

# Algorithme génétique de pondération d'attributs pour une classification non supervisée d'objets complexes

Alexandre Blansché et Pierre Gançarski

LSIIT, UMR 7005 CNRS-ULP  
Parc d'Innovation, Boulevard Sébastien Brant  
67412 ILLKIRCH  
{blanche, gancarski}@lsiit.u-strasbg.fr  
<http://lsiit.u-strasbg.fr/afd/>

**Résumé.** La classification non supervisée d'objets complexes, composés d'un nombre important d'attributs est souvent problématique. En effet il existe très fréquemment des corrélations entre les attributs ainsi que des attributs bruités ou non pertinents. Pour résoudre ce problème nous proposons une méthode de pondération d'attributs non supervisée par approche *wrapper*, cherchant des pondérations continues et locales.

Dans cette méthode, un ensemble d'extracteurs (de classe) est défini. Chaque extracteur est muni d'une stratégie qui lui permet d'extraire une unique classe. Chaque extracteur utilise une pondération différente sur les attributs. Le résultat global est obtenu par l'ensemble des classes ainsi extraites. Afin de chercher l'ensemble de poids optimal permettant d'obtenir la meilleure classification possible, nous utilisons un apprentissage par coévolution coopérative.

Dans cet article, nous précisons notre méthode, en particulier comment sont définis les extracteurs, comment est évaluée la qualité de la classification, comment se déroule l'apprentissage et comment est construit le résultat final. Nous présenterons nos premiers résultats, sur des données artificielles et sur des images de télédétection.

## 1 Introduction

La classification non supervisée consiste à créer des classes à partir d'un ensemble d'objets, telles qu'une classe regroupe des objets similaires et deux classes différentes des objets dissemblables. Contrairement à une approche supervisée, la classification non supervisée ne se base pas sur un ensemble d'exemples, mais opère uniquement à partir des propriétés intrinsèques des objets. Un panorama très complet des méthodes existantes est donné dans [Grabmeier et Rudolph, 2002].

Ces méthodes ont prouvé leur pertinence dans bien des domaines. Néanmoins, les données à traiter sont de plus en plus complexes :

- de part leurs types (intervalles, distributions, histogrammes, etc.) ;
- de part l'hétérogénéité de leurs attributs (données multi-sources) ;
- de part la qualité de ces attributs.