Abstract

K-means is considered as one of the most common and powerful algorithms in data clustering, in this paper we're going to present new techniques to solve two problems in the K-means traditional clustering algorithm, the 1st problem is its sensitivity for outliers, in this part we are going to depend on a function that will help us to decide if this object is an outlier or not, if it was
an outlier it will be expelled from our calculations, that will help the K-means to make good results even if we added more outlier points; in the second part we are going to make K-means depend on Rock links in addition to its traditional distance, Rock links takes into account the number of common neighbors between two objects, that will make the K-means able to detect shapes that can't be detected by the traditional K-means.

Reference

- Sudipto Guha, Rajeev Rastogi, Kyuseok Shim, ROCK: A Robust Clustering Algorithm for Categorical Attributes.
- Wesam Barbakh, Similarity Graphs.
- Xiaochuan Wu and Colin Fyfe, On initializing prototypes for clustering.

Index Terms

Computer Science

Data Mining

Key words

Robust K-means

Rock links

Initializing K-means electing

centroids

Optimizing K-means distance measurement