

Abstract

Clustering is a process for grouping object into a cluster such that object within a single cluster have similar characteristics, while object in different cluster are dissimilar. Clustering has been extensively implemented in many fields such as market research, pattern recognition, data analysis, image processing etc. The problem of clustering becomes more challenging when the data is categorical, that is when there is no inherent distance measure between data values. Moreover many clustering algorithm take a long time so it is not suitable for large amount of data. NabSqueezer algorithm is a clustering method for categorical data, in the first step it gives weight to each attribute value where for the uncommon attribute value are given a greater weight. NabSqueezer only need one scan of the data and decide an object will merge with the existing cluster depend on the threshold similarity value. The result shows that threshold which inputed by user influences system's accuracy based on purity measure, cohesion, and separation. Beside that it also show that NabSqueezer algorithm has good scalability with the increasing of dataset size.

Keywords : clustering, data categorical , nabsqueezer