Abstract

K-means clustering algorithm is one of the best known algorithms used in clustering; nevertheless it has many disadvantages as it may converge to a local optimum, depending on its random initialization of prototypes. We will propose an enhancement to the initialization process of k-means, which depends on using statistical information from the data set to initialize...
the prototypes. We show that our algorithm gives valid clusters, and that it decreases error and time.

**Reference**

Initializing K-Means Clustering Algorithm using Statistical Information

474-483.

Index Terms

Computer Science  Data Mining

Key words

Clustering  K-means Clustering  Initial
Prototypes Determination

Central Limit Theory

Normal Distribution

Maximum Likelihood Estimator