Data Clustering Algorithms

Clustering Algorithm Applications

1) Clustering Algorithm in Identifying Cancerous Data

Clustering algorithm can be used in identifying the cancerous data set. Initially we take known samples of cancerous and non cancerous data set. Label both the samples data set. We then randomly mix both samples and apply different clustering algorithms into the mixed samples data set (this is known as learning phase of clustering algorithm) and accordingly check the result for how many data set we are getting the correct results (since this is known samples we already know the results beforehand) and hence we can calculate the percentage of correct results obtained. Now, for some arbitrary sample data set if we apply the same algorithm we can expect the result to be the same percentage correct as we got during the learning phase of the particular algorithm. On this basis we can search for the best suitable clustering algorithm for our data samples.

It has been found through experiment that cancerous data set gives best results with unsupervised non linear clustering algorithms and hence we can conclude the non linear nature of the cancerous data set.

References

1) A Comparison of Fuzzy and Non-Fuzzy clustering Techniques in Cancer Diagnosis by X.Y. Wang and J.M. Garibaldi.

2) Probability Density Estimation from Optimally Condensed Data Samples by Mark Girolami and Chao He.

2) Clustering Algorithm in Search Engines

ⓘ

Clustering algorithm is the backbone behind the search engines. Search engines try to group similar objects in

☰ Data Clustering Algorithms

used, better are the chances of getting the required result on the front page. Hence, the definition of **similar object** play a crucial role in getting the search results, better the definition of similar object better the result is.

Most of the brainstorming activities needs to be done for defining the criteria to be used for similar object.

References

1) Clustering Billions of Images with Large Scale Nearest Neighbor Search by Ting Liu, Charles Rosenberg and H.A. Rowley.

2) Probability Density Estimation from Optimally Condensed Data Samples by Mark Girolami and Chao He.

3) Clustering Algorithm in Academics

The ability to monitor the progress of students' academic performance has been the critical issue for the academic community of higher learning. Clustering algorithm can be used to monitor the students' academic performance. Based on the students' score they are grouped into different-different clusters (using k-means, fuzzy c-means etc), where each clusters denoting the different level of performance. By knowing the number of students' in each cluster we can know the average performance of a class as a whole.

References

1) Application of k-means clustering algorithm for prediction of students' academic performance by O.J. Oyelade, O.O. Oladipupo and I.C. Obagbuwa.

4) Clustering Algorithm in Wireless Sensor Network's based Application

Clustering Algorithm can be used effectively in Wireless Sensor Network's based application. One application where it can be used is in Landmine detection. Clustering algorithm plays the role of finding the Cluster heads(or cluster center) which collects all the data in its respective cluster.

References

1) Clustering of wireless sensor and actor networks based on sensor distribution and connectivity by Kemal Akkaya, Fatih Senel and Brian McLaughlan.

2) Wireless Sensor Network based Adaptive Landmine Detection Algorithm by Abhishek Saurabh and Azad Naik.

5) Clustering Algorithm in Drug Activity Prediction

Useful Dataset Links:

1) Social Network Dataset: http://socialcomputing.asu.edu/pages/datasets

ⓘ