Abstract

Clustering is a technique used to analyze 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, the dataset will be loaded and central points are selected using the formulae. Points are assigned to clusters on the basis of Euclidian distance. The main disadvantage of k-mean is of accuracy, as in k-mean clustering, the user needs to define the number of clusters. Because of the user-defined number of clusters, some points in the dataset are remained un-clustered. In this work, improvement in the k-mean clustering algorithm is proposed which can define the number of clusters automatically and assign the 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