

Approches génétiques pour une classification non supervisée avec pondération locale d'attributs

Pierre Gañarski,
LSIIT
Strasbourg

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 différentes méthodes de pondération d'attributs non supervisée par approche wrapper, cherchant des pondérations continues et locales. Toutes ces méthodes sont basées sur une approche génétique dans laquelle des individus (qualifiés d'extracteurs de classes(s)) évoluent afin d'optimiser la classification finale.

Elles diffèrent :

- d'une part suivant la stratégie d'évolution mise en place : approche darwinienne, lamarckienne ou enfin baldwindienne;
- d'autre part en fonction du nombre de populations utilisées : une seule population d'individus cherchant chacun l'ensemble des pondérations sur l'ensemble des classes ou alors plusieurs populations dans lesquelles chaque individu cherche à extraire au mieux une et une seule classe.

Dans la première partie de l'exposé, nous montrerons les six méthodes proposées et les comparerons à la méthode de pondération d'attributs Weighting-Kmeans

Dans une deuxième partie, nous montrerons une proposition d'une méthode co-évolutive de classification non supervisée avec pondération locale automatique d'attributs dans le cas d'absence de fonction de distance : la fonction de fitness ne peut plus être basée sur un calcul simple d'inertie intra ou inter classe.