
Apprentissage de bonnes similarités dans un cadre conjoint semi-supervisé

Maria-Irina Nicolae^{*1,2}, Marc Sebban¹, Amaury Habrard¹, and Eric Gaussier²

¹Laboratoire Hubert Curien (LAHC) – CNRS : UMR5516, Université Jean Monnet - Saint-Etienne – 18 rue du Professeur LAuras 42000 SAINT-ETIENNE, France

²Laboratoire d'Informatique de Grenoble (LIG) – CNRS : UMR5217, Université Pierre-Mendès-France - Grenoble II, Institut polytechnique de Grenoble (Grenoble INP), Université Joseph Fourier - Grenoble I – UMR 5217 - Laboratoire LIG - 38041 Grenoble cedex 9 - France Tél. : +33 (0)4 76 51 43 61 - Fax : +33 (0)4 76 51 49 85, France

Résumé

Le rôle clé joué par les métriques en apprentissage automatique a généré ces dernières années un intérêt croissant pour l'optimisation de fonctions de distances et de similarités. Dans cet article, nous traitons le cas où peu de données étiquetées (et potentiellement également peu de données non-étiquetées) sont disponibles, situation qui apparaît fréquemment dans les applications réelles. Nous nous intéressons également aux garanties théoriques de la métrique apprise et du classifieur exploitant celle-ci. Notre approche s'inscrit dans le cadre théorique des fonctions de similarité (ϵ, γ, τ) -bonnes (Balcan et al., 2008) qui a été l'une des premières supervisées et régularisées. Nous fournissons une analyse théorique de cet apprentissage conjoint dérivant des bornes basées

Mots-Clés: apprentissage de similarités, similarités (epsilon gamma tau) bonnes, complexité de Rademacher

*Intervenant