

Compte-rendu sur le livre "Les forêts aléatoires avec R" par Robin Genuer et Jean-Michel Poggi

Title: Book Review: "Les forêts aléatoires avec R" by Robin Genuer and Jean-Michel Poggi

Badih Ghattas¹

Les arbres de décisions sont un des piliers des techniques d'apprentissage statistique, et les forêts aléatoires sont les techniques qui sont de nos jours parmi les plus répandues et employées dans le milieu industriel, mais aussi dans tous les domaines d'application et d'utilisation de la statistique, en particulier pour la modélisation et la prévision.

Cet ouvrage destiné à des étudiants de Licence ou de Master de toutes filières ayant des connaissances de base en statistique, décrit les techniques de type forêts aléatoires et leur mise en œuvre avec le logiciel R. Les forêts aléatoires étant des méthodes d'ensemble d'arbres, et leur succès étant partiellement justifié par la possibilité de hiérarchiser les variables explicatives, on retrouve dans ce livre quatre chapitres en plus de l'introduction

- Les arbres CART.
- Les forêts aléatoires (FA).
- L'importance des variables (dans les FA).
- La sélection de variables (par les FA).

Chacun de ces chapitres décrit une approche avec un niveau de détails techniques suffisant pour comprendre l'implémentation de celle-ci dans le *package* correspondant de R. Chaque approche est illustrée soit sur des données réelles (spam, pollution, génomique) soit sur des données simulées (*Toys*).

Bien qu'une seule approche soit décrite en détail avec le package R correspondant, les auteurs donnent des références à des variantes et à des bibliothèques de R qui permettent de les utiliser. L'ensemble des techniques citées ou décrites constituent l'état de l'art dans le domaine.

Les illustrations sur les données réelles ou simulées servent à donner des exemples approfondis d'utilisation des méthodes. Les auteurs ont pris le soin de décrire les données en détail et ont pris le temps de commenter les résultats et de les analyser en vue des différentes options et paramètres que présentent les méthodes et les fonctions disponibles.

¹ I2M, AMU, adress.

E-mail : badihghattas@gmail.com

Un autre aspect particulièrement intéressant et utile pour tout utilisateur de ces approches est analysé de manière assez complète; la sensibilité des approches décrites quant aux hyperparamètres de réglages de ces approches. Le choix des valeurs des hyperparamètres est souvent une tâche non triviale qui peut affecter de manière très significative les performances des méthodes utilisées. Les auteurs ont montré ceci à travers les exemples bien choisis, et ont souvent donné des indications précises sur la manière de procéder pour régler ou optimiser les hyperparamètres.

J'ajouterai enfin que le manuel est très clair, très bien écrit et agréable à lire. C'est un très bon guide pour l'utilisation et la mise en œuvre de ces techniques avec le logiciel R.

Les forêts aléatoires avec R, Robin Genuer et Jean-Michel Poggi, Presses Universitaires de Rennes, 112 pages, ISBN : 978-2-7535-7710-7