

DATA MINING VIVA ASSIGNMENT (111717013066)

Grid-based Method

What is Clustering?

Clustering is the process of making a group of abstract objects into classes of similar objects.

- A cluster of data objects can be treated as one group.
- While doing cluster analysis, we first partition the set of data into groups based on data similarity and then assign the labels to the groups.
- The main advantage of clustering over classification is that, it is adaptable to changes and helps single out useful features that distinguish different groups.

Clustering Methods

Clustering methods can be classified into the following categories –

- Partitioning Method
- Hierarchical Method
- Density-based Method
- Grid-Based Method
- Model-Based Method
- Constraint-based Method

Grid-based Method

Grid-based approaches are popular for mining clusters in a large multidimensional space wherein clusters are regarded as denser regions than their surroundings. In this chapter, we present some grid-based clustering algorithms.

The computational complexity of most clustering algorithms is at least linearly proportional to the size of the data set. The great advantage of grid-based clustering is its significant reduction of the computational complexity, especially for clustering very large data sets. The grid-based clustering approach differs from the conventional clustering algorithms in that it is concerned not with the data points but with the value space that surrounds the data points. In general, a typical grid-based clustering algorithm consists of the following five basic steps:

1. Creating the grid structure, i.e. partitioning the data space into a finite number of cells.
2. Calculating the cell density for each cell.
3. Sorting of the cells according to their densities.
4. Identifying cluster centers.
5. Traversal of neighbor cells.

STING

Wang et al. (1997) proposed A Statistical Information Grid-based clustering method (STING) to cluster spatial databases. The algorithm can be used to facilitate several kinds of spatial queries. The spatial area is divided into rectangle cells, which are represented by a hierarchical structure. Let the root of the hierarchy be at level 1, its children at level 2, etc. The number of layers could be obtained by changing the number of cells that form a higher-level cell. A cell in level i corresponds to the union of the areas of its children in level $i + 1$. In the algorithm STING, each cell has 4 children and each child corresponds to one quadrant of the parent cell. Only two-dimensional spatial space is considered in this algorithm).