Learning Naïve Bayes Tree for Conditional Probability Estimation

Han Liang1, Dr. Yuhong Yan2, Dr. Huajie Zhang1
1Faculty of Computer Science, University of New Brunswick
2 National Research Council of Canada

Problem Description & Motivation

Naïve Bayes Tree (NBT) is a typical classifier in machine learning. The intuition behind it is to use a kernel-based density estimator, such as naive Bayes, at each leaf and segment the data so that the conditional independence assumption is better verified. In our research, we propose a new learning algorithm to directly improve the conditional probability estimation in the diagram of NBT. The motivation of our work is that, in some cost-sensitive applications where costs are associated with conditional probabilities, the cost score function is optimized when the estimates of conditional probabilities are accurate.

Methodology

The learning algorithm is a greedy and recursive procedure where in each step, the conditional log likelihood (CLL) is used as the metric to expand the decision tree. When some bound conditions are met, the algorithm uses naïve Bayes to estimate the probabilities for leaf attributes given the class variable and the path attributes. Here is the pseudo-code of our algorithm:

□

Experimental Results

For the purpose of our study, we conducted three groups of experiments using 33 well-recognized sample sets from the UCI repository recommended by Weka. The detailed results have been showed as below: