New entropy-based method for gene selection

ABSTRACT

Dimension reduction and selection of a small number of genes with high ability, to discriminate objects, are important challenges in micro-array data analysis. Gene selection, based on top ranked genes which individually have high power to discriminate objects, is a traditional method that doesn't consider the redundancy among the genes. Some results present that subset of genes with low degree of redundancy can show a more comprehensive representation of the targeted classes than one with redundant genes. In this paper, we use Shannon theorem and penalized logistic regression (PLR) as a probability estimator to present a new algorithm for dimension reduction and collect a subset of representative genes of gene expression profile. Breast cancer, leukemia, colon and lung datasets have been classified based on proposed gene selection algorithm by PLR classifier. In most cases the results show a good performance compared to other recent researches.