Abstract

Clustering is a technique which is used to analyze the data in an efficient manner and generate required information. To cluster the dataset, there is a technique named k-mean, which is applied based on central point selection and calculation of Euclidian Distance. Here in k-mean, dataset will be loaded and from the dataset. Central points are selected using the formulae Euclidian distance and on the basis of Euclidian distance points are assigned to the clusters. The main disadvantage of k-mean is of accuracy, as in k-mean clustering user needs to define number of clusters. Because of user defined number of clusters, some points of the dataset are remained un-clustered. In this work, improvement in the k-mean clustering algorithm will be proposed which can define number of clusters automatically and assign required cluster to un-clustered points. The proposed improvement will lead to improvement in accuracy and reduce clustering time by the member assigned to the cluster to predict cancer.

Index Terms

Computer Science
Algorithms
Keywords

K-mean clustering, Prediction, clustering, Classification, Hierarchal clustering