Measures for Decision Tree Building

Tamás Gergely

Decision trees are special trees that contain some kind of decisions in the (internal) nodes and some kind of information in the leaves. Often, they are used as a method for knowledge representation in artificial intelligence. The construction of a decision tree (in most cases) can be split into two phases: building and pruning. The building of the decision trees is based on a metric called information gain (entropy-based metric), and it creates a full tree. Although the full tree contains the most precise information it also requires the most resources (the most space in memory). The aim of the pruning is to "balance" the tree between size and information by replacing some subtrees with leaves. There are several metrics used to balance the tree, and. they are obviously dependent on its functionality. In this paper I introduce some metrics we used for pruning our trees. We tried to compress specific data with some coder algorithms using decision tree models. This required new, specific metrics for tree pruning. Our experimental results show whether it is a good idea to use decision trees as models for compression algorithms.