Abstract

Clustering problem is one of the hottest research fields in microarrays data analysis. In Clustering, a set of observations are assigned into subsets (called clusters) such that observations in the same cluster are similar in some sense. One of the clustering approaches is based on the minimum spanning tree (MST). The MST-based clustering techniques consist of three main phases; MST construction, inconsistent edges identification and clusters identification. The CLUMP algorithm (Clustering through Minimum spanning tree in parallel) is one of the MST-based clustering algorithms, which have been enhanced in the iCLUMP algorithm was improved using the cover tree data structure. This paper presents another improvement called iCLUMP-2 to enhance the edge inconsistency measure employed by both CLUMP and iCLUMP. The performance of the implemented algorithm was tested on a 45 nodes cluster using cancer microarrays data sets. The results showed that the proposed algorithm outperformed both CLUMP and iCLUMP providing better speedup and efficiency. Furthermore the quality of cluster produced by the iCLUMP-2 algorithm is much better that those produced by both CUMP and iCLUMP.
Microarrays Data Analysis for Cancer Disease on a Cluster of Computers

- Olman, V. and Mao, F. and Wu, H. and Xu, Y. Parallel clustering algorithm for large

Index Terms

Computer Science

Data Mining

Keywords

Clustering microarrays cancer parallel algorithm minimum spanning tree