Uncertain Data Clustering in Distributed Peer-to-Peer Networks

Abstract:

Clustering has emerged as an essential data mining technique for statistical analysis, pattern recognition, and image segment ratio. It partitions the data into clusters according to the similarities between objects and helps in extraction of new information or discovering new patterns. In the past few decades, a large number of clustering algorithms have been proposed in which the K-means algorithm is one well-known clustering method. Then the variants of this algorithm are further discussed in the strong consistency of this method has been proved.

Existing System:

The traditional centralized clustering approaches for uncertain data have shown the weaknesses: 1) raw information sharing is discouraged due to the confidentiality and security requirements in distributed P2P networks; 2) effective data collection from all peers to the central site is not guaranteed due to the energy or bandwidth limitations; and 3) high-computational complexity with large data sets. These motivate seeking a new clustering algorithm in distributed network environments for uncertain data, i.e., the distributed uncertain data clustering.

Disadvantages:

➢ Distributed clustering algorithms on P2P networks are all implemented for certain data.

Proposed System:

We propose a novel distributed uncertain data clustering algorithm, in which the centralized global clustering solution is approximated by performing distributed clustering. To shorten the execution time, the reduction technique is then applied to transform the proposed method into its deterministic form by replacing each uncertain data object with its expected centroid. Finally, the attribute-weight-entropy regularization technique enhances the proposed
distributed clustering method to achieve better results in data clustering and extract the essential features for cluster identification. The experiments on both synthetic and real-world data have shown the efficiency and superiority of the presented algorithm.

Advantages:
➢ Distributed Uncertain Data Clustering

Modules:
➢ Distributed Uncertain Data Clustering.
➢ Attribute Weighted Distributed Uncertain Clustering.

SYSTEM REQUIREMENTS

H/W System Configuration:

Processor : Pentium –III
RAM : 256 MB (min)
Hard Disk : 20 GB
Key Board : Standard Windows Keyboard
Mouse : Two or Three Button Mouse
Monitor : SVGA

S/W System Configuration:

Operating System : Windows95/98/2000/XP
Application Server : Tomcat5.0/6.X
Front End : HTML, Jsp
Scripts : JavaScript.
Server side Script : Java Server Pages.
Database : MySQL 5.0
Database Connectivity : JDBC