

Les raisons majoritaires : des explications abductives pour les forêts aléatoires

Gilles Audemard*, Steve Bellart*
Louenas Bounia*, Frédéric Koriche*
Jean-Marie Lagniez*, Pierre Marquis***

*Univ. Artois, CNRS, CRIL, F-62300 Lens

** Institut universitaire de France

"nom"@cril.fr,

<http://www.cril.univ-artois.fr/>

Résumé. Les forêts aléatoires constituent un modèle d'apprentissage automatique efficace, ce qui explique qu'elles soient encore massivement utilisées aujourd'hui. S'il est assez facile de comprendre le fonctionnement d'un arbre de décision, il est beaucoup plus complexe d'interpréter la décision prise par une forêt aléatoire, car elle est typiquement issue d'un vote majoritaire entre de nombreux arbres. Nous examinons ici diverses définitions d'explication abductive du classement prédit par une forêt aléatoire. Nous nous intéressons au problème de leur génération (trouver une explication) et au problème de leur minimisation (trouver une explication parmi les plus courtes). Nous montrons notamment que les explications abductives irredondantes (ou raisons suffisantes) peuvent être difficiles à calculer pour les forêts aléatoires. Nous proposons à leur place les raisons majoritaires, des explications abductives en théorie moins concises mais que l'on peut calculer en temps polynomial.

1 Introduction

Les progrès remarquables réalisés durant les deux dernières décennies en matière d'apprentissage automatique ont conduit à utiliser maintes fois ces technologies. Ce faisant, des modèles de prédiction offrant de hauts degrés de précision ont été intégrés ou en cours d'intégration dans des applications issues de domaines variés. Mais la haute précision de ces modèles se fait souvent au détriment de leur interprétabilité, ce qui peut se révéler critique dans certains domaines. Dans le monde médical par exemple, si un classifieur identifie un patient comme malade suite à une radiographie, le médecin doit pouvoir demander au classifieur ce qui sur la radiographie réalisée explique le classement obtenu.

Dans cet article, nous nous concentrons sur les classements réalisés à partir du modèle des forêts aléatoires, un modèle d'apprentissage automatique bien connu et couramment utilisé, basé sur le vote majoritaire de plusieurs arbres de décision appris à partir d'instances tirées aléatoirement avec remise dans l'ensemble d'apprentissage considéré (Breiman, 2001). Ces forêts sont en pratique faciles à apprendre et elles offrent une bonne robustesse au bruit. C'est