

Abstract

Clustering gene expression data is an important task in bioinformatics research and biomedical application. Many clustering methods can be found in the literature. In cluster analysis, one wishes to partition the given data set into groups based on the influencing features such that the data objects in the same group are more similar to each other than the data objects in other groups. Clustering techniques have proven to be useful to understand gene function, gene regulation, cellular processes and sub types of cells. Microarray measures gene expression. We will refer to both the cDNA micro array and the oligo chip as micro array technology and term the measurements collected via both methods as gene expression data. Genes with similar expression patterns (co-expression genes) can be clustered together with similar cellular function. Clustering is more subjective work than classification and regression. Though classification and regression have many general validation measures, clustering have few validation measures. Also, it is difficult to develop general measure for cluster validation. So, many evaluation measures have been done for cluster validation. In this project we did an ensemble method of validation for cluster analysis. To verify our improved performance, we make experiments by some objective data sets.

Keywords: *Cluster Analysis, Cluster Validation, Internal measures, Stability measures, Ensemble method, Gene Expression Pattern.*