KYBERNETIKA — VOLUME 34 (1998), NUMBER 4, PAGES 363–368

THE BHATTACHARYYA METRIC AS AN ABSOLUTE SIMILARITY MEASURE FOR FREQUENCY CODED DATA

FRANK J. AHERNE, NEIL A. THACKER AND PETER I. ROCKETT

This paper highlights advantageous properties of the Bhattacharyya metric over the chi-squared statistic for comparing frequency distributed data. The original interpretation of the Bhattacharyya metric as a geometric similarity measure is reviewed and it is pointed out that this derivation is independent of the use of the Bhattacharyya measure as an upper bound on the probability of misclassification in a two-class problem. The affinity between the Bhattacharyya and Matusita measures is described and we suggest use of the Bhattacharyya measure for comparing histogram data. We explain how the chi-squared statistic compensates for the implicit assumption of a Euclidean distance measure being the shortest path between two points in high dimensional space. By using the square-root transformation the Bhattacharyya metric requires no such standardization and by its multiplicative nature has no singularity problems (unlike those caused by the denominator of the chi-squared statistic) with zero count-data.

1. INTRODUCTION

When we wish to compare one data set to a known distribution, or to compare two equally unknown distributions, a commonly used technique is to apply the chi-squared statistic. In this paper we propose the Bhattacharyya metric as an alternative similarity measure when comparing Poisson distributed data and we demonstrate certain advantages of this measure over the chi-squared statistic. Also we give an example where the chi-squared statistic is not an accurate comparison measure over large distances in a statistical pattern space. We demonstrate this for the case of 2-dimensional measurements having Poisson errors.

2. BHATTACHARYYA'S ORIGINAL DERIVATION

Bhattacharyya's original interpretation of the measure was geometric [2]. He considered two multinomial populations each consisting of k classes with associated probabilities $p_1, p_2, \ldots, p_k$, and $p'_1, p'_2, \ldots, p'_k$, respectively. Then, as $\sum_{i=1}^{k} p_i = 1$ and $\sum_{i=1}^{k} p'_i = 1$, he noted that $(\sqrt{p_1}, \ldots, \sqrt{p_k})$ and $(\sqrt{p'_1}, \ldots, \sqrt{p'_k})$ could be considered as the direction cosines of two vectors in k-dimensional space referred to a system of orthogonal co-ordinate axes. As a measure of divergence between the two populations Bhattacharyya used the square of the angle between the two position vectors. If θ is the angle between the vectors then:

$$\cos(\theta) = \sum_{i=1}^{k} \sqrt{p_i p_i'}. \tag{2.1}$$

Thus if the two populations are identical we have equation (2.2):

$$\cos(\theta) = \sum_{i=1}^{k} p_i = 1 \tag{2.2}$$

corresponding to $\theta = 0$, hence we see the intuitive motivation behind the definition as the vectors are collinear.

3. THE BHATTACHARYYA BOUND

In this section we consider the Bhattacharyya bound as commonly used in pattern recognition. Consider a two-class problem where each sample belongs to one of two mutually exclusive classes (the conditional density functions and the *a priori* probabilities are assumed known). The sample serves as input to a decision rule whereby we classify each sample to one of the two classes. In general, decision rules do not lead to perfect classification and in order to evaluate the performance of a decision rule we must calculate the probability of error – that is the probability that the sample is assigned to the wrong class. If we define the *a posteriori* probability of class I given x as $P(\omega_I|x)$ and similarly $P(\omega_{II}|x)$ for class II then the conditional error $r(x)$, given x, is either $P(\omega_I|x)$ or $P(\omega_{II}|x)$ (whichever is smaller), as described by Fukunaga ([4]). That is:

$$r(x) = \min[P(\omega_I|x), P(\omega_{II}|x)] \tag{3.1}$$

The total error, (or *Bayes error*) is computed by the expectation of $r(x)$, $E[r(x)]$:

$$E[r(x)] = \int_x r(x)\, P(x)\, \mathrm{d}x \tag{3.2}$$

where $P(x)$ is the probability of observing the pattern $X = x$. An upper bound on the above integrand can be obtained by making use of the fact that $\min[a, b] \le a^s b^{1-s}$ for all $0 \le s \le 1$, $a, b \ge 0$. This is commonly known as the Chernoff bound and taking the case of $s = 0.5$ gives the Bhattacharyya bound (3.3):

$$\int_x P(\omega_I|x)^{0.5} P(\omega_{II}|x)^{0.5} P(x)\, \mathrm{d}x \tag{3.3}$$

or equivalently (3.4):

$$\int_x f_1(x)^{0.5} f_2(x)^{0.5}\, \mathrm{d}x \tag{3.4}$$

where $f_1(x) = P(\omega_I|x)\,P(x)$ and $f_2(x) = P(\omega_{II}|x)\,P(x)$. Thus the Bhattacharyya bound integrates over all positions in the domain and assumes that the sample belongs to only one of the two classes. This assumption is a major restriction on the scope of the method as it should strictly only be applied to simple two class problems where this is known to be the case. It is therefore used as a relative separation measure. Below we propose an alternative interpretation of the Bhattacharyya measure which has far wider potential for application.

4. THE MATUSITA METRIC

The Matusita ([5]) distance between two probability density functions is defined by:

$$\int_{-\infty}^{\infty} \left(\sqrt{f_A(x)} - \sqrt{f_B(x)}\right)^2 \mathrm{d}x \tag{4.1}$$

and is related to the Bhattacharyya metric by (4.2):

$$\int_{-\infty}^{\infty} \left(\sqrt{f_A(x)} - \sqrt{f_B(x)}\right)^2 \mathrm{d}x = 2 - 2\int_{-\infty}^{\infty} \sqrt{f_A(x)}\sqrt{f_B(x)}\,\mathrm{d}x. \tag{4.2}$$

5. EUCLIDEAN DISTANCE ASSUMPTION OF χ^2 STATISTICS

In the field of pattern recognition we often need to determine the similarity between two observations in a high dimensional space. In such domains the way the errors on the observations or measurements vary over the space can influence the shortest path between two observations. For example consider the case of Poisson distributed measurements in two dimensions. Since the mean and variance of a Poisson distribution are equal the space can be considered as a region of smoothly changing variance as illustrated by the ellipses of Figure 1(a). For example, the bottom row of ellipses correspond to a fixed error in the y direction paired with an increasing error in the x direction. In this space the shortest distance between two observations that are close to each other can be reasonably approximated by a straight line. However, for two distant observations the shortest distance between them is not necessarily a straight line but a curved path as shown by the dotted line in Figure 1(a). This presents obvious difficulties when trying to construct a simple similarity measure in such a space. Moreover, if a chi-squared statistic (such as (6.1)) was used in such a case the Euclidean distance term in the numerator of the statistic implicitly assumes that a straight line distance is the shortest path between the observations, when in fact a curved path integral would be a more accurate description.

6. ADVANTAGES OF THE BHATTACHARYYA STATISTIC

The square-root transformation is frequently used throughout statistics for producing a homogeneous variance for count data [3] and it can be shown that for count data following a Poisson distribution, the square-root transformation maps all errors to a constant. Moreover, by mapping to a domain where all errors are constant the problem of evaluating the minimum of a curved path integral is avoided by ensuring that a straight line measure is always the minimum distance between two observations – as shown in Figure 1(b).

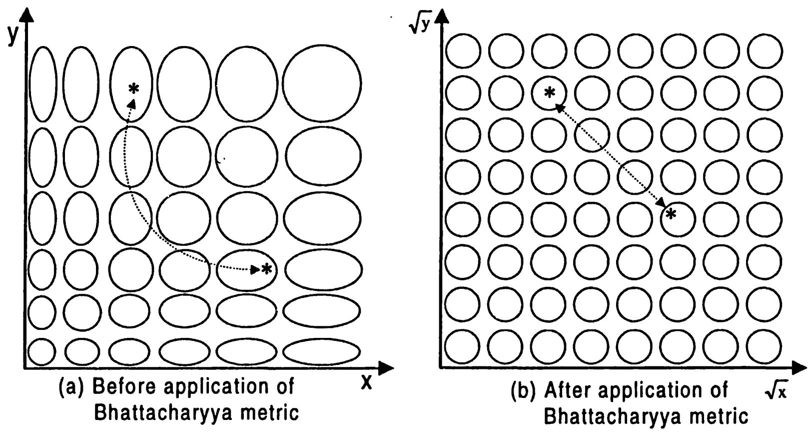

Fig. 1. Illustration of shortest distance between 2 observations having 2D Poisson errors.

$$\chi^2 = \sum_i \frac{(R_i - S_i)^2}{R_i + S_i}. \tag{6.1}$$

7. BHATTACHARYYA APPROXIMATION TO THE χ^2

We now show how the Bhattacharyya metric can approximate the chi-squared for small distances. The purpose of this section is to highlight how the Bhattacharyya statistic can be considered as an approximation to the chi-squared with the advantage that the Bhattacharyya metric has no problems comparing zero count-data (which can lead to singularities in the chi-squared statistic). For an arbitrary function, $f(\cdot)$ acting on binned, frequency coded data of unknown distribution we can approximate the chi-squared measure as (7.1):

$$\chi'^2 = \sum_i \frac{(f(R_i) - f(S_i))^2}{(\frac{\partial f}{\partial R_i})^2 R_i + (\frac{\partial f}{\partial S_i})^2 S_i}. \tag{7.1}$$

The denominator of the right-hand side of (7.1) is produced by using a first-order Taylor approximation. Now substituting $f = \sqrt{(\cdot)}$ into (7.1) gives (7.2):

$$\sum_i \frac{(\sqrt{R_i} - \sqrt{S_i})^2}{(\frac{1}{2\sqrt{R_i}})^2 R_i + (\frac{1}{2\sqrt{S_i}})^2 S_i} = 2 \sum_i \left(\sqrt{R_i} - \sqrt{S_i}\right)^2 \tag{7.2}$$

which is a scaled Matusita distance measure and rearranging this gives (7.3):

$$2\sum_i \left(\sqrt{R_i} - \sqrt{S_i}\right)^2 = \text{constant} - 4\sum_i \sqrt{R_i}\sqrt{S_i}. \tag{7.3}$$

Thus the Bhattacharyya measure approximates the chi-squared statistic. It should also be noted that by transforming all variances to be constant the Bhattacharyya measure avoids the singularity problem of the chi-squared statistic when comparing empty histogram bins.

8. THE DIMENSIONLESS EXAMPLE

In this section we highlight an example where the Bhattacharyya statistic is dimensionless. Consider two univariate Gaussian probability density functions where $\mu_A \neq \mu_B$ and $\sigma_A \neq \sigma_B$:

$$f_A(x) = \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp\left\{\frac{-(x-\mu_A)^2}{2\sigma_A^2}\right\} \tag{8.1}$$

$$f_B(x) = \frac{1}{\sqrt{2\pi\sigma_B^2}} \exp\left\{\frac{-(x-\mu_B)^2}{2\sigma_B^2}\right\}. \tag{8.2}$$

Consider the integral below as a general similarity measure (8.3):

$$\int_{-\infty}^{\infty} (f_A(x))^n (f_B(x))^n \, dx. \tag{8.3}$$

This has a solution given by:

$$\frac{\sqrt{2\pi\sigma_A^2\sigma_B^2/n}}{(2\pi\sigma_A\sigma_B)^n\sqrt{(\sigma_A^2+\sigma_B^2)}} \exp\left\{\frac{-n(\mu_A-\mu_B)^2}{2(\sigma_A^2+\sigma_B^2)}\right\} \tag{8.4}$$

and therefore the case of $n = 0.5$ gives a solution:

$$\frac{\sqrt{2\sigma_A\sigma_B}}{\sqrt{\sigma_A^2+\sigma_B^2}} \exp\left\{\frac{-(\mu_A-\mu_B)^2}{4(\sigma_A^2+\sigma_B^2)}\right\} \tag{8.5}$$

and $n = 1$ gives the solution:

$$\frac{1}{\sqrt{2\pi(\sigma_A^2+\sigma_B^2)}} \exp\left\{\frac{-(\mu_A-\mu_B)^2}{2(\sigma_A^2+\sigma_B^2)}\right\}. \tag{8.6}$$

From the negative natural logarithm of (8.6) we can consider (8.6) to be a scaled chi-squared type statistic. The constant term in (8.6) is not dimensionless, thus the statistic will depend upon the measurement scale used. In contrast the constant term of the Bhattacharyya measure in (8.5) is dimensionless, thus the Bhattacharyya metric is independent of the measurement scale.

9. BHATTACHARYYA MEASURE APPLIED TO HISTOGRAMS

The Bhattacharyya measure can be used to compare the similarity between two histograms as follows: If we let R_i be the frequency coded quantity in bin i for the first histogram and S_i a similar quantity for the second histogram. We propose the Bhattacharyya statistic $\sum_i \sqrt{R_i}\sqrt{S_i}$ as a measure of similarity between the two histograms. For the case of two identical histograms we obtain $\sum_i R_i = 1$ indicating a perfect match. The successful application of the Bhattacharyya statistic for histogram matching can be found in numerous applications, ([1, 6]).

10. SUMMARY

In this work we have presented the original geometric interpretation of the Bhattacharyya similarity measure and explained the use of the metric in the Bhattacharyya bound. We emphasize that the use of the statistic should not be limited to the upper bound on misclassification in a two-class problem and recommend its use for comparing frequency-coded data. Several examples have been described where the Bhattacharyya statistic has more desirable properties than the chi-squared metric, such as it being dimensionless. We have described the importance of the square-root variance stabilizing transformation with respect to the Bhattacharyya metric and explained how the statistic can be used to compare histogram data.

(Received December 18, 1997.)

REFERENCES

[1] F. J. Aherne, N. A. Thacker and P. I. Rockett: Optimal pairwise geometric histograms. In: Proc. 8th British Machine Vision Conf., Colchester 1997, pp. 480–490.

[2] A. Bhattacharyya: On a measure of divergence between two statistical populations defined by their probability distributions. Bull. Calcutta Math. Soc. *35* (1943), 99–110.

[3] R. Christensen: Linear Models for Multivariate Time Series and Spatial Data. Springer–Verlag, New York 1991.

[4] K. Fukanaga: Introduction to Statistical Pattern Recognition. Second edition. Academic Press, New York 1990.

[5] K. Matusita: Decision rules based on distance for problems of fit, two samples and estimation. Ann. Math. Statist. *26* (1955), 631-641.

[6] N. A. Thacker, I. Abraham, and P. G. Courtney: Supervised learning extensions to the CLAM network. Neural Network J. *10* (1997), 2, 315–326.

Dr. Frank J. Aherne, R&D Pace Micro Technology Plc, Saltaire, W. Yorkshire BD18 3LF. United Kingdom.
e-mail: frank.aherne@pace.co.uk
Dr. Neil A. Thacker, Dept. Medical Biophysics, University of Manchester. United Kingdom.
e-mail: nat@server1.smb.man.ac.uk.
Dr. Peter I. Rockett, University of Sheffield, Department EEE, Mappin Street, S1 3JD. United Kingdom.
e-mail: p.rockett@sheffield.ac.uk