

Sélection d'attributs et classification d'objets complexes

Alexandre Blansché, Pierre Gañarski

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/>

La classification de données complexes est souvent problématique lorsque les objets sont composés d'un nombre important d'attributs. En effet il existe très fréquemment de fortes corrélations entre les attributs ainsi que des attributs bruités ou non pertinents. Il est alors nécessaire de choisir correctement les attributs à utiliser pour obtenir une classification pertinente.

Pour cela, nous proposons une nouvelle méthode de sélection d'attributs non supervisée consistant à déterminer pour chaque classe à extraire une pondération sur les attributs. Les pondérations obtenues seront donc différentes, mais optimales pour chacune des classes cherchées. En effet, à notre avis, lors d'une classification, l'importance d'un attribut dépend de la classe que l'on désire mettre en évidence.

La méthode utilisera un ensemble de classifieurs/extracteurs paramétré chacun par une pondération différente des attributs. L'apprentissage consistera donc à trouver pour chacun de ces classifieurs la pondération optimale telle que la classe extraite soit la meilleure possible, mais aussi, et surtout, que l'ensemble des classes extraites forment une bonne classification des données. Il est à noter que les classes à extraire ne sont pas connues à priori.

Il est nécessaire de définir un nouveau critère de qualité d'une classification, car chaque pondération définit une métrique différente. De manière simplifiée, notre critère consiste à affecter une bonne qualité à la classification globale construite à partir des classes proposées par chacun des classifieurs si celle-ci forme une bonne répartition (idéalement une partition) dans l'espace des données, et une mauvaise qualité s'il y a des superpositions ou des objets non classifiés.

La recherche de ces pondérations se fait par coévolution coopérative. Nous définissons ainsi une population par classe cherchée. Les individus évoluent (par croisements, mutations, etc.) au sein de leur population uniquement, indépendamment des autres, mais la qualité d'un individu est définie en fonction des individus des autres populations via la classification globale construite.

Cette méthode a été validée sur des images de télédétection. Après une segmentation classique de l'image, nous avons caractérisé les régions obtenues par de très nombreux critères (texture, forme, etc.). Nous avons ensuite appliqué notre méthode à la classification de ces régions. Notre méthode s'est avérée parfaitement adaptée pour la sélection des caractéristiques les plus pertinentes. Les résultats obtenus sont bien meilleurs qu'avec une méthode classique, sans pondération ou à pondération unique sur l'ensemble des classes.