As a useful data mining technique biclustering identifies local patterns from gene expression data. A bicluster of a gene expression dataset is a subset of genes which exhibit similar expression patterns along a subset of conditions. In this paper a new method is introduced based on greedy search algorithm combined with the evolutionary technique particle swarm optimization for the identification of biclusters. Greedy methods have the possibility of getting trapped in local minima. Metaheuristic methods like particle swarm optimization have features for escaping from local minima and can find global optimal solutions. In this algorithm biclusters are identified in three steps. In the first step small disjoint tightly coregulated submatrices are generated using K-Means clustering algorithm. Then greedy search algorithm is used to enlarge the seeds. Output of greedy search algorithm is used as initial population of binary PSO. The result obtained on Yeast dataset shows that this method can generate high quality biclusters.

Reference


Index Terms

Computer Science

Database

Management
Key words

Biclustering

gene expression data

greedy search

kmeans clustering

particle swarm optimization