Abstract

Clustering high dimensional data is challenging due to its dimensionality problem and it affects time complexity and accuracy of clustering methods. This paper presents the F-measure and Euclidean distance based performance efficiency and effectiveness of K-means and Agglomerative hierarchical clustering methods on Text and Microarray datasets by varying cluster values. Efficiency concerns about computational time required to build up dataset and effectiveness concerns about accuracy to cluster the data. Experimental results on different datasets demonstrate that K-means clustering algorithm is favourable in terms of effectiveness where as Agglomerative hierarchical clustering is efficient in time for text datasets used for empirical study.

References


Index Terms

Computer Science

Databases
Keywords

Clustering, K-means, Agglomerative hierarchical, F-measure, Precision, Recall.