

Classification évidentielle avec contraintes d'étiquettes

Violaine Antoine*, Nicolas Labroche**

*Université Blaise Pascal, LIMOS, UMR CNRS 6158, BP 10125, 63000 Clermont-Ferrand,
violaine.antoine@univ-bpclermont.fr,

**Université François Rabelais de Tours, LI EA 6300, Campus de Blois, 41000 Blois,
nicolas.labroche@univ-tours.fr.

Résumé. Ce papier propose une version améliorée de l'algorithme de classification automatique évidentielle semi-supervisée SECM. Celui-ci bénéficie de l'introduction de données étiquetées pour améliorer la pertinence de ses résultats et utilise la théorie des fonctions de croyance afin de produire une partition crédale qui généralise notamment les concepts de partitions dures et floues. Le pendant de ce gain d'expressivité est une complexité qui est exponentielle avec le nombre de classes, ce qui impose en retour l'utilisation de schémas efficaces pour optimiser la fonction objectif. Nous proposons dans cet article une heuristique qui relâche la contrainte classique de positivité liée aux masses de croyances des méthodes évidentielles. Nous montrons sur un ensemble de jeux de données de test que notre méthode d'optimisation permet d'accélérer sensiblement l'algorithme SECM avec un schéma d'optimisation classique, tout en améliorant également la qualité de la fonction objectif.

1 Introduction

Ce papier propose un nouveau mécanisme d'optimisation pour l'algorithme de classification automatique évidentielle semi-supervisée SECM (Antoine et al., 2014), qui est le premier à reposer sur des contraintes exprimées sous la forme de données étiquetées. Les algorithmes de classification évidentielle (Masson et Denœux, 2008, 2009) reposent sur le cadre théorique des fonctions de croyance et permettent de représenter tous les types d'affectations partielles grâce au concept de partition crédale qui étend la notion de partition stricte, floue et possibiliste. Ces méthodes évidentielles ont été étendues dans le cadre semi-supervisé (Antoine et al., 2012, 2014) pour pouvoir tirer partie de contraintes de type Must-Link (ML) et Cannot-Link (CL) qui spécifient si deux données doivent ou non appartenir à la même classe. La transformation des informations disponibles a priori en ce type de contraintes peut néanmoins induire une perte de connaissance. L'algorithme SECM a été proposé récemment pour tirer partie de données partiellement étiquetées (Antoine et al., 2014). Cependant, l'algorithme SECM initial repose sur une optimisation stricte qui respecte l'ensemble des contraintes et notamment la positivité des masses de croyances associées à l'affectation d'un point à une classe. Cette contrainte théorique entraîne la formation d'un problème complexe. Nous proposons donc de modifier le mécanisme d'optimisation en relâchant la contrainte de positivité, à l'instar de ce