Abstract

In recent times in the classification and diagnosis of cancer nodules, gene expression profiling by micro array techniques are playing a fundamental role. A range of researchers have proposed a number of machine learning and data-mining approaches for identifying cancerous nodule using gene expression data. The process of gene selection for the cancer classification encounters with some major problems due to the properties of the data such as the small number of samples compared to the huge number of genes, irrelevant genes, and noisy data. Hence, this paper aims at selecting a near-optimal subset of informative genes that is most relevant for the cancer classification. This paper also proposes an efficient BFSS (Boost Feature Subset Selection) technique to improve the performance of single-gene based discriminative scores using bootstrapping techniques. The proposed hybrid approach
(Filter-Wrapper) will be implemented on three publicly available microarray datasets. These microarray datasets are: Acute Lymphoblastic Leukemia Cancer (ALL), Lung Cancer and Colon Cancer.

References

- Correlation-based Feature Selection for Machine Learning, By Mark A. Hall, Department of Computer Science, The University of Waikato, Hamilton, New Zealand.
- Studies on Intelligent Approaches to Select Informative Genes from Gene Expression Data for Cancer Classification, Mohd Saberi Bin Mohamad, ??????, 2009, ???, 2009.
- Xian Xu and Aidong Zhang (2010), Boost Feature Subset Selection: A New Gene Selection Algorithm for Microarray Dataset, State University of New York at Buffalo, Buffalo, NY 14260, USA.

Index Terms

Computer Science Bio-medical Sciences
Keywords
   Evolutionary Algorithms  Binary Coded Genetic Algorithm