

Forêts aléatoires

Sylvain Arlot*

Les forêts aléatoires, introduites par Breiman [Bre01], sont un parfait exemple du fossé entre théorie et pratique en apprentissage automatique : elles fonctionnent remarquablement bien (et sont donc très utilisées, elles le sont par exemple au sein de la Kinect [SFC⁺11]), mais on ne sait pas (encore) le démontrer mathématiquement.

Il s’agit d’une méthode d’ensemble : à partir d’un jeu de données, on ne cherche pas à « apprendre » directement un très bon « prédicteur », mais on construit plutôt une grande famille de prédicteurs, dont on agrège ensuite les prédictions. Ainsi, même si chacun de ces prédicteurs (qui sont des arbres de décision) n’est pas très bon, en les considérant collectivement, on obtient une « forêt » qui est bien meilleure que chacun des « arbres ». Autrement dit, pour prendre une décision, plutôt que de demander à un seul expert, on organise un vote entre mille personnes qui s’y connaissent un peu... et cela marche mieux ! Pour en savoir plus, on pourra lire le premier chapitre de [Gen10] qui propose une très bonne introduction sur ce sujet.

Le but de ce mémoire est de comprendre le fonctionnement des forêts aléatoires, en particulier en étudiant le résultat de consistance obtenu par [BDL08] et en les mettant en œuvre sur quelques exemples (la proportion exacte entre théorie et expérimentation restant à choisir).

Références

- [BDL08] Gérard Biau, Luc P. Devroye, and Gábor Lugosi. Consistency of random forests and other averaging classifiers. *J. Mach. Learn. Res.*, 9 :2015–2033, 2008.
- [Bre01] Leo Breiman. Random forests. *Machine Learning*, 45 :5–32, 2001.
- [Gen10] Robin Genuer. *Forêts aléatoires : aspects théoriques, sélection de variables et applications*. PhD thesis, University Paris-Sud 11, 2010. <http://tel.archives-ouvertes.fr/tel-00550989/>.
- [SFC⁺11] Jamie Shotton, Andrew W. Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, and Andrew Blake. Real-time human pose recognition in parts from single depth images. In *CVPR*, pages 1297–1304, 2011.

*sylvain.arlot@ens.fr