

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 segmentation. It partitions the data into clusters according to the similarities between objects and helps in extraction of new information or discovering new patterns. However, in many real applications today, like sensor monitoring and location-based services, data mostly contains inherent uncertainty due to the random nature of the data generation, measurement inaccuracy, sampling discrepancy, data staling, and other errors.

EXISTING SYSTEM

In Existing System, Uncertain data clustering has been recognized as an essential task in the research of data mining. Many centralized clustering algorithms are extended by defining new distance or similarity measurements to tackle this issue. With the fast development of network applications, these centralized methods show their limitations in conducting data clustering in a large dynamic distributed peer-to-peer network due to the privacy and security concerns or the technical constraints brought by distributive environments.

DIS ADVANTAGES

- The privacy and security problems like no full data transmission are occurred.
- Bad savings in execution time will be achieved with affecting CR in clustering.

PROPOSED SYSTEM

In Proposed System, 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

- Raw information sharing is discouraged due to the confidentiality and security requirements in distributed P2P networks.
- It improves efficiency; a reduction technique is used to optimize the DUK-means algorithm.

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