

Robbins, K. R.; Zhang, W.; Bertrand, J. K.; Rekaya, R.

The ant colony algorithm for feature selection in high-dimension gene expression data for disease classification. (English) [Zbl 1146.92318](#)

Math. Med. Biol. 24, No. 4, 413-426 (2007).

Summary: The use of gene expression data to diagnose complex diseases represents an exciting area of medicine; however, such data sets are often noisy, requiring the selection of feature subsets to obtain maximum classification accuracy. Due to the high dimensions of many expression data sets, filter-based methods are commonly used, but often yield inconsistent results. Optimization algorithms can outperform filter methods, but often require preselection of features to achieve good results. To address the problems of many commonly used feature selection methods, the ant colony algorithm (ACA) is proposed for use on data sets with large numbers of features. The ACA is an optimization algorithm capable of incorporating prior information, allowing it to search the sample space more efficiently than other optimization methods. When applied to several high-dimensional data sets, the ACA was able to identify small subsets of highly predictive and biologically relevant genes without the need for extensive preselection of features. Using the selected genes to train a latent variable model yielded substantial increases in prediction accuracy when compared to several rank-based methods and results obtained in previous studies. The superiority of the ACA algorithm was validated through simulations.

MSC:

[92C50](#) Medical applications (general)

[90C90](#) Applications of mathematical programming

[92C40](#) Biochemistry, molecular biology

[92-08](#) Computational methods for problems pertaining to biology

[62P10](#) Applications of statistics to biology and medical sciences; meta analysis

Cited in **1** Document

Keywords:

disease diagnostics; feature selection

Software:

AS 136

Full Text: [DOI](#)