Abstract

Clustering is a process of grouping data into a group or cluster, so that the objects in a cluster has a very large similarity with other objects in the same cluster, but not similar to objects in other clusters.

Many clustering algorithms are used for numerical data. One of them is hierarchical clustering algorithm that is build a hierarchy of clusters where similar objects will be placed on the same hierarchy and not similar objects that are far apart on the hierarchy. However, problems arise when the algorithm was applied to data that have boolean or categorical attribute. The data object that has a small similarity often grouped into one cluster altough the data has no common characteristic between its object.

To remedy the problem of categorical data, this final project uses ROCK (*RObust Clustering using linKs*) algorithm that perform clustering by grouping the data objects that have links at most between pairs of objects with number of clusters (k) and θ (threshold) value as parameter.

This final project shows that ROCK algorithm produces average of cohesion and separation and can handles outlier data better than hierarchical clustering algorithm.

Key word: clustering, cluster, categorical data, hierarchical clustering, robust clustering using links