

Classification automatique collaborative multi-stratégies

Application à la télédétection

**Par Pierre Gañçarski, Maître de Conférences, HDR,
Laboratoire SIIT, Université Louis Pasteur, Strasbourg**

Résumé

Alors que la classification supervisée cherche à modéliser au plus juste et au plus précis un phénomène, la classification non supervisée a pour but principal de permettre à l'utilisateur de s'appropriier ses données, de les comprendre et d'en extraire des régularités ou des propriétés cachées pouvant être grossières, incomplètes voire entachées d'erreurs. Placer l'utilisateur au cœur du processus de classification non supervisée est donc primordial. Or, cet utilisateur est bien souvent débordé par la profusion et la complexité des méthodes de classification et des données qu'il est amené à traiter.

Notre rôle en tant qu'informaticien est donc double. Si d'une part, nous nous devons d'offrir à cet utilisateur des méthodes de classification efficaces, d'autre part, celles-ci se doivent d'être les plus naturelles possible aussi bien dans leur façon de procéder (algorithmes mis en jeu) que dans leur paramétrage. Notre activité de recherche se situe dans cette optique. Notre volonté est de proposer un processus de classification non supervisée centré utilisateur.

Les travaux décrits dans ce mémoire se sont déroulés suivant deux grands axes.

Dans un premier axe de recherche, nous avons étudié et implanté un nouveau processus de classification basé sur la collaboration entre méthodes de classification existantes. Chaque classifieur collabore avec les autres classifieurs et utilise leurs résultats pour remettre en cause ses propres résultats. Suite à ce processus de collaboration, une solution unifiant peut alors être construite par un mécanisme de vote adapté. Ce processus de classification non supervisée collaborative multi-stratégies a été implémenté par la méthode Samarah (Système d'Apprentissage Multi-Agents de Raffinement Automatique de Hiérarchies).

Dans un deuxième axe, nous avons étudié et défini une famille de méthodes coévolutives de classification non supervisée avec pondération d'attributs inspirées de *K*-means. Puis, pour nous affranchir de la nécessité d'avoir une mesure de distance, nous avons étudié et défini la méthode maclaw (Modular Approach for Clustering with Local Attribute Weighting) qui consiste à construire localement un cluster par individu, la classification globale étant construite par union de ces clusters (un par population). Ces clusters peuvent être définis à partir d'algorithmes de classification différents rejoignant ainsi l'approche multi-stratégies définie précédemment.

Nous avons validé nos méthodes entre autres sur des images de télédétection. Les expériences menées ont produit des résultats très probants et très intéressants dans l'optique d'une utilisation réelle de nos méthodes. Dans ce mémoire, nous nous intéresserons uniquement aux expériences et validations faites en collaboration avec des géographes dans le cadre de la télédétection.