

# Classificateurs aléatoires Topologiques à base de graphes de voisinage

Fabien Rico\*, Djamel A. Zighed\*

\*Laboratoire Eric, Univ. de Lyon, 5, av. P. Mendès France, 69676 Bron Cedex, France  
fabien.rico@univ-lyon1.fr & zighed@univ-lyon2.fr  
<http://eric.univ-lyon2.fr>

**Résumé.** En apprentissage supervisé, les Méthodes Ensemble (ME) ont montré leurs qualités. L'une des méthodes de référence dans ce domaine est les Forêts Aléatoires (FA). Cette dernière repose sur des partitionnements de l'espace de représentation selon des frontières parallèles aux axes ou obliques. Les conséquences de cette façon de partitionner l'espace de représentation peuvent affecter la qualité de chaque prédicteur. Il nous a semblé que cette approche pouvait être améliorée si on se libérait de cette contrainte de manière à mieux coller à la structure topologique de l'ensemble d'apprentissage. Dans cet article, nous proposons une nouvelle ME basée sur des graphes de voisinage dont les performances, sur nos premières expérimentations, sont aussi bonnes que celles des FA.

## 1 Introduction

En data mining, plus particulièrement en apprentissage supervisé, on a de plus en plus recours à la combinaison d'un ensemble de prédicteurs. Le principe consiste à générer un ensemble de classifieurs selon une ou plusieurs méthodes d'apprentissage données et d'agréger ensuite leur prédiction par un vote.

Le principe commun à de nombreuses ME est que, l'espace des observations est découpé en régions soit de manière quasi aléatoire comme le proposent [Kleinberg (2000, 1996); Ho (1998); Ho et Kleinberg (1996); Fradkin et Madigan (2003)] soit selon des règles générales fixées par un algorithme comme les arbres de décision comme le fait [Breiman (2001)]. Chaque région constitue ensuite un classifieur local.

Il nous semble que la manière dont sont construites ces régions doit influencer sur chaque classifieur individuel et par conséquent peut affecter le classifieur agrégé. Nous proposons un nouveau moyen de construire une telle méthode basé sur l'utilisation de graphes de voisinage. Cette méthode présente d'ores et déjà des performances comparables aux Forêts Aléatoires (FA) considérées souvent comme très performantes.

La section suivante va donner les principales définitions et formaliser les méthodes ensemble en prenant comme exemple les forêts aléatoires. Dans la section 3, nous introduisons l'utilisation des graphes de voisinage. La section 4 montrera nos premiers résultats comparés à ceux des forêts aléatoires.