Fuzzy Set Theoretic Measures for Automatic Feature Evaluation: II

SANKAR K. PAL*

Electronics and Communication Sciences Unit, Indian Statistical Institute, Calcutta 700035, India

Communicated by Abe Kandel

ABSTRACT

The present paper is a continuation of our previous work [1] in which we developed an algorithm for automatic ranking of the individual feature importance for pattern recognition problems. Here we have extended the work by (i) evaluating the importance of any subset of features collectively (ii) providing an average feature evaluation index considering all the classes and (iii) comparing the results with those of statistical measures considering their variation with interest distance.

Effectiveness of the algorithm is demonstrated on six-class, three-feature vowel data; four-class, five-feature consonant data; and three-class fifteen-feature mango leaf data.

1. INTRODUCTION

The present study is a continuation of our previous report [1] on fuzzy set theoretic measures for automatic feature evaluation. The terms *index of fuzziness*, *entropy*, and *π-ness*, which give measures of fuzziness in a set, were used in [1] to define an index of feature evaluation in pattern recognition problems in terms of their intraclass and interclass measures. These two types of measurements were found to reflect the concepts of intraset and interset distances in classical set theory such that the index value decreases as the reliability of a feature in characterizing and discriminating different classes increases. The algorithms were implemented with success to solve a speech recognition problem where the features (formant and antiformant frequencies) were ranked according to their *individual importance* (goodness) in characterizing and discriminating classes. Both standard and approximated versions of the S and π functions were considered in developing the algorithms.

*At present, Dr. Pal is with the NASA/Johnson Space Center, PT4/Software Technology Branch, Information Technology Division, Houston, Texas 77058.

655 Avenue of the Americas, New York, NY 10010 0020-0255/92/$5.00

The present paper describes some further developments and investigations of those algorithms along with an additional application on large dimensional data (large feature set). These developments include the following:

a) Determination of the feature evaluation index (FEI) for a set of N, $N \geqslant 2$ features taken at a time out of Q, $Q > N$ measurements on patterns. In other words, automatic determination (without performing a recognition experiment) of the set of N features that collectively will characterize and discriminate best the different classes. Obviously, this collective importance has more significance in practice than the individual importance of features for recognizing a pattern.

Furthermore, since $N = 1$ corresponds to our previous work [1], the investigation just mentioned can be treated as an extension leading to a *generalized algorithm* for FEI.

b) Determining average FEI of a set of features with respect to all the classes under consideration. Weighting coefficients proportional to the number of samples in a class have been introduced here in computing the average importance.

c) Studying the variation of FEI with interset distance among classes.

d) Demonstrating the effectiveness of the generalized algorithm to three-class, fifteen-feature mango leaf data; ten-class, three-feature vowel data; and four-class, five-feature plosive consonant data.

e) Comparison of the algorithm with other statistical separability measures, namely, divergence and JM distance [2].

In developing the generalized FEI algorithm, we considered standard π membership function in $\mathbb{R}^N$ [3] in measuring fuzziness in a set. Again, among the various aforementioned fuzzy measures, the entropy is only considered here as a typical example for illustrating the results.

2. ENTROPY AND FEATURE EVALUATION INDEX

Entropy of a fuzzy set $A = \{\mu_A(x_i)/x_i,\ i = 1,2,\ldots,n\}$ is defined according to Deluca and Termini [4] as

$$H(A) = \frac{1}{n \ln 2} \sum_i S_n(\mu_A(x_i)),\ i = 1,2,\ldots,n \tag{1}$$

with

$$S_n(\mu_A(x_i)) = -\mu_A(x_i) \ln \mu_A(x_i) - (1 - \mu_A(x_i)) \ln(1 - \mu_A(x_i)). \tag{2}$$

$H(A)$ is such that

$$H(A) = 0(\min) \quad \text{for } \mu_A(x_i) = 0 \quad \text{or } 1 \quad \text{for all } i \tag{3a}$$

$$H(A) = 1(\max) \quad \text{for } \mu_A(x_i) = 0.5 \quad \text{for all } i. \tag{3b}$$

Therefore, H increases monotonically in the interval [0, 0.5] and decreases monotonically in [0.5, 1] with a maximum of unity at $\mu = 0.5$. The similar properties hold good for index of fuzziness [5], π-ness [1] and index of crispness [6].

Let $C_1, C_2, \ldots, C_j, \ldots, C_m$ be the m pattern classes in a Q dimensional $(X_1, X_2, \ldots, X_q, \ldots, X_Q)$ feature space Ω_x. Also, let n_j $(j = 1, 2, \ldots, m)$ be the number of samples available from class C_j. The feature evaluation index for the qth feature computed with the π function (Section 3) is accordingly defined as in [1].

$$(FEI)_{qjk} = \frac{H_{qj}^{\pi} + H_{qk}^{\pi}}{H_{qjk}^{\pi}} \tag{4}$$

$$j, k = 1, 2, \ldots, m, \qquad j \neq k, \qquad q = 1, 2, \ldots, Q.$$

Here, H_{qj} and H_{qk} denote the intraset entropy for the classes C_j and C_k, respdctively, on the basis of the qth measurement only. Similarly, H_{qjk} is the interset entropy between the classes C_j and C_k along the qth axis. H_{qj} and H_{qk} decrease and H_{qjk} increases (i.e., $(\text{FEI})_q$ decreases) as the reliability (goodness) of the qth feature in characterizing and discriminating the jth and kth classes increases. The method for computing Equation (4) was explained in detail in [1] and it is also mentioned for convenience in the next section.

For evaluating importance of a set s of N measurements ($N \geqslant 2$) taken together out of Q, $Q > N$ measurements, the generalized expression to be used is thus

$$(FEI)_{sjk} = \frac{H_{sj}^{\pi} + H_{sk}^{\pi}}{H_{sjk}^{\pi}}. \tag{5}$$

The π function in N-dimensional feature space is defined in the next section. $N = 1$ (set s consists of one measurement only) corresponds to Equation (4).

As seen from Equations (4) or (5), it evaluates the importance (reliability) of a feature (or a set of features) considering its (their) ability to characterize and discriminate only two classes, C_j and C_k. Considering all the m classes, the average importance of the set s can be computed from

$$(FEI)_s^{av} = \sum_j \sum_k (FEI)_{sjk} W_j W_k \tag{6}$$

where

$$W_j = \frac{n_j}{n_t},\ W_k = \frac{n_k}{n_t},\ n_t = \sum_j n_j, \qquad j,k = 1,2,\dots,m,\ k \neq j$$

W_j and W_k denote the weighting coefficients for the classes C_j and C_k and they ensure the provision of unequal weight for the different classes depending on their number of occurrences in Ω_X. This is logical in the sense that two pairs of classes having equal FEI values cannot contribute equally to the $(\mathrm{FEI})^{av}$ value; the pair having more samples (i.e., more probable classes) must have greater contribution or significance.

3. MEMBERSHIP FUNCTIONS

As stated in Section 1, fuzzy measures have been computed here with the π membership function only. The standard π functions to obtain $\mu_A(x_i)$ from x_i when $x \in \mathbb{R}$ and $x \in \mathbb{R}^N$ are defined here.

The standard π function of Zadeh [7] to obtain $\mu_A(x)$ from x when $x \in \mathbb{R}$ has the form

$$\mu_{A\pi}(x; a, c, a') = \mu_{AS}(x; a, b, c),\ x \leqslant c \tag{7a}$$

$$= 1 - \mu_{AS}(x; c, b', a'),\ x \geqslant c \tag{7b}$$

in the interval $[a, a']$ with $c = (a + a')/2$, $b = (a + c)/2$ and $b' = (a' + c)/2$. b and b' are the cross-over points, i.e., $\mu_{A\pi}(b) = \mu_{A\pi}(b') = 0.5$, and c in the central point at which $\mu_{A\pi} = 1$.

μ_{AS} is called the standard S function and is defined as

$$\begin{aligned}
\mu_{AS}(x; a, b, c) &= 0, & x \leqslant a \\
&= 2[(x-a)/(c-a)]^2, & a \leqslant x \leqslant b \\
&= 1 - 2[(x-c)/(c-a)]^2, & b \leqslant x \leqslant c \\
&= 1, & x \geqslant c
\end{aligned} \tag{8}$$

in the interval $[a, c]$ with $b = (a + c)/2$ such that $\mu_{AS}(b) = S(b; a, b, c) = 0.5$.

Pal and Pramanick [3] defined the π function when $x \in \mathbb{R}^N$ as

$$\mu_{A\hat{\pi}}(x;c,\lambda) = 2(1 - \|x-c\|/\lambda)^2, \qquad \lambda/2 \leq \|x-c\| \leq \lambda$$

$$= 1 - 2(\|x-c\|/\lambda)^2, \qquad 0 \leq \|x-c\| \leq \lambda/2 \qquad (9)$$

where c is the central point at which $\mu = 1$ and λ is the bandwidth. Equation (9) represents the membership function for the set of points $\{x\}$, $x \in \mathbb{R}^N$ clustered around c. The graphical representation of $\mu_{\hat{\pi}}$ when $x \in \mathbb{R}^2$ is shown in Figure 1.

In computing H_{qj}^{π} [Equation (4)], the parameters of Equation (7) were taken as in [1]:

$$c = (x_{qj})_{av} \qquad (10)$$

$$b' = c + \max\{|(x_{qj})_{av} - (x_{qj})_{max}|, \quad |(x_{qj})_{av} - (x_{qj})_{min}|\}$$

with

$$b = 2c - b'$$

$$a = 2b - c$$

$$a' = 2b' - c$$

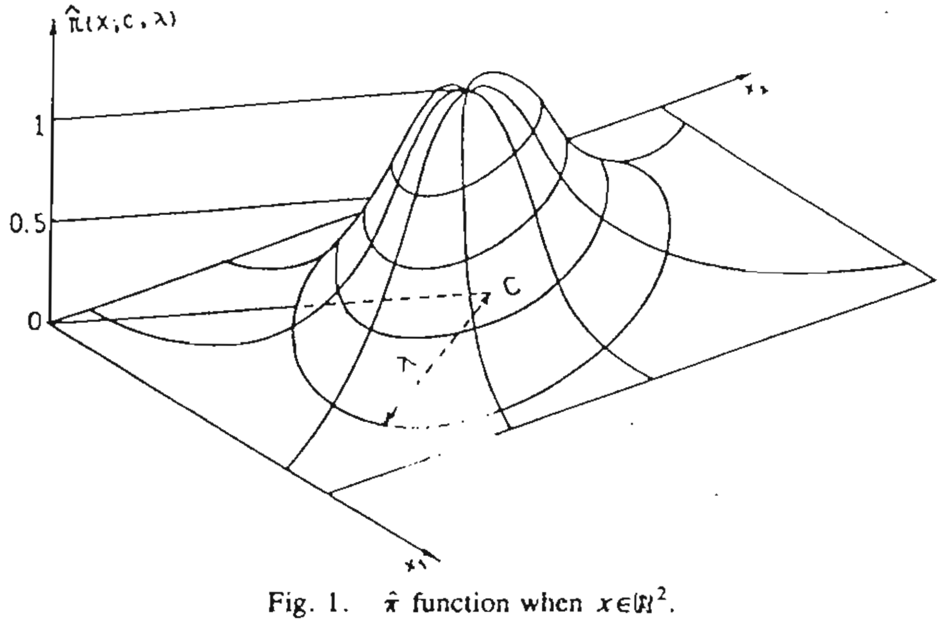

Fig. 1. $\hat{\pi}$ function when $x \in \mathbb{R}^2$.

where $(x_{qj})_{av}$, $(x_{qj})_{max}$ and $(x_{qj})_{min}$ denote the mean, maximum, and minimum values, respectively, computed along the qth coordinate axis, over all the n_j samples in C_j. This makes the $(x_{qj})_{av}$ point least ambiguous (i.e., $H = 0$) and the boundary points most ambiguous (i.e., $H = 1$) as far as their belongingness in C_j is concerned. In computing H^{*}_{qjk}, the classes C_j and C_k were pooled and the parameters a, b, and c of the π function were made to correspond to the pooled values over $(n_j + n_k)$ samples.

Similar is the case for Equation (5), where H^{*}_{sj} is computed with Equation (9) considering

$$c = (x_{sj})_{av}$$

and

$$\lambda = 2 \max_i \| x^i_{sj} - c \|, \qquad i = 1, 2, \cdots\cdots, n_j \tag{11}$$

where $\mathbf{x}$ is an N-dimensional vector and the set s contains N such features. $(\mathbf{x}_{sj})_{av}$ denotes the mean vector in $\mathbb{R}^N$ computed over n_j samples in C_j.

4. STATISTICAL MEASURES FOR FEATURE EVALUATION [2]

Divergence and JM distances are two widely used measures of separability of a pair of classes in statistical decision theory. For a pair of classes C_j and C_k characterized by normal distributions $N(\overline{X}_j, \Sigma_j)$ and $N(\overline{X}_k, \Sigma_k)$ ($\overline{X}$ and Σ being mean and covariance matrix), the divergence between the classes is defined as

$$d_{jk} = \tfrac{1}{2}\mathrm{Tr}\left\{(\Sigma_K - \Sigma_j)(\Sigma_j^{-1} - \Sigma_K^{-1})\right\}$$

$$+ \tfrac{1}{2}\mathrm{Tr}\left\{(\Sigma_K^{-1} + \Sigma_j^{-1})(\overline{X}_K - \overline{X}_j)(\overline{X}_K - \overline{X}_j)'\right\} \tag{12}$$

where $\mathrm{Tr}\{\cdot\}$ denotes trace.

The Jeffries-Matusita (JM) distance between C_j and C_K is defined as

$$J_{ij} = 2(1 - e^{-\alpha}) \tag{13a}$$

where

$$\alpha = \text{Bhattacharyya distance}$$

$$= \tfrac{1}{8}(\bar{X}_K - \bar{X}_j)'\left\{\frac{\Sigma_k + \Sigma_j}{2}\right\}^{-1}(\bar{X}_K - \bar{X}_j)$$

$$+ \tfrac{1}{2}\log\left\{\frac{|\Sigma_K + \Sigma_j|/2}{|\Sigma_K|^{1/2}|\Sigma_j|^{1/2}}\right\}. \tag{13b}$$

The higher the values of d_{jk} or J_{jk} for a feature subset, the greater is its importance as far as separability of classes is concerned.

5. IMPLEMENTATION AND RESULTS

The generalized feature evaluation algorithm has been implemented on speech and mango leaf data. Details of speech data containing ten-class, three-feature vowel and four-class, five-feature plosive consonants are given in our earlier report [1, 8]. The mango leaf data, on the other hand, have three classes with fifteen features. The problem of recognizing the kind (group) of mango from its plant-leaf has very important and significant application in forestry, agriculture, and horticulture because buying the plants without knowing the quality of fruit beforehand means risking a great amount of money.

First, three formant frequencies F_1, F_2, and F_3 are considered to be the measurements for vowel recognition; the higher formants are dependent on the speaker and are not considered for their recognition. For consonants, ΔF_1, ΔF_2 (the difference between the initial and final values of first and second formants), ΔT (duration), $\Delta F_1/\Delta T$, $\Delta F_2/\Delta T$ (the rate of transition) constitute the feature set. The speech units in consonant-vowel-consonant context were taken from Telugu (a major Indian language) vocabulary and were uttered by three male speakers in the age group 30–35 years. The data set consists of 496 vowels and 588 unaspirated plosives. The 15 measurements on the mango leaf are namely, area (A), perimeter (Pe), maximum length (L), maximum breadth (B), petiole (P), shape index (SI), length + petiole ($L + P$), length/petiole (L/P), length/maximum breadth (L/B), $(L + P)/B$, area/length (A/L), A/B, A/Pe, upper midrib/lower midrib, and upper Pe/lower Pe. The terms *upper* and *lower* are used with respect to maximum breadth position. Three types (classes) of mangoes were considered for extracting the aforesaid measurements. Leaves were collected over an entire season from three mango groves in West Bengal, the largest mango-producing state in India. The total number of samples used here is 165.

In [1], we presented the order of importance of individual vowel formants in characterizing the classes. For convenience, part of the result is included in Figure 2, in which the upper, middle, and lower entries correspond to index of fuzziness, entropy, and π-ness, respectively. Of the 10 vowel classes (∂, a, i, i:, u, u:, e, e:, o and o:), the shorter and longer categories are pooled together (since they differ only in duration), resulting in six classes namely, ∂, a, I, U, E, and O.

The average FEI values for F_1, F_2, and F_3 considering all the classes are given in Table 1. From Table 1, the order of average importance of features is found to be $F_2 > F_1 > F_3$ in characterizing and discriminating all the vowels. This conforms to the vowel diagram (Figure 3) in the F_1-F_2 plane, which indicates higher overall discrimination ability of Ω_X along the F_2 axis. F_3, being mostly speaker dependent, comes last.

Table 2 shows the FEI values when two features at a time were taken to evaluate their combined importance. The F_1-F_2 combination is seen from Table 2 to be the most important feature set in discriminating all (except the pairs /∂, *a*/ and /∂, O/) the class combinations. Now, it is a well established fact that the F_1-F_2 combination is the best characterizing feature for automatic vowel recognition. This is again verified here from their average FEI values

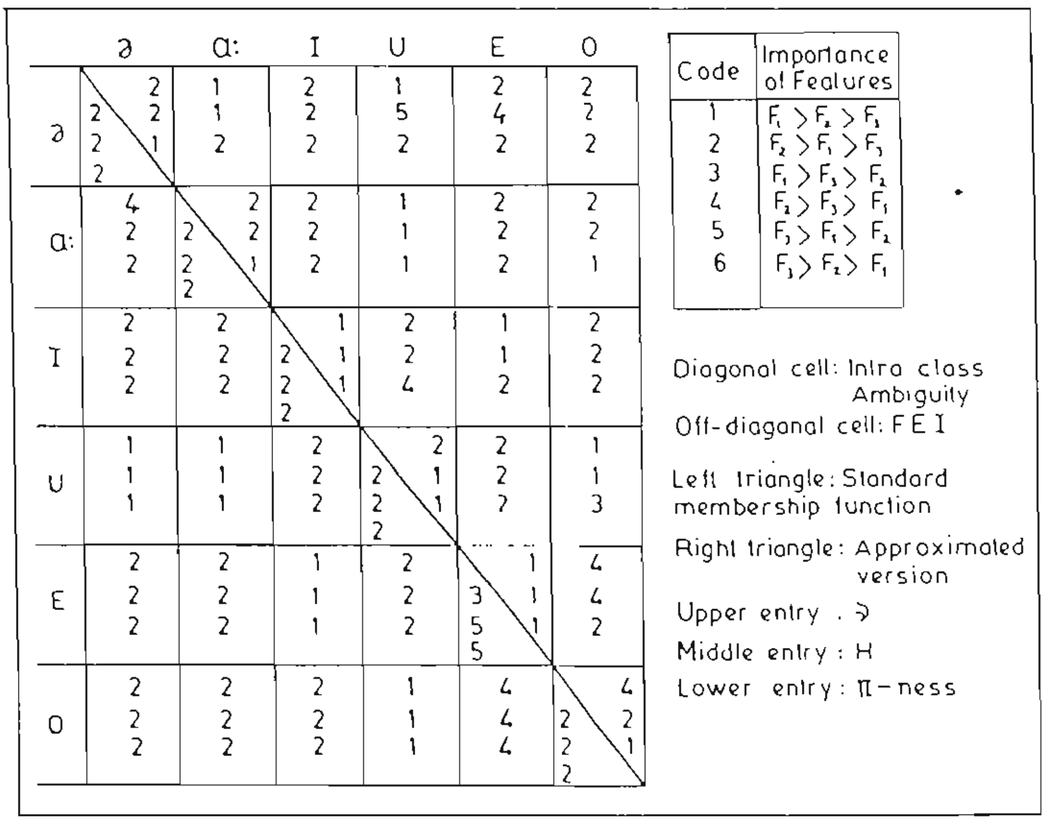

Fig. 2. Order of individual importance of features [1].

TABLE 1

Average FEI Values

Feature	$(FEI)^{av}$
F_1	0.1968
F_2	0.1482
F_3	0.2359

listed in Table 3. Between the F_1-F_3 and F_2-F_3 combinations, the latter one is found to be quantitatively more important. This again conforms to the findings in Table 1 where F_2 is seen, on the average, to be more important than F_1.

In our earlier report [1], the order of individual feature importance for plosive recognition did not seem to be regular as had been obtained in the case of the vowel recognition problem. The results on the average importance of the feature as indicated by the $(FEI)^{av}$ value and the combined importance of different feature subsets provide, on the other hand, a regular behavior and better reflection for discrimination (separability) ranking. For example, considering sets of two features, their collective importances are ranked as ΔF_1, $\Delta T > \Delta F_2$, $\Delta F_2/\Delta T > \Delta F_1$, $\Delta F_2 > \Delta F_1$, $\Delta F_2/\Delta T > \Delta T$, $\Delta F_2/\Delta T$ for

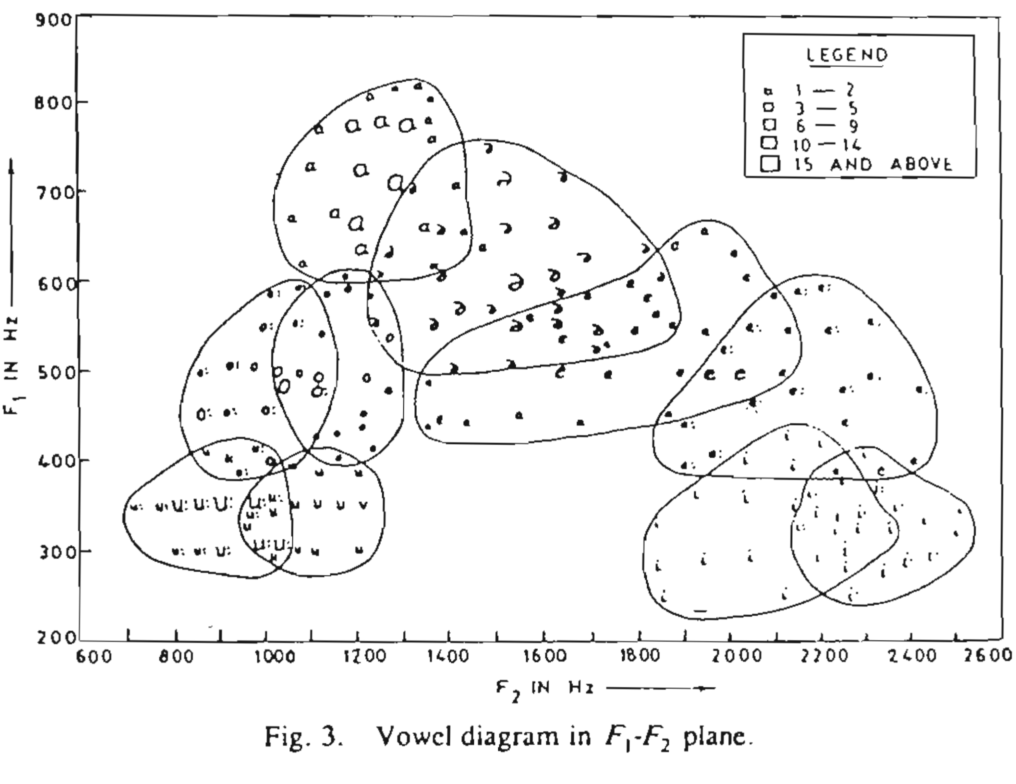

Fig. 3. Vowel diagram in F_1-F_2 plane.

TABLE 2

FEI Values for Different Feature Pairs

Classes	FEI for different feature pairs		
	F_1-F_2	F_1-F_3	F_2-F_3
δ, a	3.28	2.69	3.95
δ, I	1.46	2.87	1.81
δ, U	2.15	2.46	3.00
δ, E	2.24	3.18	2.34
δ, O	3.49	3.29	5.54
a, I	0.64	1.98	1.29
a, U	1.25	2.35	3.18
a, E	1.48	3.34	2.31
a, O	2.03	3.52	5.87
I, U	0.39	2.90	0.78
I, E	1.62	3.12	1.92
I, O	0.44	3.21	1.55
U, E	0.80	2.17	0.99
U, O	0.97	2.73	2.72
E, O	0.77	2.46	1.46

unvoiced plosives (/k/, /t/, /ṫ/, and /p/) and as

$$\Delta F_1,\ \Delta T > \Delta F_1,\ \Delta F_2 > \Delta F_2,\ \Delta T > \Delta F_2,\ \Delta F_2/\Delta T > \Delta T,\ \Delta F_2/\Delta T$$

for voiced plosives (/g/, /d/, /ḋ/, and /b/).

Referring to our earlier analysis based on individual ranking [1] and the work of Datta et al. [9] on plosive recognition, it was found overall that the features ΔF_1 and ΔT are most important in characterizing and discriminating different plosive sounds. This is exactly what is reflected quantitatively in the present study. Furthermore, considering sets of three features, the set $(\Delta F_1, \Delta F_2; \Delta T)$, came out to be the best as expected.

Table 4 illustrates the FEI values of 15 individual features of mango leaves in classifying different pairs of classes. Their ranking based on $(\mathrm{FEI})^{av}$ values

TABLE 3

Average FEI Values

Feature Set	$(\mathrm{FEI})^{av}$
F_1-F_2	0.4760
F_1-F_3	1.1169
F_2-F_3	0.8417

TABLE 4

FEI Values for Mango Leaf Data

Features		Diff. Class Pairs		
		C_1, C_2	C_1, C_3	C_2, C_3
X_1	Area (A)	2.45	2.25	2.02
X_2	Perimeter	2.16	2.18	2.06
X_3	Maximum length (L)	2.25	2.31	1.87
X_4	Maximum breadth (B)	2.05	1.84	1.99
X_5	Petiole (P)	2.93	3.36	2.01
X_6	Shape index	2.64	6.00	2.80
X_7	Length + Petiole	3.00	3.25	1.96
X_8	Length/Petiole	1.91	2.65	1.93
X_9	Length/Breadth (L/B)	7.95	8.64	1.75
X_{10}	$(L+P)/B$	3.06	2.99	2.14
X_{11}	Area/Length	2.13	2.20	2.29
X_{12}	A/B	2.21	1.70	2.31
X_{13}	Area/Perimeter	2.25	1.66	2.41
X_{14}	Upper Midrib/Lower Midrib	2.00	3.42	3.13
X_{15}	Peri Upper/Lower	2.22	2.70	2.24

(Table 5) is as follows:

$$X_4 > X_{12} > X_{13} > X_2 > X_{11} > X_8 > X_3 > X_1 > X_{15} > X_{14} > X_{10} > X_7 > X_5 > X_6 > X_9$$

For further investigation we have chosen only the first 10 important features, for the sake of simplicity. Tables 6 and 7 demonstrate the ranking in descending order of feature subsets containing two at a time and three at a time, respectively. The first five bests are included here only as an illustration. The most important feature subsets (Tables 6 and 7) are seen to contain the features X_4, X_8, X_{11}, X_{12}, and X_{13}. These measurements, according to their individual capability, are also found to lie within the first six ranks (Table 5).

The aforementioned quantitative selection of features conforms to that obtained by recognition experiment [10] on the basis of the Bayes classifier.

The last part of the study consists of comparing the results of the proposed algorithm with those of statistical measures, e.g., divergence and JM distance [Equations (12) and (13)]. The variation of $(\text{FEI})_{jk}$, d_{jk} and J_{jk} with interclass distance between C_j and C_k is investigated in this context. Tables 8 and 9 show their comparative results with distance when the formant pair (F_1-F_2) and formant triple (F_1-F_2-F_3) are considered as input feature sets,

TABLE 5

Average FEI Values

Features	$(\mathrm{FEI})^{\mathrm{av}}$
X_1	0.6476
X_2	0.6042
X_3	0.6221
X_4	0.5490
X_5	0.8328
X_6	1.1254
X_7	0.8273
X_8	0.6198
X_9	2.0672
X_{10}	0.8146
X_{11}	0.6110
X_{12}	0.5666
X_{13}	0.5708
X_{14}	0.7664
X_{15}	0.6780

respectively. This part of the study is carried out only for vowel data having the maximum number of classes. In these tables, values of 1/FEI are included so that the higher the value of 1/FEI for a feature set s, the greater its ability to discriminate classes. All these measures are found to be almost monotonically nondecreasing with interset distances between classes. Two types of distance measures are computed here, namely, Mahalanobis (squared) distance

$$D_{jk} = \left(\overline{X}_j - \overline{X}_k\right)' \left\{ \frac{\Sigma_j + \Sigma_k}{2} \right\}^{-1} \left(\overline{X}_j - \overline{X}_k\right) \tag{14}$$

involving dispersion within classes, and mean distance

$$D^{jk} = \frac{1}{n_j n_k} \sum_{i=1}^{n_j} \sum_{l=1}^{n_k} d\left(C_j^i, C_k^l\right) \tag{15}$$

TABLE 6

First Five Ranking of Feature Subset

Importance (Rank)	Feature Pair
1	X_8-X_{12}
2	X_{13}-X_{15}
3	X_4-X_{12}
4	X_{11}-X_{12}
5	X_{12}-X_{13}

TABLE 7

First Five Ranking of Feature Subset

Importance (Rank)	Feature Triple
1	X_4-X_8-X_{12}
2	X_8-X_{11}-X_{12}
3	X_8-X_{12}-X_{13}
4	X_8-X_{12}-X_{15}
5	X_4-X_{11}-X_{12}

where $d(C_j^i, C_k^l)$ is the Euclidean distance between the ith element of C_j and the lth element of C_k.

From Tables 8 and 9, it is seen that the variation of JM distance, unlike the other two, is smooth but it saturates very fast to its maximum limiting value (= 2.0) because the presence of the exponential term in Equation (13a) gives an exponential decreasing weight to increasing separation between classes. In other words, after a certain distance between classes is reached, the JM measure is unable to characterize the separability and hence the importance of a set of features quantitatively. For a large number of widely separated classes, the measures divergence and 1/FEI can provide quantitative indices of separability by the features. Again, for the divergence measure, the rate of rise is

TABLE 8

Variation of Different Measures with Interclass Distance (F_1-F_2 as feature set)

Classes	1/FEI	Divergence	JM Distance	Mean Distance	Mahalanobis Distance
δ, a	0.304	10.32	1.13	265.5	5.206
δ, I	0.681	28.84	1.95	831.9	28.654
δ, U	0.464	31.09	1.91	539.1	23.592
δ, E	0.445	5.95	0.94	525.1	4.699
δ, O	0.286	7.99	1.21	402.1	6.726
a, I	1.561	148.67	2.00	1036.0	80.876
a, U	0.798	55.48	2.00	459.0	51.053
a, E	0.674	59.59	1.84	696.3	17.280
a, O	0.492	16.65	1.73	300.0	15.551
I, U	2.518	117.75	2.00	1264.0	78.460
I, E	0.614	7.05	1.11	430.1	5.738
I, O	2.231	78.00	2.00	1168.1	61.098
U, E	1.242	57.68	1.94	953.3	24.723
U, O	1.027	10.77	1.47	219.2	10.369
E, O	1.296	28.22	1.73	838.7	14.142

TABLE 9
Variation of Different Measures with Interclass Distance
(F_1-F_2-F_3 as Feature Set)

Classes	1/FEI	Divergence	JM Distance	Mean Distance	Mahalanobis Distance
δ, a	0.267	11.57	1.27	373.53	6.151
δ, I	0.673	38.28	1.96	975.67	31.228
δ, U	0.344	33.73	1.92	630.29	23.712
δ, E	0.425	6.75	1.03	648.56	4.908
δ, O	0.184	10.34	1.38	552.22	8.054
a, I	0.784	176.46	2.00	1141.64	100.563
a, U	0.375	58.32	2.00	535.09	53.677
a, E	0.446	64.18	1.86	775.47	17.281
a, O	0.192	18.69	1.76	440.56	15.873
I, U	1.226	180.95	2.00	1332.27	84.506
I, E	0.524	11.14	1.23	536.81	5.794
I, O	0.653	128.34	2.00	1232.94	65.931
U, E	0.997	76.41	1.95	1004.57	25.773
U, O	0.392	11.52	1.48	370.14	10.430
E, O	0.700	35.15	1.79	900.36	15.141

very high at larger distances (for example, the rise in the divergence value from 117.75 to 148.67 with a small increase in distance from 78.46 to 80.876). This means that the separability measure increases significantly with little further increase in distance for widely separated classes. This is also intuitively unappealing in practice in the sense that the probability of correct classification should change slightly with a small change in class separability if we want to use d_{jk} as an indication of how successfully patterns in the classes C_j and C_k can be mutually correctly classified. Similar findings were also observed for Landsat imagery data [11] while evaluating the importance of different bands (features) for classification.

6. CONCLUSIONS AND DISCUSSION

A fuzzy set theoretic algorithm for evaluating importance of a set of features in characterizing and discriminating different classes in pattern recognition problems is described. Among the various measures of fuzziness of a set, the measure *entropy* is used, as an example, to minimize intraset ambiguity and to maximize interset ambiguity in feature space. These were implemented with a generalised $\hat{\pi}$ function that assigns the degree of belonging of a pattern to a class.

The automatic ranking of features obtained for speech recognition and mango variety classification is found to be very satisfactory. For plosive recognition, in particular, the index value provides a regular behavior in feature importance that was lacking in earlier work [1] on determining individual importance. Out of 15 measurements of mango leaves, a set of 4 to 5 features turned out to be consistently good by any kind of subset evaluation.

The proposed measure 1/FEI and the Statistical measures d_{jk} and J_{jk} are all found to increase with interset distance; thus providing a measure of separability and an indication of how successfully patterns in two classes can be correctly classified. J_{jk} saturates quickly, whereas the d_{jk} measures have a significant increase for a small increase in distance when the classes are widely separated.

The author wishes to thank Mr. Ashish Ghosh for doing the computer programming, Mr. T. Das and Mrs. S. De Bhowmick for typing the manuscript, Prof. S. K. Sen, Agriculture Department, Calcutta University, for providing data, and Professor D. Dutta Majumder for his interest in this work.

REFERENCES

1. S. K. Pal and B. Chakraborty, Fuzzy set theoretic measure for automatic feature evaluation, *IEEE Trans. Syst. Man Cybern.* SMC-16 (5):754-760 (1986).
2. J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*, Reading, Massachusetts, Addison-Wesley, 1974.
3. S. K. Pal and P. K. Pramanik, Fuzzy measures in determining seed points in clustering, *Patt. Recog. Lett.* 4:159-164 (July 1986).
4. A. Deluca and S. Termini, A definition of a nonprobabilistic entropy in the setting of fuzzy set theory, *Inform. Contr.* 20:301-312 (1972).
5. A. Kaufmann, *Introduction to the Theory of Fuzzy Subsets — Fundamental Theoretical Elements*, Vol. 1, Academic, New York, 1975.
6. S. K. Pal, A measure of edge ambiguity using fuzzy sets, *Patt. Recog. Lett.* 4:51-56 (February 1986).
7. L. A. Zadeh, Calculus of fuzzy restrictions, in *Fuzzy Sets and Their Application to Cognitive and Decision Process*, (Zadeh et al., Eds.), Academic, London, 1975, pp. 1-39.
8. S. K. Pal and D. Dutta Majumder, *Fuzzy Mathematical Approach to Pattern Recognition*, Wiley (Halsted Press), New York, 1986.
9. A. K. Datta, N. R. Ganguli, and S. Ray, Recognition of unaspirated plosives—A statistical approach," *IEEE Trans. Acoustic, Speech, Signal Process.*, ASSP-28:85-91 (Feb. 1980).
10. A. Bhattacharjee, Some aspects of mango (*Mangifora indica L*) leaf growth features in varietal recognition, M. Sc. Dissertation, Horticulture, Calcutta University, 1986.
11. J. A. Richards, *Remote Sensing Digital Image Analysis: An Introduction*, Springer-Verlag, Berlin 1986.

Received 1 August 1990; revised 15 October 1990