

"Implementing & Improvisation of K-means Clustering Algorithm" article.

This paper about usage of Machine Learning K-means algorithm for clustering analysis problem. Authors discuss traditional K-means algorithm, advantages and disadvantages of it, areas of concern for improving k-means, propose enhanced k-means, and in the end of the paper they review proposed enhanced k-means by other authors.

First of all described clustering analysis problem and such traditional solution method as K-means algorithm. Clustering analysis - this is process of group partitioning a large data set according on some similarity function. So some data objects in one group, which calls "cluster", are similar. In paper, for clarity, data objects - point on two-dimensional space and function of similarity is distance between these points.

Traditional K-means algorithm including next steps:

1. Choosing k points from data set randomly as centroids of clusters.
2. For each point in data set computing distance to centroids and point assigning to those cluster, which centroid is nearest.
3. For each cluster computing centroid and if none centroid have changed algorithm ends, else computing distance for each point to new centroids.

This algorithm maybe faster clustering technique, but choosing initial centroid randomly sometimes have bad accuracy and can have infinite iterations of algorithm.

Authors proposed some improvement for this algorithm using something like binary tree for choosing centroids in first iterations. They tested their algorithm and traditional with proved input data sets, than compared them. Results are very impressive, because enhanced algorithm have better accuracy lesser time complexity. Test results presented as output program screenshots and tables with accuracy in percentage and running time in seconds.

Also in paper reviewed other good articles, which describes some methods to enhance traditional algorithm. Best time complexity of proposed algorithms by other authors was $O(N)$, but authors of this paper have doubts about the authenticity of the results in article.

So this paper lit up improvement methods of k-means and this is not the end of researching in this directions, because, however, improvement problem of k-means is not solved completely.