Acomparative study of K-Means and KMIX Clustering methods

No No Saw Lwin

University of Computer Studies, Yangon
nonosawlwin88@gmail.com

Abstract

Clustering is the processof grouping the similar objects into the same group and produce characteristic for each cluster. Partition-based clustering algorithms are simple to implement. K-Means is one of the partition based clustering methods but it only work well on conditions include categorical data like medical fields. K-Means clustering need to transform categorical value to appropriate numeric form. KMIX clustering use simple matching dissimilarity measure for categorical value and replace mode instead of using mean for center vector. This system implements K-Means and KMIX algorithms on Zoo Small data set and Heart disease data set from UCI Machine Learning Repository. And then the clustering performance of these two algorithms are compared. According to experimental results, KMIX clustering is effective on data domain containing mixed type of numerical and categorical attributes.