

Clustering incrémental pour un apprentissage distribué et robuste: vers un système évolutif...

Yann Prudent, Abdel Ennaji

PSI - FRE CNRS 2645 - Université de Rouen
{yann.prudent, abdel.ennaji}@univ-rouen.fr

Résumé. Ce papier présente un système d'apprentissage multi-classifieurs dont la conception est pilotée par la topologie des données d'apprentissage. Il présente un nouvel algorithme d'apprentissage de carte neuronale auto-organisée incrémentale en données. Cette carte est utilisée pour distribuer les tâches de classification sur un ensemble de classifieurs. Seuls certains classifieurs sont activés en fonction de l'activité des neurones de la carte pour une donnée. De plus, le système proposé permet d'établir un critère de confiance s'affranchissant totalement du type de classifieurs utilisés. Ce coefficient permet de contrôler efficacement le compromis *Erreur/Rejet*. Des résultats comparatifs de cette architecture sont donnés sur la base de segmentation d'images de l'UCI et en reconnaissance de chiffres manuscrits.

1 Introduction

En reconnaissance de formes, les phases d'apprentissage et de classification constituent des étapes fondamentales qui conditionnent en grande partie les performances. Les techniques d'apprentissage artificiel ont connu ces dernières décennies des avancées fondamentales à travers des modèles tels que les réseaux de neurones et les SVM (*Support Vector Machines* ou *Séparateurs à Vaste Marge*) [Vapnik, 1995], qui montrent globalement de bonnes performances. Ces méthodes, basées sur différentes théories et méthodologies, ont été considérées comme autant de solutions possibles à un même problème. Cependant, leur développement n'a pas permis de mettre en évidence la supériorité incontestable d'une méthode sur une autre face aux contraintes et à la complexité des applications pratiques. De plus, peu de modèles permettent de fournir une estimation de la fiabilité de la décision ou un coefficient de confiance qui ne soit pas simplement une probabilité *a posteriori* comme c'est souvent le cas. Une solution admise dans la communauté pour résoudre ce type de problèmes consiste à adopter une conception modulaire et distribuée du système de classification [Duin, 2002] [Kuncheva, 2002] [Gori et Scarselli, 1998].

Dans ce contexte, ce papier présente un système d'apprentissage multi-classifieurs dont la conception est pilotée par la topologie des données d'apprentissage. Plusieurs contributions sont à noter dans le domaine de la décomposition de tâches de classification. Nous pouvons distinguer les approches purement supervisées des approches hybrides combinant à la fois des algorithmes supervisés et non supervisés.

Dans [Jacobs *et al.*, 1991], les auteurs proposent une approche qui décompose l'ensemble d'apprentissage en plusieurs sous-ensembles puis entraîne plusieurs réseaux sur