L_1-based compression of random forest models

Arnaud Joly A.JOLY@ULG.AC.BE
François Schnitzler FSCHNITZLER@ULG.AC.BE
Pierre Geurts P.GEURTS@ULG.AC.BE
Louis Wehenkel L.WEHENKEL@ULG.AC.BE
University of Liège - Department of EE and CS & GIGA-research
Liège, Sart-Tilman, B-28, B-4000, Belgium

Keywords: Ensemble of randomized trees; Pruning; L1-norm regularization; LASSO

The goal of this one page abstract is to present the following article (Joly et al., 2012).

High-dimensional supervised learning problems, *e.g.* in image exploitation and bioinformatics, are more frequent than ever. Tree-based ensemble methods, such as random forests (Breiman, 2001) and extremely randomized trees (Geurts et al., 2006), are effective variance reduction techniques offering in this context a good trade-off between accuracy, computational complexity, and interpretability.

The number of nodes of a tree ensemble grows as nM (n being the size of the learning sample and M the number of trees in the ensemble). Empirical observations show that the variance of individual trees increases with the dimension p of the original feature space used to represent the inputs of the learning problem. Hence, the number $M(p)$ of ensemble terms yielding near-optimal accuracy, which is proportional to this variance, also increases with p. The net result is that the space complexity of these tree-based ensemble methods will grow as $nM(p)$, which may jeopardize their practicality in large scale problems, or when memory is limited.

While pruning of single tree models is a standard approach, less work has been devoted to pruning ensembles of trees. To further investigate the feasibility of reducing the space complexity of tree-based ensemble models, we consider the following experiment: (i) build an ensemble of trees; (ii) apply to this ensemble a 'compression step' by reformulating the tree-ensemble based model as a linear model in terms of node indicator functions and by using an L_1-norm regularization approach - à la Lasso (Tibshirani, 1996) - to select a minimal subset of these indicator functions while maintaining predictive accuracy.

We propose an algorithmic framework and an empirical investigation of this idea, based on three complementary datasets, and we show that indeed it is possible to so compress significantly tree-based ensemble models, both in regression and in classification problems. We also observe that the compression rate and the accuracy of the compressed models further increase with the ensemble size M, even beyond the number $M(p)$ of terms required to ensure convergence of the variance reduction effect.

Acknowledgments

F. Schnitzler is supported by a F.R.I.A. scholarship. This work was funded by the Biomagnet IUAP network of the Belgian Science Policy Office and the Pascal2 network of excellence of the EC. The scientific responsibility is the authors'.

References

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5–32.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*, 3–42.

Joly, A., Schnitzler, F., Geurts, P., & Wehenkel, L. (2012). l_1-based compression of random forest models. *20th European Symposium on Artificial Neural Networks*.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.