

Une approche ensembliste inspirée du boosting en classification non supervisée

Romain Billot (*,**,*), Henri-Maxime Suchier (*,****)
Stephane Lallich (*)

* Université Lyon 2, Laboratoire ERIC, 5 avenue Pierre Mendès-France,
69676 Bron Cedex, France

** Laboratoire d'Ingénierie Circulation Transports (LICIT), INRETS-ENTPE
25 Avenue François Mitterand Case 24, 69675 Bron Cedex, France

*** Laboratoire de Mathématiques Appliquées aux Systèmes (MAS), Ecole Centrale Paris,
92295 Châtenay-Malabry- France

**** Laboratoire informatique, Agrocampus Rennes, 65 rue de Saint-Brieuc, CS 84215
35042 Rennes Cedex - France

Contacts : billotro@gmail.com , hmsuchier@gmail.com , stephane.lallich@univ-lyon2.fr

Résumé. En classification supervisée, de nombreuses méthodes ensemblistes peuvent combiner plusieurs hypothèses de base afin de créer une règle de décision finale plus performante. Ainsi, il a été montré que des méthodes comme le *bagging* ou le *boosting* pouvaient se révéler intéressantes, tant dans la phase d'apprentissage qu'en généralisation. Dès lors, il est tentant de vouloir s'inspirer des grands principes d'une méthode comme le *boosting* en classification non supervisée. Or, il convient préalablement de se confronter aux difficultés connues de la thématique des ensembles de regroupements (correspondance des classes, agrégation des résultats, qualité) puis d'introduire l'idée du *boosting* dans un processus itératif. Cet article propose une méthode ensembliste inspirée du *boosting*, qui, à partir d'un partitionnement flou obtenu par les c-moyennes floues (fuzzy-c-means), va insister itérativement sur les exemples difficiles pour former une partition dure finale plus pertinente.

1 Introduction

Il est courant de séparer le domaine de l'apprentissage automatique en deux domaines distincts. D'un côté, l'apprentissage supervisé désigne un cadre où les exemples sont reliés à une information relative à leur classe, à un concept. Les méthodes supervisées produisent par la suite, à partir d'une base d'exemples d'apprentissage pour lesquels la classe est connue, une règle de décision visant à prédire la classe de nouvelles observations. Cette règle de décision, appelée aussi classifieur ou hypothèse, peut être considérée géométriquement comme une hypersurface séparant les exemples représentés dans un espace multidimensionnel.