

## **A Heuristic Search for Discovering Classifier Gene-Sets**

Dataset used: cancer data

Authors: Goutham Kurra, Wen Niu, Raj Bhatnagar  
University of Cincinnati,  
Cincinnati, OH  
Raj.Bhatnagar@uc.edu

The objective of the algorithms proposed in this study is to use a combination of statistical class scatter metrics and heuristic search algorithms to determine all those minimal combinations of genes that can discriminate sufficiently among samples from known classes. The knowledge of all those subsets that have similar performance as class predictors can shed useful light on the relative importance and interactions of various individual genes, as far as their prediction capability is concerned.

The main problem in discovering all the good discriminant sets is computational in nature. In a typical gene-expression test one may identify a couple of hundred genes that appear to be good candidates for class discriminants. To uncover the role of all inter-gene interactions one must examine all possible subsets of the few hundred individual genes. This results in a computationally intractable set of candidates that must be examined by an algorithm.

We present a heuristic search based algorithm that prunes off the unpromising candidates and outputs a number of good candidates that can act as discriminants between the classes. These discriminants, however, may consist of some overlapping and some distinct genes. A relative study of the composition of various equally well-performing discriminants can provide us additional insights into the roles of individual genes.