Pattern Recognition Letters 9 (1989) 321–325
North-Holland June 1989

On looseness of error bounds provided by the generalized separability measures of Lissack and Fu

S. RAY
Electronics and Communication Sciences Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Calcutta 700035, India

Received 25 October 1988

Abstract: An expression is obtained for maximum difference between the upper and the lower bounds to Bayesian probability of error in terms of the generalized separability measures of Lissack and Fu (L_α). The expression gives the magnitude of looseness of error bounds for different values of α.

Key words: Pattern recognition, feature evaluation, Bayesian probability of error, error bounds, probabilistic criteria, separability measures.

1. Introduction

The Bayesian probability of error (P_e) is an optimum measure of effectiveness of a set of features selected for the purpose of pattern recognition. Owing to the difficulty involved in computation (or estimation) of P_e, various probabilistic separability criteria have been suggested in the past as indirect measures of feature effectiveness [1, Ch. 7]. The generalized separability measures (L_α, $0 < \alpha < \infty$), suggested by Lissack and Fu [2], are one such series of feature effectiveness measures defined in terms of the difference between the a posteriori probabilities of pattern classes.

It is worth noting that both the upper and the lower bounds to P_e in terms of a measure are indicative of how closely the measure approximates P_e. If the resulting upper bound is sufficiently low, then the set of features under consideration are 'acceptable'. On the other hand, a sufficiently high lower bound leads to a 'rejection' decision. Difference between the upper bound and the lower bound is an indicator of the overall closeness of a measure to P_e. In this letter some results are proved from which one can know the magnitude of the looseness of the existing P_e bounds provided by L_α.

2. Error bounds in terms of L_α

Suppose the a priori probabilities of the two classes ω_1 and ω_2 are π_1 and π_2, respectively ($0 < \pi_1, \pi_2 < 1$, $\pi_1 + \pi_2 = 1$). Let $p(x\,|\,\omega_1)$ and $p(x\,|\,\omega_2)$ be the class-conditional probability density functions of the feature vector X, assumed to be continuous, in the two classes ω_1 and ω_2, respectively. Then the Bayesian error probability [1, Ch. 2] is given by

$$P_e = \int_{\Omega_X} \min[\pi_1\, p(x\,|\,\omega_1),\ \pi_2 p(x\,|\,\omega_2)]\, dx \tag{1}$$

and the generalized seperability measure proposed by Lissack and Fu [2] is defined by

$$L_\alpha = \int_{\Omega_X} |P(\omega_1|x) - P(\omega_2|x)|^\alpha p(x)\,dx, \quad 0 < \alpha < \infty \tag{2}$$

where $P(\omega_i|x)$, $i = 1,2$ is the a posteriori probability of ω_i given $X = x$, Ω_X denotes the sample space of X and

$$p(x) = \pi_1 p(x|\omega_1) + \pi_2 p(x|\omega_2) \tag{3}$$

denotes the mixture density of X.

It can be seen that L_α is a straightforward generalization of the Kolmogorov variational distance [3] defined by

$$K = \tfrac{1}{2} \int_{\Omega_X} |\pi_1 p(x|\omega_1) - \pi_2 p(x|\omega_2)|\,dx \tag{4}$$

and, for $\alpha = 1$, the measure L_α reduces to $2K$. In this case,

$$L_1 = 2K = 1 - 2P_e. \tag{5}$$

Lissack and Fu [2] obtained the following error bounds. For $0 < \alpha \le 1$,

$$\tfrac{1}{2}\{1 - L_\alpha\} \le P_e \le \tfrac{1}{2}\{1 - [L\alpha]^{1/\alpha}\} \tag{6}$$

and for $1 \le \alpha < \infty$,

$$\tfrac{1}{2}\{1 - [L_\alpha]^{1/\alpha}\} \le P_e \le \tfrac{1}{2}\{1 - L_\alpha\}. \tag{7}$$

3. Looseness of error bounds

As indicated above, for $\alpha = 1$ the lower and the upper bounds coincide. An increase or decrease in the value of α loosens the bounds. From the following theorem one can obtain information about the magnitude of the loosening of the bounds depending on the value of α.

Theorem. (i) *For a given $\alpha > 1$ the maximum value of δ (= upper bound − lower bound) is given by*

$$\delta_{\max} = \tfrac{1}{2}\{\alpha^{-1/(\alpha-1)} - \alpha^{-\alpha/(\alpha-1)}\}. \tag{8}$$

(ii) *And the value of $\delta_{\max}$ increases with increase in α.*

Proof. (i) For $\alpha > 1$,

$$\delta = \tfrac{1}{2}\{1 - L_\alpha\} - \tfrac{1}{2}\{1 - [L_\alpha]^{1/\alpha}\} = \tfrac{1}{2}\{[L_\alpha]^{1/\alpha} - L_\alpha\}. \tag{9}$$

Differentiating δ with respect to L_α one gets

$$\frac{\mathrm{d}\delta}{\mathrm{d}L_\alpha} = \tfrac{1}{2}\{(1/\alpha)[L_\alpha]^{1/\alpha - 1} - 1\}. \tag{10}$$

Equating the above expression to zero leads to

$$L_\alpha = \alpha^{-\alpha/(\alpha - 1)}. \tag{11}$$

It is easy to see that

$$\frac{\mathrm{d}^2\delta}{\mathrm{d}L_\alpha^2} < 0. \tag{12}$$

Thus, the maximum value of δ occurs at the value of L_α given in equation (11). Putting this value of L_α in (9) gives

$$\delta_{\max} = \tfrac{1}{2}\{[\alpha^{-\alpha/(\alpha-1)}]^{1/\alpha} - \alpha^{-\alpha/(\alpha-1)}\} = \tfrac{1}{2}\{\alpha^{-1/(\alpha-1)} - \alpha^{-\alpha/(\alpha-1)}\}.$$

(ii) Differentiating $\delta_{\max}$ with respect to α,

$$\frac{\mathrm{d}\delta_{\max}}{\mathrm{d}\alpha} = \frac{1}{2}\left\{\alpha^{-1/(\alpha-1)}\left[\frac{\log\alpha}{(\alpha-1)^2} - \frac{1}{(\alpha-1)\alpha}\right] - \alpha^{-\alpha/(\alpha-1)}\left[\frac{\log\alpha}{(\alpha-1)^2} - \frac{1}{\alpha-1}\right]\right\}$$

$$= \frac{1}{2}\left\{\frac{\log\alpha}{(\alpha-1)^2}\left[\alpha^{-1/(\alpha-1)} - \alpha^{-\alpha(\alpha-1)}\right] + \frac{1}{\alpha-1}\left[\alpha^{-\alpha/(\alpha-1)} - \frac{1}{\alpha}\alpha^{-1/(\alpha-1)}\right]\right\}. \tag{13}$$

Using the identity $\alpha/(\alpha - 1) = 1 + 1/(\alpha - 1)$ in (13) leads to

$$\frac{\mathrm{d}\delta_{\max}}{\mathrm{d}\alpha} = \frac{1}{2}\left\{\frac{\log\alpha}{(\alpha-1)^2}\alpha^{-1/(\alpha-1)}\left(1 - \frac{1}{\alpha}\right) + \frac{1}{\alpha-1}\left[\frac{1}{\alpha}\alpha^{-1/(\alpha-1)} - \frac{1}{\alpha}\alpha^{-1/(\alpha-1)}\right]\right\}$$

$$= \frac{1}{2}\left\{\frac{\log\alpha}{(\alpha-1)^2}\alpha^{-1/(\alpha-1)}\left(1 - \frac{1}{\alpha}\right)\right\}. \tag{14}$$

It is easy to see that the expression in the right hand side of (14) is positive. Hence the desired result is proved. □

For a given $\alpha > 1$ the upper and the lower bounds of P_e corresponding to the maximum difference between the two bounds are given by

$$P_e^U = \tfrac{1}{2}\,(1 - \alpha^{-\alpha/(\alpha-1)}) \tag{15}$$

and

$$P_e^L = \tfrac{1}{2}\,(1 - \alpha^{-1/(\alpha-1)}). \tag{16}$$

Figure 1 shows how the values of P_e^U and P_e^L vary with α. It may be noted that as α increases from 1 to

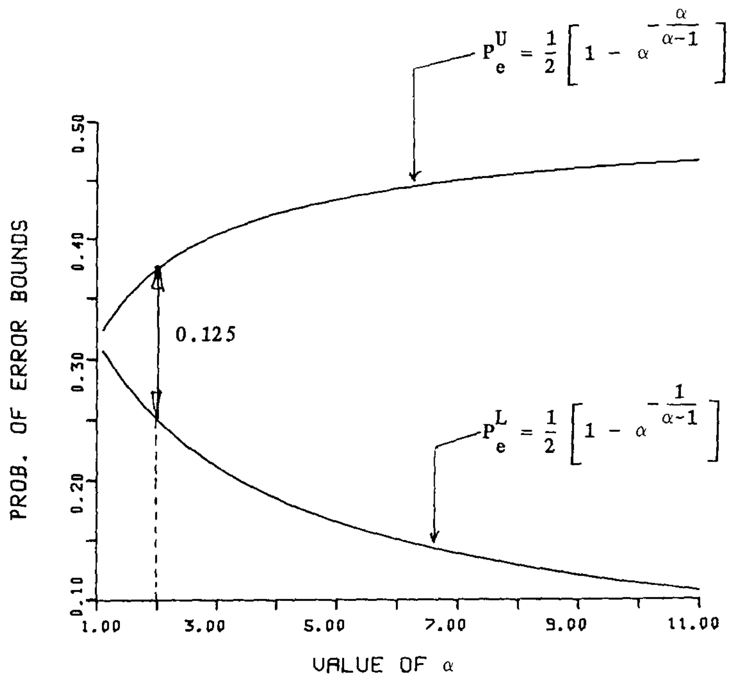

Figure 1. Looseness in probability of error (P_e) bounds given by L_α for different values of $\alpha \geq 1$.

∞ the maximum difference between the two bounds increases from 0 to 0.5. This shows how the bounds loosen with increasing α. With increasing α the computation of L_α becomes more demanding. Therefore, it appears that there is no advantage in going for high values of α. The bounds corresponding to L_2 ($\alpha = 2$) are tighter than most of the existing bounds associated with the other two-class measures. As can be seen from Figure 1 the maximum difference between the two bounds in this case is 0.125. L_2 has the advantage over L_1 in that L_2 involves the operation of raising $P(\omega_1|x) - P(\omega_2|x)$ to the power of 2 which is mathematically more handy to deal with than the difference operation involved in L_1.

It is easy to verify that, following a similar procedure as in the theorem above, for $0 < \alpha < 1$, the maximum difference between the two bounds increases from 0 to 0.5 with decrease in the value of α from 1 to 0.

4. Concluding remarks

The maximum difference between the upper and the lower bounds to P_e in terms of L_α monotonically increases from 0 to 0.5 as the value of α increases from 1 to ∞ or it decreases from 1 to 0. L_1 is directly related to P_e. In a two-class pattern recognition problem, therefore, it makes no difference whether we use P_e or L_1. Mathematical treatment of L_2 is more convenient than that of L_1. Moreover, L_2 has computational advantage over other L_α's ($\alpha \neq 2$). As a result of closer relationship with P_e and computational advantages, in feature evaluation L_1 and L_2 are favoured over other L_α's.

Acknowledgement

Typing of the manuscript by Mr. J. Gupta is gratefully acknowledged. Thanks are due to Prof. D. Dutta Majumder for his interest in the work.

References

[1] Devijver, P.A. and J. Kittler (1982). *Pattern Recognition: A Statistical Approach*. Prentice-Hall, London.

[2] Lissack, T. and K.S. Fu (1956). Error estimation in pattern recognition via L^{α}-distance between posterior density functions. *IEEE Trans. Inform. Theory* 22, 34–54.

[3] Kailath, T. (1967). The divergence and Bhattacharyya distance measures in signal selection. *IEEE Trans. Commun. Technol.* 15, 52–60.