Abstract

A microarray measures the expression levels of thousands of genes at the same time. Clustering helps to analyze microarray gene expression data. The characteristic of gene expression data is its coherent structure with regards to genes and samples. In this paper we have implemented a biclustering algorithm to identify subgroups of data which shows correlated behavior under specific experimental conditions. In the process of finding biclusters, Fuzzy C-means clustering is used to cluster the genes and samples with maximum membership function. Dimensionality and reducing the gene shaving is done using principal component analysis & gene filtering with the function respectively. This method obtains highly correlated sub matrices of the gene expression dataset. It is also observed that it identifies important co-regulated genes and samples at the same time. Principal component analysis is also verified the concatenation of small biclusters into bigger one. Biclustering is a NP-hard problem [10] therefore we have implemented biclustering in multi-core parallel environment to reduce the computational time of the algorithm. Data level and task level parallelism is used to develop the algorithm on MATLAB Parallel computing toolbox with multicore platform. We have compared
the results with other parallel & sequential algorithm to show the effectiveness of the algorithm.

References

18. Dwitiya Tyagi, Sujoy Das, and Namita Srivastava, Parallel Two-way Clustering for


Index Terms

Computer Science Fuzzy Systems

Keywords

Microarray, gene expression, Multicore platform, Biclustering, MATLAB parallel computing, PCA, gene entropy.