Pattern Recognition Letters 2 (1984) 275–279
North-Holland
September 1984

Intraclass and interclass ambiguities (fuzziness) in feature evaluation

S.K. PAL and B. CHAKRABORTY
Electronics and Communication Sciences Unit, Indian Statistical Institute, Calcutta 700035, India

Received 13 February 1984

Abstract: The terms Index of fuzziness, Entropy and π-ness which give measures of fuzziness (ambiguity) in a set are used here to define an Index of Feature Evaluation in pattern recognition problems in terms of intraclass and interclass ambiguity. The index is seen to possess a lower value for the feature having more importance in characterising a class. The algorithm has been implemented on a speech recognition problem.

Key words: Pattern recognition, feature evaluation, fuzzy sets.

1. Introduction

The process of selecting what information to present to the decision rule is called Feature Selection. The criterion of a good feature is that it should be invariant to within-class variation while emphasizing differences that are important in discriminating between patterns of different types. One of the useful techniques to achieve this is clustering transformation (Fukunaga (1972), Tou and Gonzalez (1974)) which maximises/minimises the interset/intraset distance using a diagonal transformation such that smaller weights are given to features having larger variance (less reliable). Other separability measures based on an information theoretic approach include Divergence, Bhattacharyya co-efficient, the Kolmogorov variational distance (Fukunaga (1972), Tou and Gonzalez (1974)).

The present work demonstrates an application of the theory of fuzzy sets to the prohlem of evaluating feature quality. The terms Index of fuzziness (Kaufmann (1975)), Entropy (DeLuca and Termini (1972)) and π-ness (Pal (1982)) provide measures of ambiguity (Fuzziness) in a set and are used here to define an Index of Feature Evaluation in terms of interclass and intraclass ambiguity. The lower the value of this index for a feature, the greater is the importance (quality) of the feature in separating classes in the feature space.

Effectiveness of the algorithms is demonstrated on a vowel sound recognition problem using the first three formants along with S and π functions.

2. Fuzzy sets and measures of fuzziness

A fuzzy set A with its finite number of supports $x_1, x_2, \ldots, x_n$ in the universe of discourse U is formally defined as

$$A = \{(\mu_A(x_i), x_i)\} = \{\mu_i / x_i\}, \quad i = 1, 2, \ldots, n, \tag{1}$$

where the characteristic function $\mu_A(x_i)$ known as membership function having positive value in the interval [0,1] denotes the degree to which an event x_i may be a member of A. The point x_i for which $\mu_A(x_i) = 0.5$ is said to be the cross-over point of the fnzzy set A. Let $\underline{A}$ be defined as the nearest ordinary set to A, such that

$$\mu_A(x_i) = \begin{cases} 0 & \text{if } \mu_A(x_i) \le 0.5, \\ 1 & \text{if } \mu_A(x_i) > 0.5. \end{cases}$$

Index of Fuzziness. The linear index of fuzziness $\gamma_\ell(A)$ of A is defined as (Kaufmann (1975))

$$\gamma_\ell(A) = \frac{2}{n} \sum_{i=1}^{n} |\mu_A(x_i) - \mu_{\underline{A}}(x_i)| = \frac{2}{n} \sum_{i=1}^{n} \mu_{A \cap \bar{A}}(x_i). \tag{2}$$

$\gamma_\ell(A)$ measures the linear distance between the fuzzy set A and its nearest ordinary set $\underline{A}$. $A \cap \bar{A}$ is the intersection between $A = \{\mu_i / x_i\}$ and its complement $\bar{A} = \{(1-\mu_i)/x_i\}$ such that

$$\mu_{A \cap \bar{A}}(x_i) = \mu_i \cap \bar{\mu}_i = \min\{\mu_i, (1-\mu_i)\} \quad \text{for all } i. \tag{3}$$

Entropy. The term entropy of A as defined by DeLuca and Termini (1972) is

$$H(A) = \frac{1}{n \ln 2} \sum_{i=1}^{n} \mathrm{Sn}(\mu_A(x_i)), \tag{4a}$$

with the Shannon function

$$\mathrm{Sn}(\mu_A(x_i)) = -\mu_A(x_i) \ln \mu_A(x_i) - (1-\mu_A(x_i)) \ln(1-\mu_A(x_i)). \tag{4b}$$

$\gamma(A)$ and $H(A)$ reflect the ambiguity present in A, such that

$$\gamma_{\min} = H_{\min} = 0 \quad \text{for } \mu_A(x_i) = 0 \text{ or } 1, \tag{5a}$$

and

$$\gamma_{\max} = H_{\max} = 1 \quad \text{for } \mu_A(x_i) = 0.5. \tag{5b}$$

It is to be mentioned here that the terms γ and H need an S function (Zadeh (1975)) for their implementation.

π-ness. As its name implies, π-ness uses, unlike γ and H, a π function to compute the ambiguity in A and is defined as (Pal (1982))

$$\pi(A) = \frac{1}{n} \sum_{i=1}^{n} G_\pi(x_i), \tag{6}$$

where G_π is any π function (Zadeh (1975)).

3. Membership functions

To obtain $\mu_A(x_i)$ from x_i, let us consider the standard S function (for computing $\gamma(A)$ and $H(A)$) and π function (for computing $\pi(A)$) which have the forms (Zadeh (1975))

$$\mu_{AS}(x_i) = \begin{cases} 0, & x_i \le a, \\ 2\left(\dfrac{x_i - a}{c-a}\right)^2, & a \le x_i \le b, \\ 1 - 2\left(\dfrac{x_i - c}{c-a}\right)^2, & b \le x_i \le c, \\ 1, & x_i \ge c, \end{cases} \tag{7}$$

in the interval $[a, c]$ with $b = \frac{1}{2}(a+c)$ and

$$\mu_{A\pi}(x_i) = \begin{cases} \mu_{AS}(x_i), & a \le x_i \le c, \\ 1 - \mu_{AS}(x_i), & c \le x_i \le a', \end{cases} \tag{8}$$

in the interval $[a, a']$ with

$$b' = \tfrac{1}{2}(a' + c), \qquad c = \tfrac{1}{2}(a + a').$$

b and b' are cross-over points, i.e.

$$\mu_{AS}(b) = \mu_{AS}(b') = 0.5,$$

and $\mu_{A\pi}(c) = 1$.

4. Intraset and Interset Ambiquity: Feature Evaluation Index

Let $c_1, c_2, \ldots, c_j, \ldots, c_m$ be the m pattern classes in an N dimensional $(x_1, x_2, \ldots, x_q, \ldots, x_N)$ feature space Q_X. Also consider that the class c_j contains a number of n_j $(j = 1, 2, \ldots m)$ samples.

For computing γ and H (equations (2) and (4)) of the qth component in the jth class, let us consider the parameters of the S function (equation (7)) as

$$b = (x_{qj})_{\mathrm{av}}, \tag{9a}$$

$$c = b + \max\{|(x_{qj})_{\mathrm{av}} - (x_{qj})_{\max}|, |(x_{qj})_{\mathrm{av}} - (x_{qj})_{\min}|\}, \tag{9b}$$

with

$$a = 2b - c,$$

where $(x_{qj})_{\mathrm{av}}$, $(x_{qj})_{\max}$ and $(x_{qj})_{\min}$ denote the mean, maximum and minimum values respectively computed along the qth co-ordinate axis over all the n_j samples in c_j.

Since $\mu(b)=\mu((x_{jq})_{av})=0.5$, the values of γ and H are 1 at $b=(x_{jq})_{av}$ and tend to zero (equation (5)) as we move away from b towards either c or a of the S function. The higher the value of γ or H, the larger would the number of samples having $\mu(x)\simeq 0.5$ be and hence the greater would the tendency of the samples be to cluster around its mean value, resulting in lesser internal scatter within the class. Therefore, the reliability (goodness) of a feature in characterising a class increases as its corresponding γ or H value within the class increases. The value of γ or H thus obtained gives a measure of Intraset Ambiguity along the qth co-ordinate axis in c_j and is denoted here by γ_{qj} or H_{qj}.

Let us now pool together the classes c_j and c_k ($j=k=1, 2, \ldots m, j\neq k$) and compure the mean, maximum and minimum values of qth dimension over the samples (n_j+n_k). The value of γ or H so computed with equation (9) would therefore decrease as the goodness of the qth feature in discriminating pattern classes c_j and c_k increases; because there would be lesser samples around the mean of the classes c_j and c_k resulting in γ or $H\simeq 1$ and more samples far from the mean giving γ or $H\simeq 0$. Let us denote this measure as γ_{qjk} or H_{qjk} and call it Interset Ambiquity along the qth dimension between the classes c_j and c_k.

Similarly, for computing π_{qj} along the qth dimension in c_j, the parameters of the π function are set as follows:

$$c=(x_{qj})_{av}, \tag{10a}$$

$$a'=c+\max\{|(x_{qj})_{av}-(x_{qj})_{max}|,\; |(x_{qj})_{av}-(x_{qj})_{min}|\}, \tag{10b}$$

with

$$a=2c-a', \quad b=\tfrac{1}{2}(a+c), \quad b'=\tfrac{1}{2}(a'+c). \tag{10c}$$

For computing π_{qjk}, the classes c_j and c_k are pooled together and the above parameters (equation (10)) are obtained from (n_j+n_k) samples.

Considering the two types of ambiguities as discussed above, the problem of evaluating feature quality therefore reduces to maximising intraset ambiguities while minimising the interset ambiguities in Q_X. The Feature Evaluation Index (FEI) for the qth feature is accordingly defined as

$$(\mathrm{FEI})_q=\frac{d_{qjk}}{d_{qj}+d_{qk}}, \tag{11}$$

$$j=k=1,2,\ldots,m,\; j\neq k,\; q=1,2,\ldots,N.$$

where d stands for γ or H or π-ness. The lower the value of $(\mathrm{FEI})_q$, the higher therefore is the quality (importance) of the qth feature in recognising and discriminating different classes.

5. Implementation and results

The above algorithm has been implemented on a set of 496 Indian Telugu vowel sounds in a consonant–vowel–consonant context of ten vowel classes

(δ,a, i,i:, u,u:, e,e:, o,o:)

including shorter and longer categories uttered by three speakers in the age group of 30 to 35 years (Pal and Dutta Majumder (1977)). First three vowel formant fragnencies F_1, F_2 and F_3 obtained through spectrum analysis are considered as the feature set for their recognition. Fig. 1 shows the feature space of vowels corresponding to F_1 and F_2.

Order of importance of formants in recognising a class as obtained with the intraclass measure based on γ, H and π-ness is illustrated in Table 1. In a part of the experiment, the shorter and longer counterparts of a vowel (differing mainly in duration) were pooled together and the corresponding results are included for comparison (Table 1). For recognising vowels (except for /E/), the first two formants are found to be much more important than F_3 (which is mainly responsible for speaker identification). The result agrees well with our previous investigations (Pal and Dutta Majumder (1977)).

Table 2 shows the order of importance of formants according to FEI values (equation (11)) for different pairs of classes. From Table 2, F_1 is seen to be more important than F_2 in discriminating the class combinations /U,O/, /I,E/, /a, U/ and /δ,U/, i.e. between /front and front/ or /back and back/ vowels. For the other combinations, i.e. discriminating between /front and back/ vowels, F_2 is found to be the strongest feature. The above

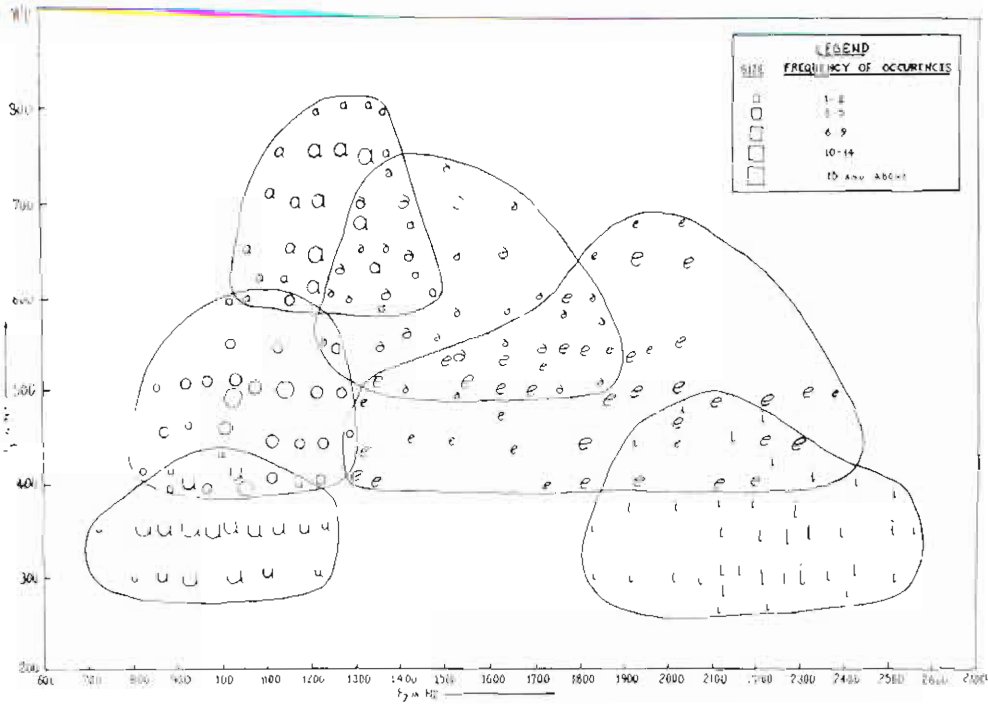

Fig. 1. Vowel diagram iu $F_1 - F_2$ space.

Table 1
Intraclass Ambiguities for feature evaluation

Vowel class		Order of importance (left to right) of formants according to Intraclass Ambiguity					
		Index of fuzziness		Entropy		π-ness	
δ		F_2 F_1 F_3	F_2 F_1 F_3	F_2 F_1 F_3	F_2 F_1 F_3	F_2 F_1 F_3	F_2 F_1 F_3
a		F_2 F_1 F_3	F_2 F_1 F_3	F_2 F_1 F_3	F_2 F_1 F_3	F_2 F_1 F_3	F_2 F_1 F_3
I	i	F_2 F_1 F_3	F_2 F_1 F_3	F_2 F_1 F_3	F_2 F_1 F_3	F_2 F_1 F_3	F_2 F_1 F_3
	i:		F_1 F_2 F_3		F_1 F_2 F_3		F_1 F_2 F_3
U	u	F_2 F_1 F_3	F_2 F_1 F_3	F_2 F_1 F_3	F_2 F_1 F_3	F_2 F_1 F_3	F_2 F_1 F_3
	u:		F_1 F_3 F_2		F_3 F_2 F_1		F_3 F_2 F_1
E	e	F_1 F_3 F_2	F_2 F_3 F_1	F_3 F_1 F_2	F_3 F_2 F_1	F_3 F_1 F_2	F_3 F_2 F_1
	e:		F_1 F_2 F_3		F_1 F_2 F_3		F_1 F_2 F_3
O	o	F_2 F_1 F_3	F_2 F_1 F_3	F_2 F_1 F_3	F_2 F_1 F_3	F_2 F_1 F_3	F_2 F_1 F_3
	o.		F_2 F_1 F_3		F_2 F_1 F_3		F_2 F_1 F_3

findings can readily be verified from Fig. 1.

Typical FEI values for F_1, F_2 and F_3 are

0.3399 0.2280 0.5418 for /a,I/

and

0.2093 0.3649 0.5970 for /a,U/,

using the parameter γ. This is shown to illustrate the relative difference in importance among the formants in characterising a class.

Similar investigations have also been made in case of a speaker identification problem using the same data set (Fig. 1) and

$$\{F_1, F_2, F_3, F_3 - F_2, F_3 - F_1, F_3/F_2, F_3/F_1\}$$

as the feature set. FEI values have been computed

Table 2
FEI values for feature evaluation

Pair of vowel classes	Order of importance (left to right) of formants according to FEI								
	Index of fuzziness			Entropy			π-ness		
δ, a	F_2	F_3	F_1	F_2	F_1	F_3	F_2	F_1	F_3
δ, I	F_2	F_1	F_3	F_2	F_1	F_3	F_2	F_1	F_3
δ, U	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
δ, E	F_2	F_1	F_3	F_2	F_1	F_3	F_2	F_1	F_3
δ, O	F_2	F_1	F_3	F_2	F_1	F_3	F_2	F_1	F_3
a, I	F_2	F_1	F_3	F_2	F_1	F_3	F_2	F_1	F_3
a, U	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
a, E	F_2	F_1	F_3	F_2	F_1	F_3	F_2	F_1	F_3
a, O	F_2	F_1	F_3	F_2	F_1	F_3	F_2	F_1	F_3
I, U	F_2	F_1	F_3	F_2	F_1	F_3	F_2	F_1	F_3
I, E	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
I, O	F_2	F_1	F_3	F_2	F_1	F_3	F_2	F_1	F_3
U, E	F_2	F_1	F_3	F_2	F_1	F_3	F_2	F_1	F_3
U, O	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
E, O	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3	F_1

for each of the three speakers individually for all the vowel classes. Contrary to the vowel recognition problem, F_3 and its combinations were found here to yield lower FEI values (i.e. more important) than F_1 and F_2 – resembling well our earlier report (Pal and Dutta Majumder (1977)).

6. Conclusion

An algorithm for automatic evaluation of feature quality in pattern recognition has been described by minimising/maximising the interclass/intraclass fnzziness as measured by the terms index of fuzziness, entropy and π-ness in a set. The algorithm is found to provide satisfactory order of importance of the formant freqnencies in characterising a vowel class, in discriminating a pair of vowel classes and also in identifying a speaker.

Since F_3 and its higher formants ($F_4, F_5, \ldots$) are mostly responsible for identifying a speaker, we have considered in our experiment only F_3 in addition to F_1 and F_2 for evaluating feature quality in vowel recognition problems.

Acknowledgement

The authors wish to acknowledge Professor D. Dutta Majumder for his interest in this work and Mr. N. Chatterjee for typing the manuscript.

References

Fukunaga, K (1972). *Introduction to Statistical Pattern Recognition and Machine Learning*. Academic Press, New York.

Tou, J.T. and R.C. Gonzalez (1974). *Pattern Recognition Principles*. Addison-Wesley, Reading, MA.

Kaufmann, A. (1975). *Introduction to the Theory of Fuzzy Subsets – Fundamental Theoretical Elements*, Vol. 1. Academic Press, New York.

De Luca, A and S. Termini (1972). A definition of a nonprobabilistic entropy in the setting of fuzzy set theory. *Inform. and Control* 20, 301–312.

Pal, S.K. (1982) Fuzzy set theoretic approach – A tool for speech and iuage recognition. In: J. Kittler, et al., Eds., *Pattern Recognition Theory and Applications*. Reidel, Dordrecht, 103–117.

Zadeh, L.A. (1975). Calculus of fuzzy restrictions. In: L. Zadeh, et al., Eds., *Fuzzy Sets and their Application to Cognitive and Decision Processes*. Academic Press, London, 1–39.

Pal, S.K. and D. Dutta Majumder (1977). Fuzzy sets and decision making approaches in vowel and speaker recognition. *IEEE Trans. Systems Man Cybernet.* 7, 625–629