Abstract

This paper identifies the scope for improvement in the execution of baseline kNN join algorithms in a distributed environment. Improvements are suggested and the improved methods are applied in performing kNN joins on R*-Trees. The effectiveness of the proposed improvements have been experimentally verified and presented.

References

Improvisation to the R*-Tree kNN Join Principles in Distributed Environment

- Daniar Achakeev, Marc Seidemann, Markus Schmidt and Bernhard Seeger, 2012, &quot;Sort-based parallel loading of R-trees,&quot; Proceedings of the 1st ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data, pp. 62-70.
- Faloutsos and I. Kamel, 1994, &quot;Beyond uniformity and independence: Analysis of R-trees using the concept of fractal dimension,&quot; Proceedings of the 13th ACM SIGMOD Conference, pp. 4-13.
- Chi Zhang, Feifei Li and Jeffrey Jestes, 2012, &quot;Efficient parallel kNN joins for large data in Mapreduce,&quot; Proceedings of the 15th International Conference on Extending Database Technology, pp. 38-49.

Index Terms

Computer Science
Distributed Systems
Keywords
R*-Trees  kNN join  Hadoop  Mapreduce  z-values.