

Une nouvelle stratégie d'Apprentissage Bayésienne

Alexis Bondu*, Vincent Lemaire**
Marc Boullé**

*EDF R&D ICAME/SOAD, 1 avenue du Général de Gaulle, 92140 Clamart.
alexis.bondu@edf.fr,

**OrangeLabs, 2 avenue Pierre Marzin, 22300 Lannion.
prenom.nom@orange-ftgroup.com

Résumé. Dans cet article, une nouvelle stratégie d'apprentissage actif est proposée. Cette stratégie est fondée sur une méthode de discrétisation Bayésienne semi-supervisée. Des expériences comparatives sont menées sur des données unidimensionnelles, l'objectif étant d'estimer la position d'un échelon à partir de données bruitées.

1 Notations

Les données D sont composées de deux sous-ensembles T et U qui correspondent respectivement aux données étiquetées et non-étiquetées, avec $D = T \cup U$. L'ensemble T contient des couples (x, y) , où $x \in \mathbb{R}$ et $y \in \mathbb{Y}$ est une valeur discrète représentant la classe de l'exemple x . L'ensemble U contient des réels. Les notations suivantes sont adoptées : N , le nombre d'exemples observables ($N = |D|$); N^l , le nombre d'exemples étiquetés ($N^l = |T|$); J , le nombre de classes observées dans les données ($J = |\mathbb{Y}|$).

2 Discrétisation semi-supervisée Bayésienne

L'approche MODL discrétise les variables explicatives dans le but d'estimer les distributions conditionnelles aux classes. Le problème de la discrétisation d'une variable numérique est transposé en un problème de sélection de modèles. Un modèle de discrétisation $M(I, \{N_i\}, \{N_{ij}\})$ est défini par les paramètres suivants : i) I est le nombre d'intervalles; ii) $\{N_i\}$ est le nombre d'exemples dans chaque intervalle qui définit les bornes du modèle; iii) $\{N_{ij}\}$ est le nombre d'exemples de chaque classe dans chaque intervalle, qui définit les distributions conditionnelles localement à chaque intervalle. Une démarche Bayésienne maximisant $P(M|D)$ est appliquée pour sélectionner le meilleur modèle de discrétisation, noté \mathcal{M}_{map} (*Maximum a posteriori*). Cette démarche revient à maximiser $P(M)P(D|M)$. La distribution a priori des modèles $P(M)$ et la vraisemblance des données $P(D|M)$ sont calculées analytiquement en exploitant le caractère discret de la famille de modèles, et en adoptant des hypothèses faiblement informatives sur les données. Finalement, le \mathcal{M}_{map} minimise l'Equation 1 dont les deux termes correspondent à la distribution a priori des modèles et à la vraisemblance des données.