

ABSTRACT

Data mining refers to extracting and analyzing large amounts of data to get meaningful knowledge. The different between data mining and other information extraction are analyze large amounts of data, extract potensial knowledge, and do the complex task that is difficult to perform with manual way. Clustering as a functionality of data mining is a process to grouping data objects into clusters based on relation between data atribut.

This final project apply pCluster as an algorithm to find cluster in multidimensional data. The pCluster algorithm is the algorithm which perfoms clustering by pattern similarity in microarray data analysis. In DNA microarray analysis, the expression levels of two genes may rise and fall in response to a set of environmental conditions. Under the pCluster model, two object are similar if they exhibit a coherent pattern on a subset of dimensions. Although the magnitude of their expression levels may not be close, the patterns they exhibit can be very much alike.

The pCluster model, thought developed in the study of microarray data cluster analysis, can be applied to many other applications that require finding similar or coherent pattern involving a subset of numerical dimensions in large, high-dimensional data sets.

pCLuster algorithm have good capability for finding cluster with a set of dimension in high dimensional data. pCluster scalability for grow of data object and atributte is linear for respon time. For total of cluster that found and respon time that needed are return equal with grow of minimal rows or minimal coloms. and straight equal with grow of threshold.

Keywords : *data mining, multidimensional data, clustering, subspace clustering, frequent pattern, pCluster algorithm.*