BAYESIAN EVIDENCE FRAMEWORK FOR DECISION TREE LEARNING

Ratthachat Chatpatanasiri and Boonserm Kijsirikul
Department of Computer Engineering, Chulalongkorn University,
Pathumwan, Bangkok, 10330, Thailand.
Email: Ratthachat.c@student.chula.ac.th, Boonserm.k@chula.ac.th

Abstract

This work is primary interested in the problem of learning a *single* decision (or classification) tree from the observed data. Inducing a single decision tree though has a high risk to be *overfitted*, the induced tree is easily interpretable. Researchers have invented various methods such as tree pruning or tree averaging for preventing the induced tree from overfitting (and from underfitting) the data [1,3,4]. In this paper, instead of using those conventional approaches, we apply the *Bayesian evidence framework* of Gull, Skilling and Mackay [2] to the selection process of a decision tree. We derive a formal function to measure appropriateness for each decision tree given a set of observed data. Our method, in fact, is analogous to a well-known Bayesian inference for interpolating noisy continuous-value data [2]. As in regression problems, this derived function reasonably deals with the issue of underfitting-overfitting tradeoff and automatically quantifies the principle of *Ockham's razor*. Different probability models for hyperparameters are also investigated. Comparative results to those conventional methods are shown in experiments.

References:

[1] H. A. Chipman, E. I. Georege and R. E. McCulloch. Bayesian CART Model Search. Journal of the American Statistical Association (1998).

[2] D. J. C. Mackay. Bayesian Interpolation. Neural Computation (1992).

[3] R. Quinlan. C4.5:Programs for Machine Learning. Morgan Kaufmann Publishers (1993).

[4] C. Scott and R. Nowak. Dyadic classification trees via structural risk minimization. In Advances in Neural Information Processing Systems, 14. MIT Press (2002).

Key Words: Decision Tree Learning, Underfitting and Overfitting, Evidence Framework, Ockham's Razor.