Abstract

In the past decade there have been advance in technologies, the amount of biological data such as DNA sequences and microarray data have been increased tremendously. To obtain knowledge from the data, explore relationships between genes, understanding severe diseases and development of drugs for patterns from the databases of large size and high dimensionality. Information retrieval and data mining are powerful tools to extract information from the databases and/or information repositories. The integrative cluster analysis of both clinical and gene expression data has shown to be an effective alternative to overcome the abovementioned problems. In this paper, we focus on how to improve the searching and the clustering performance in genomic data from commonly used clustering techniques. In the proposed gene clustering technique, firstly, the high dimensionality of the microarray gene data is reduced using LPP. The LPP is chosen for the dimensionality reduction because of its ability of preserving locality of neighborhood relationship. Secondly, through performance experiments on real data sets, the proposed method fuzzy C-means is shown to achieve higher efficiency, clustering quality and automation than other clustering method.

References
Efficient Clustering for Gene Expression Data


- P. Valarmathie, Dr MV Srinath, Dr T. Ravichandran. "Hybrid Fuzzy C-Means Clustering Technique for Gene Expression Data," International Journal of Research and Reviews in Apld Sci, Vol 1, No 1, pp. 33-37, October 09

- Jian Wen, "Ontology Based Clustering for Improving Genomic IR," Twentieth IEEE International Symposium on Computer-Based Medical Systems, pp. 225 – 230, June 07


Efficient Clustering for Gene Expression Data

- Microarray gene samples of human acute leukemia and colon cancer data
http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi

Index Terms

Computer Science

Information Sciences

Keywords

Clustering  Microarray  Locality Preserving Projection (lpp)  Fuzzy C-means (fcm)
K-means