

MORE POWERFUL DISCRIMINANTS FOR CLASSIFYING PHYLOGENETIC SIGNALS IN DINUCLEOTIDE FREQUENCIES

Robert H. Baran and Hanseok Ko

Department of Electronics and Computer Engineering, Korea University, Seoul, Korea
hsko@korea.ac.kr

ABSTRACT

Microbial DNA fragments are classified according to species using compositional features and "genomic signatures" the oldest of which is the dinucleotide relative abundance profile defined by Karlin et al. More informative features, including higher order signatures, have demonstrated greater species-specificity in comparison to the baseline established by the dinucleotide signature using "delta-distance" to assess dissimilarity; but lack of standard methods has precluded rigorous comparison. We describe a new method for classifier evaluation that reduces any number of pairwise inter-genomic comparisons to a single performance measure. To illustrate the method, we compare delta-distance to quadratic and linear discriminants prescribed by elementary pattern recognition theory, and find that the quadratic form is significantly more powerful.