An Analysis on Clustering Based Gene Selection and Classification for Gene Expression Data

Authors : K. Sathishkumar,V. Thiagarasu

Abstract : Due to recent advances in DNA microarray technology, it is now feasible to obtain gene expression profiles of tissue samples at relatively low costs. Many scientists around the world use the advantage of this gene profiling to characterize complex biological circumstances and diseases. Microarray techniques that are used in genome-wide gene expression and genome mutation analysis help scientists and physicians in understanding of the pathophysiological mechanisms, in diagnoses and prognoses, and choosing treatment plans. DNA microarray technology has now made it possible to simultaneously monitor the expression levels of thousands of genes during important biological processes and across collections of related samples. Elucidating the patterns hidden in gene expression data offers a tremendous opportunity for an enhanced understanding of functional genomics. However, the large number of genes and the complexity of biological networks greatly increase the challenges of comprehending and interpreting the resulting mass of data, which often consists of millions of measurements. A first step toward addressing this challenge is the use of clustering techniques, which is essential in the data mining process to reveal natural structures and identify interesting patterns in the underlying data. This work presents an analysis of several clustering algorithms proposed to deals with the gene expression data effectively. The existing clustering algorithms like Support Vector Machine (SVM), K-means algorithm and evolutionary algorithm etc. are analyzed thoroughly to identify the advantages and limitations. The performance evaluation of the existing algorithms is carried out to determine the best approach. In order to improve the classification performance of the best approach in terms of Accuracy, Convergence Behavior and processing time, a hybrid clustering based optimization approach has been proposed.

Keywords : microarray technology, gene expression data, clustering, gene Selection
Conference Title : ICCSE 2015 : International Conference on Computer Science and Engineering
Conference Location : Kuala Lumpur, Malaysia
Conference Dates : August 24-25, 2015