MASTINO: a suite of R functions to learn Bayesian Networks from data.

Massimiliano Mascherini, PhD
Department of Statistics
University of Florence
V.le Morgagni 59, 50134, Florence, Italy
mascher@ds.unifi.it

Abstract

Bayesian Networks (BNs), [2], are a widespread tool in many areas of artificial intelligence and automated reasoning because they perform probabilistic inference through very efficient algorithms. However, the problem of searching the BN that best depicts the dependence relations entailed in a database of cases it is hard to solve. Structural learning exploits algorithms which typically combine expert's knowledge with the information gathered in a database.

In this paper I present MASTINO: a suite of R functions to learn BNs from data. MASTINO is built on the top of the DEAL package, [1], and it provides several functions to learn Bayesian Networks from data in the score-and-search framework. In particular, the *P*-metric, [3], a new score to evaluate Bayesian Networks encoding prior information on structures, and the MGA algorithm, [4], an innovative genetic algorithm to search for the best Bayesian Networks, are implemented in MASTINO as well as many utility functions to work with BNs.

MASTINO has been successfully tested on several well-known Machine Learning benchmark datasets with excellent results. The package is freely available for use with R and it can be downloaded from the web site of the author: *http://www.ds.unifi.it/mascherini/*

References

[1] S. G. Bøttcher and C. Dethlefsen. DEAL: A package for learning bayesian networks. *Journal of Statistical Software*, 8(20):1–40, 2003.

[2] F. V. Jensen. *An introduction to Bayesian Networks*. Springer Verlag, New York, N.Y., 1996.

[3] M. Mascherini and F. M. Stefanini. Encoding structural prior information to learn large bayesian networks. *WP of the Department of Statistics - University of Florence*, 13, 2005.

[4] M. Mascherini and F. M. Stefanini. M-GA: A genetic algorithm to learn conditional gaussian bayesian networks. *Proceedings of the IEEE International Conference on Computational Intelligence for Modelling, Control and Automation*, 2005.