Abstract

Modern microarray chips can hold gene information from thousands of genes and hundreds of individuals and the main challenge of an effective feature selection method is to identify most useful genes from the whole dataset. Removal of less informative genes helps to alleviate the effects of noise and redundancy and simplifies the task of disease classification and prediction of medical conditions such as cancer. Genetic Algorithm (GA) based wrapper model performs well but suffers from over-fitting problem and the initial population is large and random. Traditional approaches use a filter based preprocessing step to reduce the dimension of the data on which GA operates and as filtering methods on its own has shown to introduce
BFSSGA: Enhancing the Performance of Genetic Algorithm using Boosted Filtering Approach

redundant features, in this paper Boosted Feature Subset Selection (BFSS) which is a boosted t-score filter method, is used as a preprocessing step. The gene subset provided by BFSS is fed to a Genetic Algorithm which reduces the feature subset in smaller numbers and helps to generate a better optimal subset of genes. The proposed hybrid approach is applied on leukemia, colon and lung cancer benchmarked datasets and have shown better results than other well-known approaches.

References


BFSSGA: Enhancing the Performance of Genetic Algorithm using Boosted Filtering Approach


Index Terms

Computer Science

Pattern Recognition

Keywords