Abstract

Relational databases are based on the theory of relational algebra because all the operations of RDBMS draw their functioning from the operations in relational algebra. The operations of relational algebra are defined on the sets, however, In general, the datamining algorithms requires databases which adopts the multiset philosophy to give better and more accurate results. Unrealized datasets ensures confidentiality of the actual datasets in the datamining process. C4.5 is a classic algorithm which works on mixed real world datasets. This paper proposes the application of Relational algebra for multisets to find the split criterion to be used in classification by the C4.5 algorithm. The results are shown by making the changes in original C4.5 algorithm in the weka tool setting.

References

International Conference A multi-set extended relational algebra: a formal approach to a practical issue
- Williams, J. 2010. Unrealization Approaches for Privacy Preserving Data Mining, A Thesis submitted in Department of Computer Science, University of Victoria.

Index Terms

Computer Science
Algorithms

Keywords
Multirelational algebra relation gain ratio