SIMULATION OF ENHANCED KMEAN CLUSTERING MECHENISM ON MATLAB

Nancy, nancymalik339@gmail.com

ABSTRACT: K-means clustering is very Fast, robust & easily understandable. If data set is separated from one other data set, then it gives best results. Clusters do not having overlapping character & are also non-hierarchical within nature. Some challenges are related to visualization & querying of data. Scientist has faced several challenges in e-Science such as meteorology, complicated physics simulation & environmental researches. Lot of challenges has been faced due to big data in case of biology & genomics. Problems with existing system were search, sharing, storage, transfer, and visualization, querying-updating. These problems can be reduced by using proposed algorithm. In this paper we have explain clustering & proposed algorithm is discussed. We have simulated the enhanced K-Mean clustering using MATLAB.
Keywords: Clustering, K-Mean, Data Mining, MATLAB

1. INTRODUCTION

Clustering is[1] a process of data into a group of meaningful sub-classes is called clustering. Used either as a stand-alone tool to get insight into data distribution or as a pre processing step for other algorithms.
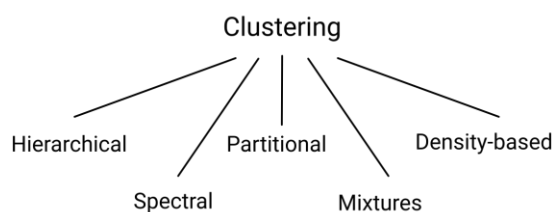

Fig 1 Clustering

By examining one or more attributes or classes, you may group individual pieces of data value together to form a structure opinion. At a simple stage, clustering is using one or more attributes as your basis for identifying a cluster of correlating results. Clustering is valuable to identify dissimilar info since this correlates with other examples so you may see where similarities & ranges agree[2]. Clustering may work both ways. You may assume that there is a cluster at a certain point & then use our credentials criteria to see if you are correct.

2. REQUIREMENTS OF CLUSTERING

The following points throw light on why clustering is required in data mining[3]

Scalability

We need highly scalable clustering algorithms to deal with large databases.

Capacity to contract within different type of attributes

The K-Mean Algorithms would be able to be implement on some kind of information like as interval based data binary data & categorical,.

Discovery of clusters with attribute shape:- clustering algorithm should be capable of detecting clusters of arbitrary shape[4]. We have not to be bounded to distance measures that care of to find round cluster of small sizes.