

Analyse des données et imagerie cérébrale

Hervé Abdi

The University of Texas at Dallas

Les expériences d'imagerie cérébrale génèrent des ensembles de données qui posent des problèmes délicats d'analyse statistique parce que le nombre de variables (appelées voxels) dépasse très largement le nombre d'observations (e.g., quelques millions de voxels opposés à quelques milliers ou centaines de scans). Un problème fréquemment rencontré dans ce type d'expériences revient à prédire une variable expérimentale (e.g., ce que le sujet scanné regarde, le sujet scanné est-il un sujet à risque pour une la maladie d'Alzheimer, ...) à partir des données d'imagerie obtenues sur ce sujet. Dans l'approche statistique classique, ce style de problème s'analyse souvent avec le modèle linéaire général (e.g., ANOVA, MANOVA, Analyse Discriminante), mais les données d'imageries ne permettent pas cette approche (à cause de la multicollinéarité conséquente du nombre de variables supérieure au nombre d'observations). Récemment, les techniques de l'apprentissage automatique commencent à être utilisées pour analyser ces données. Les résultats de ces analyses se présentent souvent comme des cartes factorielles mais ne fournissent que rarement une étape inférentielle. Dans cet exposé, nous présentons une approche basée sur le bootstrap pour estimer des intervalles de confiance pour l'analyse des résultats d'imagerie.

Dans un premier temps nous exposons les problèmes généraux de l'imagerie cérébrale et de son traitement statistique. Puis, nous illustrons notre approche par un exemple dans lequel nous essayons de prédire la catégorie des images (e.g., visages, chats, chaussures, ...) que regarde un sujet à partir de ses données d'imagerie.