Abstract

Classification is an important problem in data mining. Given a database of records, each with a class label, a classifier generates a concise and meaningful description for each class that can be used to classify subsequent records. A number of popular classifiers construct decision trees to generate class models. These classifiers first build a decision tree and then prune subtrees from the decision tree in a subsequent pruning phase to improve accuracy and prevent “overfitting”. In this paper, the different pruning methodologies available & their various features are discussed. Also the effectiveness of pruning is evaluated in terms of complexity and classification accuracy by applying C4.5 decision tree classification algorithm on Credit Card Database with pruning and without pruning. Instead of classifying the transactions either fraud or non-fraud the transactions are classified in four risk levels which is an innovative concept.
Reference

- Jiawei Han, Micheline Kamber, “Data Mining Concepts and Techniques”, pp. 279-328, 2001.
- Arun Poojari, “Data Mining techniques”, pp 150 -200, 1999
- Manish Mehta, Rakesh Agrawal et al., “SLIQ- A Fast Scalable Classifier for Data Mining.”, In 5th Intl. Conf. on Extending Database Technology, March 1996
- Salvatore, Philip et al.; “Meta learning agents for fraud and intrusion detection in Financial Information Systems.”, Inv paper Proceedings in International conference of Knowledge Discovery and Data mining, 1996.

Index Terms

Computer Science

Machine Learning

Key words

Decision tree classification

Pruning

Data Mining