Abstract

Data mining is a analize process to get important information wich is contained ing some unknow collected data. Clustering is a kind of data mining's fungsionality wich is used to do collectivity data into class or cluster. The principle of clustering is gruoping object into a class wich have big same other object in some class (similarity) , but its very difference with object in diffrent class (disimilarity). Almost all of clustering algorithm only can solve data with low dimensinality so that result of clustering have less quality. Aubspace clustering is a method to dicided cluster wich built in different subspace so that prefer handling multidimension data to common clustering. Algorithm that implemention subspace clustering such as: COSA, FINDIT, ENCLUS, PROCLUS, CLIQUE, MAFIA, ect.

This final project is implemented with MAFIA (Merging of adaptive Finite Intervals). MAFIA is clustering algorithm which use subspace clustering methodby integrity density based and adaptive grid based method so that this algorithm is good enough to handle multidimension data. MAFIA is a density based method due to clustering implementasion based on the density of data that use histogram for analyzing data range on each dimension. This algorithm use bottom-up search to find dense unit depend on the dimension to reduce searching space. Bottom-up use same princip with Apriori algorithm to procused assosoation rule. MAFIA is not need the input number of cluster fund in cluastering so that this algorithm is more fleksible on finding cluster from dataset.

Mafia algorithms tested using syntetic dataset. Datasets are generated online at www.datasetgenerator.com. Test results show that MAFIA is good for clustering data multidimensional. Quality is reflected in the maximum accuracy that is equal to that found 0.89.

Key word : *clustering, MAFIA algorithm, data mining, subspace clustering.*