Abstract

Classification is a data analysis technique. The decision tree is one of the most popular classification algorithms in current use for data mining because it is more interpretable. Training data sets are not error free due to measurement errors in the data collection process. Traditional decision tree classifiers are constructed without considering any errors in the values of attributes of the training data sets. We extend such classifiers to construct effective decision trees with error corrected training data sets. It is possible to build decision tree classifiers with higher accuracies especially when the measurement errors in the values of the attributes of the training data sets are corrected appropriately before using those training data sets in decision tree learning. Error corrected data sets can be used not only in decision tree learning but also in many data mining techniques. In general, values of attributes in training datasets are always inherently associated with errors. Data errors can be properly handled by using appropriate error models or error correction techniques. Also, sometimes for preserving data privacy, attribute values in the original training data sets are modified so that modified data sets contain data values with some errors. Later on, these modified data sets are reconstructed before applying those tuples to data mining technique. This paper introduces an effective decision
Effective Decision Tree Learning

tree (EDT) construction algorithm that uses a new error adjusting technique (NEAT) in constructing more accurate decision tree classifiers. The idea behind this new error adjusting technique is that "many data sets with numerical attributes containing point data values have been collected via repeated measurements" and the process of repeated measurements is the common source of data errors in the training data sets. EDT describes an approach to correct the errors in the values of attributes of the training data sets and then error corrected attribute values of the data sets are used in decision tree learning.

References

- Jiawei Han, Micheline Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2006.
- Introduction to Machine Learning Ethem Alpaydin

Index Terms

Computer Science Information Sciences

Keywords

Decision tree Classification Data mining