Abstract

Microarray data has been widely applied to cancer classification, where the purpose is to classify and predict the category of a sample by its gene expression profile. DNA microarray is a gene chip which consists of expression levels for a huge number of genes on a relatively small number of samples. However, only a small number of genes contribute in accurate classification of cancer. Therefore, the challenging task is to identify a small subset of informative genes which has maximum amount of information about the class. Moreover, it also minimizes the classification errors. In this paper, we propose a hybrid negative correlated method, which combines the features from various correlation based feature selection techniques, for the generation of mutually exclusive informative feature sets. We test the effectiveness of the proposed approach using a neural network based classifier on two benchmark gene expression data sets - colon dataset and leukemia dataset. The obtained results are encouraging as hybrid negative correlated method based features give better recognition accuracy than positive correlated and other negative correlated features.
Hybrid Correlation based Gene Selection for Accurate Cancer Classification of Gene Expression Data


- http://microarray.princeton.edu/oncology/affydata
- http://www. genome.gov

Index Terms
Computer Science
Artificial Intelligence

Keywords
Dna Microarray  Classification  Correlation  Neural Network  Backpropagation Algorithm