Abstract

In Data mining applications, very large training data sets with several million records are common. Decision trees are very much powerful and excellent technique for both classification and prediction problems. Many decision tree construction algorithms have been proposed to develop and handle large or small training data. Some related algorithms are best for large data sets and some for small data sets. Each algorithm works best for its own criteria. The decision tree algorithms classify categorical and continuous attributes very well but it handles efficiently only a smaller data set. It consumes more time for large datasets. Supervised Learning In Quest (SLIQ) and Scalable Parallelizable Induction of Decision Tree (SPRINT) handles very large datasets. But SLIQ requires that the class labels should be available in main memory beforehand. SPRINT is best suited for large data sets and it removes all these memory restrictions. The research work deals with the automatic selection of decision tree algorithm based on training dataset size. This proposed system first prepares the training dataset size using the mathematical measure. The result training set size problem will be checked with the available memory space. If memory is very sufficient then the tree construction will continue. After the classifying the data, the accuracy of the classifier data set
is estimated. The main advantages of the proposed method are that the system takes less time and avoids memory problem.

References

- Amir Bar-Or, Daniel Keren, Assaf Schuster, and Ran Wolff, &quot;Hierarchical Decision Tree Induction in Distributed Genomic Databases&quot;, IEEE Transactions on Knowledge and Data Engineering, vol. 17, No. 8, August 2005.
- Rakesh Agrawal, Tomasz Imielinski, and Arun Swami, &quot;Database mining: A performance perspective. IEEE Transactions on Knowledge and Data Engineering&quot;, 5(6):914(925, December 1993.
- Deborah R. Carvalho, Alex A. Freitas , &quot;A hybrid decision tree/genetic algorithm for coping with the problem of small disjuncts in data mining&quot; 2004.

Index Terms

Computer Science

Artificial Intelligence
Keywords
Decision Tree Algorithm  Classification  Data Mining  Data Set