On Speed of Stochastic Search on Decision Trees

Márton Ispány, Ilona Krasznahorkay

Dept of Applied Mathematics and Probability Theory
University of Debrecen, Hungary
e-mail: krasnil@inf.unideb.hu

The construction of decision trees is a commonly used and easily applied way of supervised learning. The aim is the prediction of a binary target variable on the basis of many predictor variables. This technique divides the field of predictors getting the target variable more and more homogeneous along the resulting partition. We modified the CART algorithm developed by Breiman et al. [1], improving with a stochastic search on the set of decision trees applying the Markov Chain Monte Carlo method. It was first proposed in a Bayesian framework by Chipman et al. [2].

We prove sharp rate of convergence for the Markov chain behind the algorithm. Namely, we show that $cn \log n$ steps are sufficient to reach the target distribution, where n is the number of the nodes in the corresponding decision tree. The technique of the proof is based on powerful estimation of the second largest eigenvalue developed in [3], [4] and [5].

References

[1] Breiman, L., Friedman, J. H., Olsen, A. O., Stone, C. J. (1984). *Classification and Regression Trees*, Wadsworth International Group.

[2] Chipman, H. A., George, E. I., McCulloch, R. E. (1998). *Bayesian CART Model Search*, Journal of the American Statistical Association, **93**, 443, 935-960.

[3] Jerrum, M. (1998). *Mathematical Foundations of the Markov Chain Monte Carlo Method.* Probabilistic Methods for Algorithmic Discrete Mathematics, Algorithm Combin., **16**, Springer, Berlin, 116-165.

[4] Jerrum, M., Sinclair, A. (1996). *The Markov Chain Monte Carlo Method: An Approach to Approximate Counting and Integration*, Approximation Algorithm for NP–hard Problems, (Dorit Hochbaum ed.), PWS.

[5] Diaconis, P., Holmes, S. P. (2002). *Random Walks on Trees and Matchings*, Electronic Journal of Probability, **7**, 6, 1-17.