The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data.

F.J. Aherne; N.A. Thacker; Peter Rockett

Abstract:

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.