Une application des Support Vector Machines à la neuro-imagerie

Alexandre Laurent[1], Pierre-Yves Hervé[2], Laure Zago[1], Marie Chavent[3], Jérôme Saracco[3], Philippe Boutinaud[2] et Marc Joliot[1]

[1] GIN, IMN, UMR5293, CNRS, Université de Bordeaux, CEA
[2] CADESIS, Lyon
[3] CQFD, Inria Bordeaux Sud Ouest, IMB, UMR 5251, Université de Bordeaux & Bordeaux INP

Dans le contexte de l'imagerie fonctionnelle cérébrale, les questions posées nécessitent désormais le traitement de masses de données qui s'apparentent à celles rencontrées en Big-Data. Pour avoir des temps d'analyse raisonnable, il est devenu indispensable d'utiliser les ressources de calcul massivement parallèles telles que celles installées au Mésocentre de Calcul Intensif Aquitain (MCIA).

Dans le cadre de ce projet, une méthode de classification supervisée par « Support Vector Machine » (SVM, Noble 2006) a été implémentée en langage R sur le MCIA. Une interface web a été développée avec pour objectif de rendre la procédure utilisable par des neuro- imageurs non spécialistes de l'informatique ou du traitement de données.

L'application de SVM consiste en deux phases, la phase d'entraînement et la phase de classification. La première phase, utilisant le plus de ressources, a pour objectif de produire le jeu de paramètres optimal (« classifier ») utilisé pour faire la prédiction ensuite. Elle utilise un processus itératif (RFE : « Recursive Feature Elimination ») pour sélectionner les variables (« features ») les plus informatives et une procédure de validation croisée (« Leave-One-Subject-Out, » LOSO) pour mesurer le taux d'erreur du « classifier ». La validation croisée est décomposable en N calculs indépendants, N étant le nombre de sujets utilisés à l'entraînement. Il en découle une procédure de calcul parallèle sur N cœurs. La phase de classification a été implémentée sur le même système, bien qu'elle ne sollicite pas d'importantes capacités de calcul.

Deux études ont été réalisées en imagerie de résonance magnétique fonctionnelle (IRMf). La première étude visait à prédire la préférence manuelle des sujets à partir des données d'IRMf de connectivité intrinsèque du repos. Chaque sujet était représenté par sa matrice de connectivité calculée sur l'atlas AICHA (Joliot et al., 2015), donnant lieu à plus de 73000 « features ». L'apprentissage effectué sur un ensemble de 231 sujets en 40 heures/cœur sur 231 cœurs a permis de sélectionner un « classifier » de 713 features avec un taux de réussite de l'ordre 69%. Le taux de réussite n'étant pas satisfaisant, des paramétrisations différentes de l'algorithme sont actuellement en cours de test. La deuxième étude visait à prédire la dominance hémisphérique cérébrale pour le langage. A la base de cartes de 60000 voxels d'activation dans une tâche de langage et un ensemble d'apprentissage de 250 sujets, le « classifier » optimal de 789 features a été calculé en 50 heures/cœur en mobilisant 250 cœurs. Les résultats de la procédure LOSO ont permis de calculer un score de 97% de réussite dans la prédiction de la dominance hémisphérique (Zago et al., 2014).

En conclusion, la procédure d'apprentissage utilisant le LOSO est une procédure générique pour ce type de problème. Nous envisageons de tester d'autres techniques, telles que les techniques d'apprentissage profond, dans le même environnement. L'application sera mise en ligne en 2017 sur la plateforme d'Imagerie en sciences du vivant (UMS 3767). Le projet, nommé TIMIC, a été soutenu financièrement par le labex CPU, dans le cadre de l'appel d'offres sur les projets émergents interdisciplinaires.

Noble, W. S. (2006). What is a support vector machine?. *Nature biotechnology*, *24*(12), 1565-1567.

Joliot, M., Jobard, G., Naveau, M., Delcroix, N., Petit, L., Zago, L., ... & Tzourio-Mazoyer, N. (2015). AICHA: An atlas of intrinsic connectivity of homotopic areas. *Journal of neuroscience methods*, *254*, 46-59.

Zago L, Hervé P, Genuer R, Jobard G, Crivello F, Petit L, Mellet E, Mazoyer B, Tzourio-Mazoyer N, Joliot M (2014) Hemispheric pattern of dominance: machine learning of sentence production in 297 healthy volunteers. In: OHBM. Hamburg, Germany.