Abstract

This paper presents a genetic k-means algorithm for clustering high dimensional objects in subspaces. High dimensional data faces data sparsity problem. In this algorithm, we present the genetic k-means clustering process to calculate a weight for each dimension in each cluster and use the weight values to identify the subsets of important dimensions that categorize different clusters. This is achieved by including the weight entropy in the objective function that is minimized in the k-means clustering process. Further, the use of genetic algorithm ensure for converge to the global optimum. The experiments on UCI data has reported that this algorithm can generate better clustering results than other subspace clustering algorithms.

Reference

**Index Terms**

Computer Science  
Data Mining

**Key words**

Genetic Algorithm  
Clustering  
Subspace clustering