Pattern Recognition Letters 9 (1989) 315–319
North-Holland June 1989

On a theoretical property of the Bhattacharyya coefficient as a feature evaluation criterion

S. RAY
Electronics & Communication Sciences Unit, Indian Statistical Institute, Calcutta 700035, India

Received 19 October 1988

Abstract: In the context of evaluation of features in a two-class pattern recognition problem it is shown that, irrespective of the values of the a priori probabilities of the two classes, the maximum difference between the existing lower and upper bounds to Bayesian probability of error in terms of the Bhattacharyya coefficient is approximately equal to 0.2071.

Key words: Bhattacharyya coefficient, Bayesian probability of error, pattern recognition, feature evaluation, error bounds.

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 indirect measures of feature effectiveness have been suggested in the past. The Bhattacharyya coefficient (ρ), which was originally defined as a measure of overlap between two probability distributions [1], has become a popular feature evaluation criterion in pattern recognition [2–12]. The two main reasons behind the popularity of ρ as a feature evaluation criterion are that (a) lower and upper bounds to P_e exist in terms of ρ [13–14] and that (b) closed-form expressions are available for ρ in the case of the exponential family of distributions, a few special cases of the family being the Gaussian distribution, the Multinomial distribution and the Poisson distribution [2]. 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 pattern recognition system under consideration is '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 it is shown that the maximum difference between the existing upper bound and the Hudimoto lower bound to P_e in terms of ρ remains the same for all values of the a priori probabilities and this maximum difference is $(\sqrt{2}-1)/2$.

2. Derivation of the span between error bounds

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 [15] is given by

$$\rho_e = \int_{\Omega_X} \min[\pi_1 p(x\,|\,\omega_1),\ \pi_2 p(x\,|\,\omega_2)]\,\mathrm{d}x \qquad (1)$$

and the Bhattacharyya coefficient [1] is defined by

$$\rho = \int_{\Omega_X} [p(x \mid \omega_1)\, p(x \mid \omega_2)]^{1/2}\, dx \tag{2}$$

where Ω_X denotes the sample space of X.

Clearly, $0 \le \rho \le 1$.

Hudimoto [13,14] showed that P_e is bounded above and below by the following relationships:

$$\tfrac{1}{2} - \tfrac{1}{2}\sqrt{1 - 4\pi_1\pi_2\rho^2} \le P_e \le \sqrt{\pi_1\pi_2}\,\rho. \tag{3}$$

These are the tightest bounds to P_e in terms of ρ available in the literature.

The span between the above upper and lower bounds to P_e (i.e., the difference between them) is given by

$$\delta = \sqrt{\pi_1\pi_2}\,\rho - \tfrac{1}{2} + \tfrac{1}{2}\sqrt{1 - 4\pi_1\pi_2\rho^2}. \tag{4}$$

In the following theorem the maximum value of δ is determined which is found to be independent of the value of π_1 (and so π_2).

Theorem. *Whatever be the value of the a priori probability* π_1, (i) *the maximum value of* δ *is* $\tfrac{1}{2}(\sqrt{2} - 1)$ *and* (ii) *this maximum can be attained for* π_1*-values (and so* π_2-values) *lying inside the interval*

$$\left[\frac{2-\sqrt{2}}{4}, \frac{2+\sqrt{2}}{4}\right].$$

Proof. (i) Taking the first derivative of δ with respect to ρ one gets

$$\frac{d\delta}{d\rho} = \sqrt{\pi_1\pi_2} - \frac{2\pi_1\pi_2\rho}{\sqrt{1 - 4\pi_1\pi_2\rho^2}}.$$

Equating the above expression to 0 leads to

$$\rho = \frac{1}{\sqrt{8\pi_1\pi_2}}. \tag{5}$$

The second derivative of δ with respect to ρ is

$$\frac{d^2\delta}{d\rho^2} = -\frac{2\pi_1\pi_2}{(1 - 4\pi_1\pi_2\rho^2)^{3/2}}. \tag{6}$$

It is easy to see that

$$\frac{d^2\delta}{d\rho^2} < 0. \tag{7}$$

Thus, the maximum value of δ is attained at the value of ρ given in equation (5). Substituting this value of ρ in equation (4) gives,

$$\begin{aligned}\delta_{\max} &= \sqrt{\pi_1\pi_2}\cdot\frac{1}{\sqrt{8\pi_1\pi_2}} - \tfrac{1}{2} + \tfrac{1}{2}\sqrt{1 - 4\pi_1\pi_2\cdot\frac{1}{8\pi_1\pi_2}}\\ &= \frac{1}{\sqrt{8}} - \tfrac{1}{2} + \tfrac{1}{2}\sqrt{1 - \tfrac{1}{2}}\\ &= \tfrac{1}{2}(\sqrt{2} - 1).\end{aligned}$$

Hence the first part of the theorem is proved.

(ii) The above mentioned maximum value occurs at

$$\rho = \frac{1}{\sqrt{8\pi_1\pi_2}}.$$

But ρ is restricted by the condition $\rho \le 1$. Substitution of the above value of ρ in the inequality $\rho \le 1$ leads to the desired result.

In Figures 1 to 4 the P_e bounds for different values of π_1, namely, $\pi_1 = 0.500$, 0.625, 0.750 and 0.875, are shown. It can be seen from these figures that the value of ρ, for which the maximum value of δ $(= \tfrac{1}{2}(\sqrt{2} - 1) \simeq 0.2071)$ is attained, gets shifted towards the right with increases in the value of π_1. For values of $\pi_1 > (2 + \sqrt{2})/4$ the maximum would occur at a value of ρ outside its range (Figure 4). The same situation would arise for $\pi_1 < (2 - \sqrt{2})/4$.

3. Conclusion

For all values of the a priori probabilities of the two classes in the range $(2 - \sqrt{2})/4$ to $(2 + \sqrt{2})/4$ the maximum difference between the existing tightest lower and upper bounds to P_e in terms of the Bhattacharyya coefficient (ρ) is the same and it is

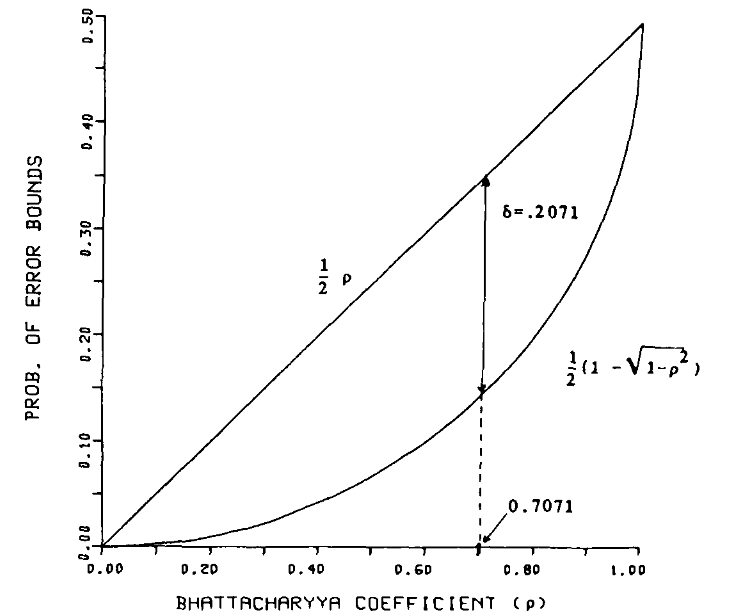

Figure 1. Probability of error (P_e) bounds in terms of the Bhattacharyya coefficient (ρ) for $\pi_1 = \pi_2 = 0.500$

equal to $(\sqrt{2}-1)/2$. In other words, irrespective of the values of the two a priori probabilities, the span of the two P_e bounds in terms of the Bhattacharyya coefficient cannot exceed 0.2071 (approximately). The span is smaller for a priori probabilities outside the range mentioned above. This result would give further insight into the applicability of Bhattacharyya coefficient as a feature evaluation criterion.

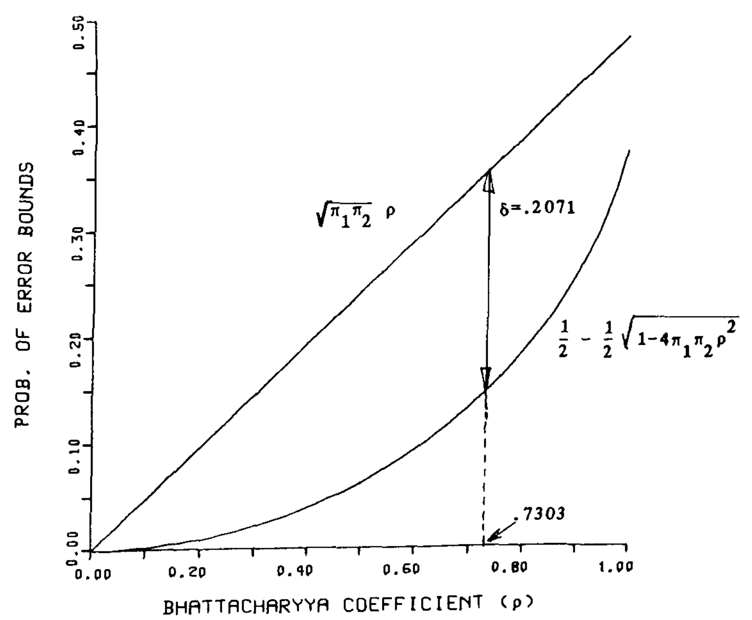

Figure 2. Probability of error (P_e) bounds in terms of the Bhattacharyya coefficient (ρ) for $\pi_1 = 0.625$, $\pi_2 = 0.375$

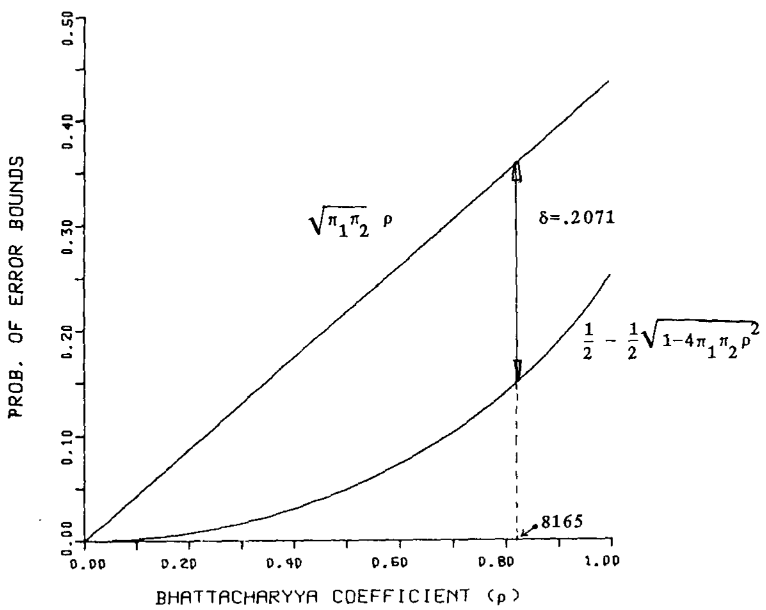

Figure 3. Probability of error (P_e) bounds in terms of the Bhattacharyya coefficient (ρ) for $\pi_1 = 0.750$ and $\pi_2 = 0.250$.

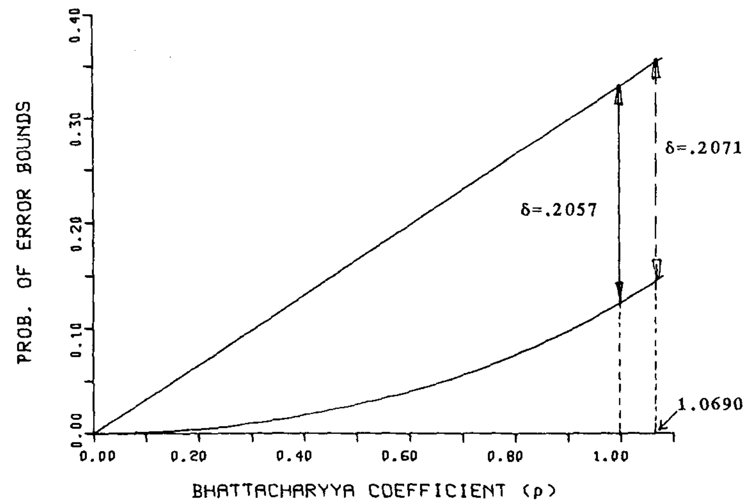

Figure 4. Probability of error (P_e) bounds in terms of the Bhattacharyya coefficient (ρ) for $\pi_1 = 0.875$ and $\pi_2 = 0.125$.

Acknowledgement

Typing of this manuscript by Smt. S. De Bhowmick is gratefully acknowledged.

References

[1] Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc.* 35, 99–109.

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

[3] Lainiotis, D.G. and T.L. Henderson (1968). Application of state variable techniques to optimal feature extraction. *Proc. IEEE* 5, 2175–2176.

[4] Lainiotis, D.G. (1969). A class of upper bounds on the probability of error for multihypothesis pattern recognition. *IEEE Trans. Inform. Theory* 15, 730–731.

[5] Lainiotis, D.G. and S.K. Park (1971). Probability of error bounds. *IEEE Trans. Syst. Man Cybernet* 1, 175–178.

[6] Matusita, K. (1971). Some properties of affinity and applications. *Ann. Inst. Stat. Math.* 23, 137–155.

[7] Swain, P.H., T.V. Robertson and A.G. Wacker (1971). Comparison of the divergence and B-distance in feature selection. *LARS Information Note 020871*, Purdue Univ., West Lafayette, IN.

[8] Swain, P.H. and R.C. King (1973). Two effective feature selection criteria for multispectral remote sensing. *LARS Information Note 042673*, Purdue Univ., West Lafayette, IN.

[9] Toussaint, G.T. (1972). Feature evaluation with a proposed generalization of Komolgorov's variational distance and the Bhattacharyya coefficient. *Proc. Annual Canadian Conference*, Montreal, June 1–3, 1972, 422401–422413.

[10] Toussaint, G.T. (1974). On some measures of information and their application in pattern recognition. *Proc. Measures of Information and their Applications*, Indian Institute of Tech., Bombay, August 16–18, 1974, 21–28.

[11] Chittineni, C.B. (1972). On the probability of error and the expected Bhattacharyya distance in multiclass pattern recognition. *Proc. IEEE* 11, 1451–1452.

[12] Chittineni, C.B. (1980). Efficient feature subset selection with probabilistic distance criteria. *Information Sciences* 22, 19–35.

[13] Hudimoto, H. (1956–57). On the distribution-free classification of an individual into one of two groups. *Ann. Inst. Stat. Math.* 8, 105–112.

[14] Hudimoto, H. (1957–58). A note on the probability of the correct classification when the distributions are not specified. *Ann. Inst. Stat. Math.* 9, 31–36.

[15] Fukunaga, K. (1972). *Introduction to Statistical Pattern Recognition*. Academic Press, New York.