Abstract

Microarray technology demands the development of data mining algorithms for extracting useful and novel patterns. A bicluster of a gene expression dataset is a local pattern such that the genes in the bicluster exhibit similar expression patterns through a subset of conditions. In this study biclusters are detected in two steps. In the first step high quality bicluster seeds are generated using K-Means clustering algorithm. These seeds are then enlarged using a multistart metaheuristic method Greedy Randomized Adaptive Search Procedure (GRASP). In GRASP there are two phases, construction and local search. The Experimental results on the benchmark datasets demonstrate that GRASP is capable of identifying high quality biclusters compared to many of the already existing biclustering algorithms. Moreover far better biclusters are obtained in this algorithm compared to the already existing algorithms based on the same GRASP metaheuristics. In this study GRASP is applied for the first time to identify biclusters from Human Lymphoma dataset.

Reference

Application of Greedy Randomized Adaptive Search Procedure to the Biclustering of Gene Expression Data

- SGD GO Termfinder [http://db.yeastgenome.org/cgi bin/ GO/ goTermFinder]
Index Terms

Computer Science

Database

Management

Key words

Gene expression data

Greedy randomized adaptive search procedure

K-Means clustering

Biclustering

Data mining