ABSTRACT

Microarray is one of the bioinformatics technologies that allows monitoring of gene expression levels in huge dimension size in parallel. Microarray is used to help researchers for diagnosing diseases or finding drugs to cure them. The microarray process involves grouping genes based on ratio which shows in all genes. In Data Mining, this process is called clustering.

Generally, microarray data is collected on huge proportion. However when gathering a large amount of data, it can be difficult to analyse. Therefore, clustering is neccesary to get clusters until it generates new information from the data.

The method that is used in this research is clique partition that is based on branch and bound procedure and DFS to browse all nodes of graph. Clique algorithm searches for every node recursively dan backtracking until all nodes have passed. In oerder to get clusters, the microarray data needs to be tranformed into a graph which is formed into adjacency matrix. It is a matrix of connectivity between points which only consists of two binary values, 1 if there is any correlation between nodes, 0 otherwise. The threshold value is used to determine correlation in this research. Finding clusters using clique partition means finding maximal clique. The results obtained show that threhold changes affect the number of clusters obtained. The results of cluster analysis using microarray data indicates that selection of smaller thresholds give a smaller margin of error SSE.

Keywords: *clique partition, clustering, microarray, PCA, threshold*