Abstract

Clustering techniques have more importance in data mining especially when the data size is very large. It is widely used in the fields including pattern recognition system, machine learning algorithms, analysis of images, information retrieval and bio-informatics. Different clustering algorithms are available such as Expectation Maximization (EM), Cobweb, FarthestFirst, OPTICS, SimpleKMeans etc. SimpleKMeans clustering is a simple clustering algorithm. It partitions n data tuples into k groups such that each entity in the cluster has nearest mean. This paper is about the implementation of the clustering techniques using WEKA interface. This paper includes a detailed analysis of various clustering techniques with the different standard online data sets. Analysis is based on the multiple dimensions which include time to build the model, number of attributes, number of iterations, number of clusters and error rate.
Analysis of SimpleKMeans with Multiple Dimensions using WEKA


- National Informatics Centre (NIC), Irrigation census data of water lifts in all villages of country, http://data.gov.in/.
- K-Means clustering Tutorial- By Kardi Teknomo, Ph. D.
- Privacy-Preserving K-Means clustering over vertically Partitioned Data-By Jaideep Vaidya and Chris Clifton, Dept. of Computer Sciences, Purdue University, 2050 N University St, West Lafayette, IN 47907-2066.

Index Terms

Computer Science Information Science

Keywords
Data mining  SimpleKMeans Clustering  WEKA.