

Analysis of Expression Data Using Unsupervised Techniques

Authors : M. A. I Perera, C. R. Wijesinghe, A. R. Weerasinghe

Abstract : his study was conducted to review and identify the unsupervised techniques that can be employed to analyze gene expression data in order to identify better subtypes of tumors. Identifying subtypes of cancer help in improving the efficacy and reducing the toxicity of the treatments by identifying clues to find target therapeutics. Process of gene expression data analysis described under three steps as preprocessing, clustering, and cluster validation. Feature selection is important since the genomic data are high dimensional with a large number of features compared to samples. Hierarchical clustering and K Means are often used in the analysis of gene expression data. There are several cluster validation techniques used in validating the clusters. Heatmaps are an effective external validation method that allows comparing the identified classes with clinical variables and visual analysis of the classes.

Keywords : cancer subtypes, gene expression data analysis, clustering, cluster validation

Conference Title : ICBM 2020 : International Conference on Bioinformatics and Metagenomics

Conference Location : Venice, Italy

Conference Dates : November 12-13, 2020