
Forêts aléatoires pour l'apprentissage de données massives

Robin Genuer*^{†1}, Jean-Michel Poggi², Christine Tuleau-Malot³, and Nathalie Villa-Vialaneix⁴

¹INRIA, SISTM team ISPED, INSERM U-897 – Université de Bordeaux (Bordeaux, France) – France

²Laboratoire de Mathématiques d'Orsay (LM-Orsay) – CNRS : UMR8628, Université Paris XI - Paris Sud – France

³Laboratoire Jean Alexandre Dieudonné (JAD) – Université Nice Sophia Antipolis (UNS), CNRS : UMR7351 – Université de Nice - Sophia Antipolis U.M.R. no 6621 du C.N.R.S. Parc Valrose 06108 Nice Cedex 02 France, France

⁴INRA, UR875, MIAT (MIAT) – Institut national de la recherche agronomique (INRA) : UR875 – F-31326 Castanet Tolosan, France

Résumé

Les forêts aléatoires (Breiman, 2001) sont une méthode d'apprentissage largement utilisée dans le cadre de problèmes de régression ou de classification en raison de leur flexibilité et de leurs bonnes performances de prédiction. Dans cette proposition de communication, nous présentons les résultats de simulations comparant les diverses approches permettant d'utiliser l'algorithme de forêts aléatoires dans le contexte de données massives.

*Intervenant

[†]Auteur correspondant: robin.genuer@isped.u-bordeaux2.fr