**ABSTRACT**

Significant research has been carried out in past to find efficient clustering technique to form clusters of data set consisting of both numerical and categorical attributes. Data mining is used to find the similarity and dissimilarity measures between the data objects which are represented by a collection of attributes both categorical and numerical. The data coming to the sensors comprises of data redundancy and distributed heterogeneously. So we need to have an efficient technique for forming clusters containing homogenous information together in clusters.

Different clustering techniques are used to solve this purpose like k-means clustering, k-mode clustering and k-prototype clustering.

K-means clustering is a conventional approach which forms cluster which consists of numeric data only like age, height. Clustering is done based on the Euclidian distance between the cluster centres and the data objects. Initially k clusters are formed by selecting k cluster centres and data objects are assigned to these clusters depending upon the minimum distance obtained. But the problem with k-means is that it works with numerical values only but the real data set consists of objects having both numerical and categorical values like to depict rainfall patterns, humidity and weather conditions.

So we need to have a technique which solves this purpose. K-mode algorithms and k-prototype algorithms are used to form clusters based on the numerical and categorical values.

The k-modes algorithm uses a simple matching dissimilarity measure to deal with categorical objects, replaces the means of clusters with modes, and uses a frequency-based method to update modes in the clustering process to minimize the clustering cost function. With the help of these extensions the k–modes algorithm enables the clustering of categorical data same as done in k-means algorithm. Further, the k-prototypes algorithm integrates the k-means and k -modes algorithms to allow for clustering objects described by mixed numeric and categorical attributes.

K-prototype uses distribution centroid to represent the prototype of categorical attributes in a cluster. Then it integrates the mean with distribution centroid to represent the prototype of a cluster with mixed attributes, and propose a new dissimilarity measure, which takes account of the significance of each attribute, to evaluate the dissimilarity between data objects and prototypes.