Abstract

Data Analysis plays an important role for understanding different events. Cluster Analysis is widely used data mining technique for knowledge discovery. Clustering has wide applications in the field of Artificial Intelligence, Pattern Matching, Image Segmentation, Compression, etc. Clustering is the process of finding the group of objects such that objects in one group will be similar to one another and different from the objects in the other group. k-Means clustering algorithm is one of the popular algorithm which has gained a lot of attraction because of its simplicity and ease of implementation. k-Means algorithm's efficiency is limited because of random selection of k initial centers. Therefore, we have surveyed different approaches for initial centers selection for k-Means algorithm. We have also shown comparative analysis of Original K-Means and Data Clustering with Modified k-Means Algorithm using MATLAB R2009b. We chose Euclidean distance as the similarity measure for our implementation and results are evaluated.

References

- Ran Vijay Singh and M. P. S Bhatia, “Data Clustering with Modified K-Means Algorithm”
Analysis of Initial Centers for k-Means Clustering Algorithm

- Data Mining Concepts and Techniques, Second edition Jiawei Han and Micheline Kamber.
- D. T Pham, S. S Dimov, C. D Nguyen, "Selection of k in k means clustering."

Index Terms

Computer Science
Algorithms
Keywords

k-Means  Clustering  Initial Centers  Similarity measures