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 $\pi$-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 problem of evaluating feature quality. The terms Index of fuzziness (Kaufmann (1975)), Entropy (DeLuca and Termini (1972)) and $\pi$-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 algorithm is demonstrated on a vowel sound recognition problem using the first three formants along with $S$ and $\pi$ 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

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

where the characteristic function $\mu_A(x)$ known as membership function having positive value in the interval $[0,1]$ denotes the degree to which an event $x$ 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 fuzzy set $A$. Let $A$ be defined as the nearest ordinary set to $A$, such that
\[ \mu_A(x_i) = \begin{cases} 0 & \text{if } \mu_A(x_i) \leq 0.5, \\ 1 & \text{if } \mu_A(x_i) > 0.5. \end{cases} \]

**Index of Fuzziness.** The linear index of fuzziness \( \gamma_i(A) \) of \( A \) is defined as (Kaufmann (1975))

\[ \gamma_i(A) = \frac{2}{n} \sum_{i=1}^{n} |\mu_A(x_i) - \mu_A(x_i)| = \frac{2}{n} \sum_{i=1}^{n} \mu_A \cap \bar{A}(x_i). \quad (2) \]

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

\[ \mu_{A \cap \bar{A}}(x_i) = \mu_i \cap \bar{\mu}_i, \]

\[ = \min \{ \mu_i, (1 - \mu_i) \} \quad \text{for all } i. \quad (3) \]

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

\[ H(A) = \frac{1}{n \ln 2} \sum_{i=1}^{n} S_n(\mu_A(x_i)), \quad (4a) \]

with the Shannon function

\[ S_n(\mu_A(x_i)) = -\mu_A(x_i) \ln \mu_A(x_i) - (1 - \mu_A(x_i)) \ln (1 - \mu_A(x_i)). \quad (4b) \]

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

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

and

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

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

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

\[ \pi(A) = \frac{1}{n} \sum_{i=1}^{n} G_{\pi}(x_i), \quad (6) \]

where \( G_{\pi} \) is any \( \pi \) 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 \( \pi \) function (for computing \( \pi(A) \)) which have the forms (Zadeh (1975))

\[ \mu_{AS}(x_i) = \begin{cases} 0, & x_i \leq a, \\ 2 \left( \frac{x_i - a}{c - a} \right)^2, & a \leq x_i \leq b, \\ 1 - 2 \left( \frac{c - x_i}{c - a} \right)^2, & b \leq x_i \leq c, \\ 1, & x_i \geq c. \end{cases} \quad (7) \]

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

\[ \mu_{AS}(x_i) = \begin{cases} \mu_{AS}(x_i), & a \leq x_i \leq c, \\ 1 - \mu_{AS}(x_i), & c \leq x_i \leq a'. \end{cases} \quad (8) \]

in the interval \([a, a']\) with

\[ b' = \frac{1}{2}(a' + c), \quad c' = \frac{1}{2}(a + a'). \]

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

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

and \( \mu_{AS}(c) = 1. \)

### 4. Intraset and Interse! Ambiquity: Feature Evaluation Index

Let \( c_1, c_2, \ldots, c_m \) be the \( m \) pattern classes in an \( N \) dimensional \((x_1, x_2, \ldots, x_N)\) 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 \( \gamma \) and \( H \) (equations (2) and (4)) of the \( q \)th component in the \( j \)th class, let us consider the parameters of the \( S \) function (equation (7)) as

\[ b = (x_{qj})_{av}, \quad (9a) \]

\[ c = b + \max \{|(x_{qj})_{av} - (x_{qj})_{max}|, \quad (9b) \]

\[ |(x_{qj})_{av} - (x_{qj})_{min}| \],

with

\[ a = 2b - c, \]

where \((x_{qj})_{av}, (x_{qj})_{max}\) and \((x_{qj})_{min}\) denote the mean, maximum and minimum values respectively computed along the \( q \)th co-ordinate axis over all the \( n_j \) samples in \( c_j \).
Since \( \mu(b) = \mu((x_q)_b) = 0.5 \), the values of \( \gamma \) and \( H \) are 1 at \( b = (x_q)_b \), and tend to zero (equation (5)) as we move away from \( b \) towards either \( c \) or \( a \) of the \( \Sigma \) function. The higher the value of \( \gamma \) or \( H \), the larger would be the number of samples having \( \mu(x) = 0.5 \) be and hence the greater would be 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 \( \gamma \) or \( H \) value within the class increases. The value of \( \gamma \) or \( H \) thus obtained gives a measure of Inraset Ambiguity along the \( q \)th co-ordinate axis in \( c \) and is denoted here by \( \gamma_q \) or \( H_q \).

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

Similarly, for computing \( \pi_{qj} \) along the \( q \)th dimension in \( c_j \), the parameters of the \( \pi \) function are set as follows:

\[
\begin{align*}
  c &= (x_{qj})_n, \\
  a' &= c + \max \left\{ \left( \frac{x_{qj}}{a'} - (x_{qj})_{\max} \right), \left( (x_{qj})_{\max} - (x_{qj})_{\min} \right) \right\}, \\
  a &= 2c - a', \quad b = \frac{1}{2}(a + c), \quad b' = \frac{1}{2}(a' + c).
\end{align*}
\]

For computing \( \pi_{qjk} \), 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 interset ambiguities while minimising the inrerset ambiguities in \( Q \). The Feature Evaluation Index (FEI) for \( q \)th feature is accordingly defined as

\[
(FEI)_q = \frac{d_{qjk}}{d_{qj} + d_{qk}},
\]

\( j = k = 1, 2, \ldots, m, j \neq k, \quad q = 1, 2, \ldots, N, \)

where \( d \) stands for \( \gamma \) or \( H \) or \( \pi \)-ness. The lower the value of \( (FEI)_q \), the higher therefore is the quality (importance) of the \( q \)th 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

\((\delta, a, i, i:, u, u:, e, e:, o, o:)\)

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 frequencies \( F_1, F_2 \) 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 \( \gamma \), \( H \) and \( \pi \)-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 \( /\delta, U/ \), 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
Table 1
Intraclass Ambiguities for feature evaluation

<table>
<thead>
<tr>
<th>Vowel class</th>
<th>Order of importance (left to right) of formants according to Intraclass Ambiguity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Index of fuzziness</td>
</tr>
<tr>
<td>( \delta )</td>
<td>( F_3, F_2, F_1 )</td>
</tr>
<tr>
<td>( a )</td>
<td>( F_3, F_2, F_1 )</td>
</tr>
<tr>
<td>( i )</td>
<td>( F_2, F_1, F_3 )</td>
</tr>
<tr>
<td>( u )</td>
<td>( F_2, F_1, F_3 )</td>
</tr>
<tr>
<td>( E )</td>
<td>( F_1, F_3, F_2 )</td>
</tr>
<tr>
<td>( o )</td>
<td>( F_1, F_3, F_2 )</td>
</tr>
</tbody>
</table>

findings can readily be verified from Fig. 1. Typical FEI values for \( F_1, F_2 \) and \( F_3 \) are

\[
0.3399 \quad 0.2280 \quad 0.5418
\]

for \( /a, i/ \).

and

\[
0.2093 \quad 0.3649 \quad 0.5970
\]

for \( /a, U/ \), using the parameter \( \gamma \). 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

\[ \{F_1, F_2, F_3, F_3 - F_2, F_2 - F_1, F_3/F_2, F_2/F_1 \} \]

as the feature set. FEI values have been computed.
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 fuzziness as measured by the terms index of fuzziness, entropy and $\pi$-ness in a set. The algorithm is found to provide satisfactory order of importance of the formant frequencies 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


