutta, India.

Currently, he is a Senior Engineer of Information Technology Industry at Alumnus Software Ltd., Calcutta. He was a Senior Research Fellow with the University of Calcutta, doing collaborative research work with Indian Statistical Institute, Calcutta, under the fellowship of University Grant Commission from 1994 to 1999. His research interest includes pattern recognition, fuzzy sets and systems, genetic algorithms, neural networks.

Mr. Dinda is a member of Indian Unit for Pattern Recognition and Artificial Intelligence (IUPRAI) and the Indian Society for Fuzzy Mathematics and Information Processing (ISFUMIP).