Abstract

Clustering is a process of partitioning a set of data (or objects) into a set of meaningful sub-classes, called clusters. This paper analyzes three major clustering algorithms: K-Means, Hierarchical clustering and Density based clustering. The performance of these three clustering algorithms is compared using the clustering toolkit Weka.

References

- Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth "From Data Mining to Knowledge Discovery in Databases".
Pros and Cons of Clustering algorithms using Weka Tools

- MacQueen J. B, University of California-Los Angeles &quot;Some Methods for classification and Analysis of Multivariate Observations&quot;.
- Jiawei Han and Micheline Kamber, &quot;Data Mining: Concepts and Techniques&quot; Morgan Kaufmann Publishers, second Edition, (2006).
- Timothy C. Havens. &quot;Clustering in relational data and ontologies&quot; July 2010.

Index Terms

Computer Science Algorithms

Keywords

Data Mining Algorithms  Weka Tools  K-means Algorithms  Hierarchical Clustering And Density Based Clustering.