Abstract

In the field of bioinformatics, selection of genes in multiclass sample classification can be done by filtering methods using microarray data. Such approaches usually contribute to bias towards a few classes that are easily recognizable from other classes due to imbalances of strong features and sample sizes of distinct classes in a microarray data. Many methods have been used for the filter methods, as they are very commonly used in gene ranking from microarray data in multiclass problems. In this research, we discuss various methods to decompose multiclass ranking statistics into class specific statistics and then need of Pareto-front analysis for selection of genes. This mitigates the bias induced by class intrinsic characteristics of dominating classes. The need of Pareto-front analysis is to indicate on two filter criteria commonly used for gene selection: F-score and KW-score. A significant development in classification performance and reduction in redundancy among top ranked genes were achieved in experiments with both synthetic and real-benchmark data sets. The following work is analysis over the traditional and improved filter methods used for gene selection of various classes through various mechanisms available in the literature.

References
Index Terms

Computer Science

Applied Sciences

Keywords

Aggregation statistics  filter methods  gene selection  multiobjective evolutionary optimization  Pareto-front analysis.