The Bhattacharyya Measure

Konstantinos G. Derpanis
kosta@cs.yorku.ca

Version 1.0

March 20, 2008

1 Introduction

An important problem in computer vision is measuring the dissimilarity between distributions of features, such as colour and texture (cf. (Rubner, Puzicha, Tomasi & Buhmann, 2001)). The focus of this note is on the *Bhattacharyya measure* and its derivatives. For a discussion of the statistical foundations of the Bhattacharyya measure, the reader is referred to (Aherne, Thacker & Rockett, 1997).

2 Bhattacharyya measure

Let $p(i)$ and $p'(i)$ represent two multinomial populations, each consisting of N classes with respective probabilities $p(i=1),\ldots,p(i=N)$ and $p'(i=1),\ldots,p'(i=N)$. Since $p(i)$ and $p'(i)$ represent probability distributions, $\sum_{i=1}^{N} p(i) = \sum_{i=1}^{N} p'(i) = 1$. The Bhattacharyya measure (Bhattacharyya, 1943) (or coefficient) is a divergence-type measure between distributions, defined as,

$$\rho(p,p') = \sum_{i=1}^{N} \sqrt{p(i)p'(i)}. \tag{1}$$

The Bhattacharyya measure has a simple geometric interpretation as the cosine of the angle between the N-dimensional vectors $(\sqrt{p(1)},\ldots,\sqrt{p(N)})^\top$ and $(\sqrt{p'(1)},\ldots,\sqrt{p'(N)})^\top$. Thus, if the two distributions are identical, we have:

$$\cos(\theta) = \sum_{i=1}^{N} \sqrt{p(i)p'(i)} = \sum_{i=1}^{N} \sqrt{p(i)p(i)} = \sum_{i=1}^{N} p(i) = 1, \tag{2}$$

and consequently $\theta = 0$. Furthermore, based on *Jensen's inequality* (Cover & Thomas, 1991), we have,

$$0 \leq \rho(p,p') = \sum_{i=1}^{N} \sqrt{p(i)p'(i)} = \sum_{i=1}^{N} p(i)\sqrt{\frac{p'(i)}{p(i)}} \leq \sqrt{\sum_{i=1}^{N} p'(i)} = 1. \tag{3}$$

A potentially undesirable property of the coefficient is that it does not impose a *metric* structure since it violates at least one of the distance metric axioms (Fukunaga, 1990). In (Comaniciu, Ramesh & Meer, 2003), the authors propose the following modification of the Bhattacharyya coefficient that does indeed represent a metric distance between distributions:

$$d(p,p') = \sqrt{1-\rho(p,p')}, \tag{4}$$

where $\rho(\cdot,\cdot)$ denotes the Bhattacharyya coefficient (1). For the proof that this distance is indeed a metric (i.e., obeys all of the metric axioms), see Appendix in (Comaniciu, Ramesh & Meer, 2003).

Next, let us consider a related measure, the *Hellinger discrimination* (Hellinger, 1909) (also known as the *Matusita* measure (Matusita, 1955)). This measure defines the distance between two probability distributions, as:

$$\sum_{i=1}^{N}\left(\sqrt{p(i)}-\sqrt{p'(i)}\right)^2. \tag{5}$$

This measure is related to the Bhattacharyya coefficient (1), $\rho(\cdot,\cdot)$, and distance (4), $d(\cdot,\cdot)$, by:

$$\begin{aligned}
&\sum_{i=1}^{N}\left(\sqrt{p(i)}-\sqrt{p'(i)}\right)^2 &(6)\\
&=\sum_{i=1}^{N}\sqrt{p(i)}\sqrt{p(i)}-2\sum_{i=1}^{N}\sqrt{p(i)}\sqrt{p'(i)}+\sum_{i=1}^{N}\sqrt{p'(i)}\sqrt{p'(i)} &(7)\\
&=\sum_{i=1}^{N}p(i)-2\sum_{i=1}^{N}\sqrt{p(i)}\sqrt{p'(i)}+\sum_{i=1}^{N}p'(i) &(8)\\
&=2-2\sum_{i=1}^{N}\sqrt{p(i)}\sqrt{p'(i)} &(9)\\
&=2-2\rho(p,p') &(10)\\
&=2(1-\rho(p,p')) &(11)\\
&=2d(p,p')^2. &(12)
\end{aligned}$$

Finally, let us turn our attention to the relationship between the the Bhattacharyya coefficient (1) and the *chi-square* (χ^2) measure. The chi-square measure is used to provide a measure of similarity between two distributions (cf. (Leung & Malik, 2001)):

$$\chi^2(p,p')=\frac{1}{2}\sum_{i=1}^{N}\frac{(p(i)-p'(i))^2}{p(i)+p'(i)}. \tag{13}$$

In (Aherne, Thacker & Rockett, 1997) it is shown that the Bhattacharyya coefficient (1) approximates the χ^2-measure (13), while avoiding the singularity problem that occurs when comparing instances of the distributions that are both zero.

References

Aherne, F., Thacker, N. & Rockett, P. (1997). The Bhattacharyya metric as an absolute similarity measure for frequency coded data. *Kybernetika*, *32(4)*, 1–7.

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distribution. *Bulletin of the Calcutta Mathematical Society*, *35*, 99–110.

Comaniciu, D., Ramesh, V. & Meer, P. (2003). Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25(5)*, 564–577.

Cover, T. & Thomas, J, A. (1991). *Elements of Information Theory*. Wiley-Interscience.

Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition (Second Edition)*. Academic Press.

Hellinger, E. (1909). Neaue begr undung der theorie der quadratischen formen von unendlichen vielen ver anderlichen. *136*, 210–271.

Leung, T. & Malik, J. (2001). Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, *43(1)*, 29–44.

Matusita, K. (1955). Decision rules based on distance for problems of fit, two samples and estimation. *Annals of Mathematical Statistics*, *26*, 631–641.

Rubner, Y., Puzicha, J., Tomasi, C. & Buhmann, J. (2001). Empirical evaluation of dissimilarity measures for color and texture. *Computer Vision and Image Understanding*, *84(1)*, 25–43.