

# Classification of Clustering Techniques

K.Ranjith kumar<sup>1</sup> M.Praveena<sup>2</sup>

<sup>1</sup>MSc Student <sup>2</sup>Assistant Professor

<sup>1,2</sup>Department of Computer Science

<sup>1,2</sup>Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore, India-641 049

**Abstract**— Data mining extract the knowledge from large amount of data which store in multiple heterogeneous databases. The overall goal of the data mining process is to extract information from a large data set and transform it into an understandable form for further use. Clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar to each other than to those in other groups (clusters). Clustering is the one of data mining techniques in which data is divided into the groups of similar objects Clustering is a suitable example of unsupervised classification. The grid based methods use the single uniform grid mesh to partition the entire problem domain into cells.

**Key words:** Data Mining, Clustering, Types of Clustering, Classification Clustering Techniques

## I. INTRODUCTION

The purpose of the data mining technique is to mine information from a bulky data set and make over it into a reasonable form for supplementary purpose. It requires accessing and preparing data for a data mining algorithm, mining the data, analyzing results and taking appropriate action. The objective of data mining is designed for, and work best with large data sets. Cluster analysis is an important data mining technique which is used to find data segmentation and pattern information. Cluster Analysis, an automatic process to find similar objects from a database.

### A. Supervised Learning

In this training data includes both the input and the desired results. The correct results are known and are given in inputs to the model during the learning processes. The goal of the analysis is to specify a relationship between the dependent variable and explanatory variables the as it is done in regression analysis.

### B. Unsupervised Learning

The model is not provided with the correct results during the training. It can be used to cluster the input data in classes on the basis of their statistical properties only.

Pre-processing is necessary to analyze the multivariate data set.

- Anomaly detection: This is the identification of the unusual records or data errors.
- Association rule learning: It Searches the relationships between variables. This is sometimes referred to as market basket analysis.
- Clustering: This is the process of finding groups and structures in the data that are in some way or another "similar", without using known structures in the data.

In general, there are two types of attributes associated with input data in clustering algorithms, i.e., numerical attributes, and categorical attributes.

## II. GENERAL TYPES OF CLUSTERS

### A. Well-Separated Clusters:

A cluster is a set of points so that any point in a cluster is nearest (or more similar) to every other point in the cluster as compared to any other point that is not in the cluster.

### B. Center-Based Clusters:

A cluster is a set of objects such that an object in a cluster is nearest (more similar) to the "center" of a cluster, than to the center of any cluster other than it.

### C. Contiguous Clusters:

A cluster is a set of points so that a point in a cluster is nearest (or more similar) to one or more other points in the cluster as compared to any point that is not in the cluster.

### D. Density-based Clusters:

A cluster is a dense region of points, which is separated by according to the low-density regions, from other regions that is of high density.

### E. Shared Property (or) Conceptual Clusters:

Clusters that share some common property or represent a particular concept

## III. CLASSIFICATION OF CLUSTERING:

Clustering is the main task of Data Mining. And it is done by the number of algorithms. The most commonly used algorithms in Clustering are Hierarchical, Partitioning, Density Based and Grid based algorithms.

### A. Partitioning Clustering:

Data objects are divided into non overlapping clusters so that each and every object is in exactly in one subset.

### B. Hierarchical Algorithms:

Hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters. This hierarchy is created using two algorithms which are:

- Agglomerative
- Divisive.

#### 1) Agglomerative:

The method starts with as many clusters as there are records where each cluster contains just one record.

#### 2) Divisive:

The technique takes the opposite approach from agglomerative techniques. They start with all the records in one cluster.

### C. Density Based Clustering:

Density-based clustering algorithms try to find clusters based on density of data points in a region. Objects in these sparse areas - that are required to separate cluster. .

Representative algorithms are DBSCAN, GDBSCAN, OPTICS, and DBCLASD

*D. Grid Based Algorithms:*

Grid-based clustering where the data space is quantized into finite number of cells which form the grid structure and perform clustering on the grids.

#### IV. CONCLUSIONS

The overall goal of the data mining process is to separate the information from a large data set and transform it into an understandable form for further use. It first defines the data mining process which is the method of finding predictive information from a huge amount of databases. . Hierarchical-based clustering is the connectivity based clustering. Partitioning-based algorithm is the centroid based clustering. . Partitioning-based algorithm is the centroid based clustering. Density based clusters are defined as area of higher density then the remaining of the data set.

#### REFERENCES

- [1] Amandeep Kaur Mann & Navneet Kaur, Review Paper on Clustering Techniques, Volume 13 Issue 5 Version 1.0 Year 2013
- [2] Pradeep Rai, "A Survey of Clustering Techniques", International Journal of Computer Applications Volume 7– No.12, October 2010.
- [3] Anoop Kumar Jain, Prof. Satyam Maheswari "Survey of Recent Clustering Techniques in Data Mining", International Journal of Computer Science and Management Research, pp.72-78, 2012.
- [4] M.Vijayalakshmi, M.Renuka Devi, "A Survey of Different Issue of Different clustering Algorithms Used in Large Data sets" , International Journal of Advanced Research in Computer Science and Software Engineering, pp.305-307, 2012.
- [5] OdedMaimon, LiorRokach, "Data Mining and Knowledge Discovery Handbook", Springer Science+BusinessMedia.Inc, pp.321-352, 2005.