An Optimized Approach for k-means Clustering

Abstract

Cluster analysis method is one of the most analytical methods of data mining. The method will directly influence the result of clustering. This paper discusses the standard of k-mean clustering and analyzes the shortcomings of standard k-means such as k-means algorithm calculates distance of each data point from each cluster centre. Calculating this distance in each iteration makes the algorithm of low efficiency. This paper introduces an optimized algorithm which solves this problem. This is done by introducing a simple data structure to store some information in every iteration and used this information in next iteration. The introduced algorithm does not require calculating the distance of each data point from each cluster centre in each iteration due to which running time of algorithm is saved. Experimental
results show that the improved algorithm can efficiently improve the speed of clustering and accuracy by reducing the computational complexity of standard k-means algorithm.

References

- Kiri Wagstaff and Claire Cardie Department of computer science, Cornell University, USA "Constrained k-means algorithm with background knowledge."
- Ahmet Alken, Department of Electrical and Electronics Engineering, KSU, Turkey, "Use of K-means clustering in migraine detection by using EEG records under flash stimulation" International Journal of the Physical Sciences Vol. 6(4), pp. 641-650, 18 February, 2011

Index Terms

Computer Science Data Mining

Keywords
Cluster Analysis  K-means Clustering  Kd-tree  Lloyd's Algorithm  Standard K-means Algorithm

Constrained K-means Algorithm.