Classification Trees based on Proportional-Reduction-in-Impurity Measures

Sung Jin Ahn
Gyeong Sang National University, Department of Information Statistics
#900 Gazwa-Dong
Jinju, 660-701, Rep. Of Korea
ahnsj@nongae.gsnu.ac.kr

1. Introduction

Classification trees are commonly constructed by recursively partitioning each associated feature space into mutually exclusive and exhaustive subsets in a systematic way of tree generation. Four essential elements in a classification tree design include 1) a data set consisting of feature vectors and their corresponding class labels, 2) a node-splitting criterion, 3) a rule for determining terminal nodes (leaves), and 4) a class-assigning rule.

The performance of a classification tree is generally evaluated in terms of the expected 'classification-effectiveness' loss. Due to the inherent computational complexity in constructing trees with the minimum expected loss (Hyafil and Rivest (1976)), heuristics are usually considered by employing steepest-descent greedy procedures. Each step of such a greedy procedure operates on a partially grown tree such that for splitting each node, a collection of splits permitted at that node are all evaluated to make the performance of the resulting tree improved most under the assumption that the branched nodes (children) will be terminal nodes. This implies that the resulting partitioning may become locally (one-step) optimal.

Chou (1991) has proposed an iterative descent algorithm to find a k-way partition for an arbitrary loss function, whose computational complexity is linearly proportional to the number of categories of each feature. The algorithm is a k-means like clustering algorithm that uses, as its distance measure, a generalized Kullback's information divergence. However, the algorithm does not have any test procedure for the best k selection, while it can find a partition to reduce most when k is given.

Zhou and Dillon (1991) have proposed the Goodman-Kruskal's symmetrical tau measure as a statistical-heuristic feature selection criterion. They showed that the measure is consistent with a Bayesian classifier, and discussed that the built-in statistical test (called Catanova or Catergorical ANOVA test, Light and Margolin (1971)) could similarly be employed. They also used the tau criterion to extend the classification tree to a probabilistic classifier, which provided a natural basis for prepruning (in contrast to postpruning, Brieman et al. (1984) and Gelfand, Ravishankar, and Delp (1991)). However, they didn't concern any feature formation procedure. Ahn (1995) introduced a tree-structured classification method based on impurity reduction measure of divergence for categorical distributions. Sung and Ahn (1998) presented PRI measures for categorical association as a generalization of PRE measures and derived some of their statistical properties. They also presented some application possibilities of PRI measures to clustering and merging problems, and to measuring of compositional associations.

In this paper we propose a unified method (in the sense of feature formation and selection) for optimal partitioning by use of PRI measure, which practically seeks a partition characterized as of being both the most explanatory and the most significant (as in Chou (1991) and Biggs et al. (1991), respectively).

REFERENCE

Ahn, S. J. (1995), 'Tree-structured Classification based on Variable Splitting,' *Korean Commun. Statist.*, 2(1), pp 74-88.

Biggs, David, Ville, Barry de, and Suen, Ed (1991), 'A method of choosing multiway partitions for classification and decision trees,' J. of *Appl. Statist.*, vol. 18, pp 49-62.

Brieman, L., Friedman, J. H., Olsen, R. A., and Stone, C. G. (1984), *Classification and regression trees*. Belmont, CA: Wadsworth.

Chou, P. A. (1991), 'Optimal partitioning for classification and regression trees,' *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 13, pp 340-354.

Gelfand, S. B., Ravishankar, C. S., and Delp, E. J. (1991), 'An iterative growing and pruning algorithm for classification tree design,' *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp 163-174.

Hyafil, L. and Rivest, R. L. (1976), 'Constructing optimal binary decision trees is NP-complete,' *Inform. Processing Lett.*, vol. 5, pp. 15-17.

Light, R. J. and Margolin, B. H. (1971), 'An analysis of variance for categorical data,' *Jour. of American Stat. Asso.*, vol. 66, pp 534-544.

Sung, C. S. and Ahn, S. J. (1998), 'A Proportional-Reduction-in- Impurity Measure of Association for Categorical Variables,' *Commun. Statist.-Theory Meth.*, 27(8) pp. 2083-2110.

Zhou, X. J. and Dillon, T. S. (1991), 'A statistical-heuristic feature selection criterion for decision tree induction,' *IEEE Trans. on Pattern Anal. and Mach. Intell.*, vol. 13, pp 834-841.

RESUME

1980.2. B.S. in Statistics, Korea University

1982.2. M.E. in OR, KAIST

1994.2. Ph.D. in Decision Science, KAIST

1983.4- From Full-time Lecturer, now Professor of Statistics, GyeongSang National University