A New Dissimilarity Measure between Feature-Vectors

Abstract

Distance measures is very important in some clustering and machine learning techniques. At present there are many such measures for determining the dissimilarity between the feature-vectors, but it is very important to make a choice that depends on the problem to be solved. This paper proposes a simple but robust distance measure called Reference Distance Weighted, for calculating distance between feature-vectors with real values. The basic attribute that distinguishes it from other measures is that the distance is measured from one of the feature-vector, considered as a reference system, to other feature-vectors. In fact this reference vector belongs to a class of a classification system. A second distinctive attribute is that its value does not depend on the orders of magnitude of the different characteristics of vectors. In addition, through a parameter called factor of relevance, each feature receives a weight in terms of its influence, because different features have different influence on dissimilarity estimation depending on the final problem to be solved. An extension of the proposed distance allows working with hybrid vectors, ie real and logical values. Future research directions are also provided.

References

- R. Hamming, "Error detecting and error correcting codes," The BellSystem
- A. Bhattacharya, P. Kar, M. Pal, "On Low Distortion Embeddings of Statistical Distance Measures into Low Dimensional Spaces". Springer, LNCS, Vol. 5690, pp 164-172, 2009

**Index Terms**

Computer Science  
Artificial Intelligence
Keywords
classification distance dissimilarity features