Abstract

The precious information is embedded in large databases. To extract them has become an
interesting area of Data mining. Clustering, in data mining, is useful for discovering groups and identifying interesting distributions in the underlying data [5]. Among several clustering algorithms, we have considered CURE method from hierarchical clustering. CURE (Clustering usage Representatives) method finds clusters from a large database that is more robust to outliers, and identifies clusters having non-spherical shapes and wide variances in size. CURE employs a combination of data collection, data reduction by using random sampling and partitioning. With the availability of large data sets in application areas like bioinformatics, medical informatics, scientific data analysis, financial analysis, telecommunications, retailing, and marketing, it is becoming increasingly important to execute data mining tasks in parallel. At the same time, technological advances have made shared memory parallel machines commonly available to organizations and individuals. Although CURE provide high quality clustering, a parallel version was not available. Our new algorithm enabled it to outperform existing algorithms as well as to scale well for large databases without declining clustering quality.

References

- J. Han and M. Kamber; 2000, &quot;Data Mining: Concepts and Techniques&quot; Morgan Kaufmann.
- Pavel Berkin, Survey Of Clustering Data Mining Techniques &quot; 2000
- Richard J. Roiger, Michael W. Geatz, 2007, Data Mining A tutorial-based Primer &quot; Pearson Education, New Delhi
Enhancement of CURE Clustering Technique in Data Mining


- Hinneburg, Keim. Clustering Techniques for Large Data Sets. First publ. in: ACM SIGKDD 1999 Int. Conf. on Knowledge Discovery and Data Mining (KDD’99), San Diego, CA, September, 1999, pp. 141-181

Index Terms
Computer Science
Emerging Trends in Technology

Keywords
Data Mining  Kdd  Clustering  Issues  Parallelism