

Partitionnement des données pour les problèmes de classement difficiles: Combinaison des cartes topologiques mixtes et SVM

Mustapha Lebbah*, Mohamed Ramzi Temanni*,**, Christine Poitou-Bernert**,***,****
Karine Clement**,***,****, Jean-Daniel Zucker*,**

* Université Paris 13, UFR de Santé,
Médecine et Biologie Humaine (SMBH) - Léonard de Vinci- LIM&BIO
74, rue Marcel Cachin 93017 Bobigny Cedex France
nom@limbio-paris13.org,
<http://www.limbio-paris13.org>

** Inserm, U755 Nutriomique, 75004 Paris, France;

*** University Pierre and Marie Curie-Paris 6, Faculty of Medicine,
Les Cordeliers, 75004 Paris, France;

**** AP-HP, Hôtel-Dieu Hospital, Nutrition department,
1 Place du parvis Notre-Dame, 75004 Paris, France
prénom.nom@htp.aphp.fr

Résumé. Dans ce papier, nous présentons un modèle pour aborder les problèmes de classement difficiles. Ces problèmes ont souvent la particularité d'avoir des taux d'erreurs en généralisations très élevés et ce quelles que soient les méthodes utilisées. Pour ce genre de problème, nous proposons d'utiliser un modèle de classement combinant le modèle de partitionnement des cartes topologiques mixtes et les machines à vecteur de supports (SVM). Le modèle non supervisé est dédié à la visualisation et au partitionnement des données composées de variables quantitatives et/ou qualitatives. Le deuxième modèle supervisé, est dédié au classement. Dans ce papier, nous présentons une combinaison de deux modèles qui permettent d'améliorer la visualisation des données et d'augmenter les performances en classement. Ce modèle consiste à entraîner des cartes auto-organisatrices pour construire une partition organisée des données, constituée de plusieurs sous-ensembles qui vont servir à reformuler le problème de classement initial en sous-problème de classement. Pour chaque sous-ensemble, on entraîne un classifieur SVM spécifique. Pour la validation de notre modèle (CT-SVM), nous avons utilisé quatre jeux de données. La première base est un extrait d'une grande base médicale sur l'étude de l'obésité à l'Hôpital Hôtel-Dieu de Paris, et les trois dernières bases sont issues de la littérature. Les résultats obtenus montrent l'apport de ce modèle dans la visualisation et le classement de données complexes.