Abstract

Data mining techniques such as clustering, in gene expression data analysis often land up with results which have a large number of highly related groups depending on various correlation measures. For biologists, to extract biological information from such a huge possible option is cumbersome. Information from gene ontology repository can be of significant importance to filter such results into a smaller number of relevant candidates which can be easily and readily used by biologists to derive unknown biological facts.

The main objective of this project is divided into two tasks. The first task under this objective is to develop an effective unsupervised measure that divides expression of genes into a set of groups of meaningful clusters which have some biological significance. The second task is to work out a framework for integrating GO knowledge to improve the biological significance of the results and finally to validate the results in light of some benchmark datasets.