

Proposition de stage Master – 2013

Codage binaire de signaux EEG

RESPONSABLES: Michèle Sebag et Jamal Atif

LABORATOIRE: Laboratoire de Recherche en Informatique CNRS UMR 8623 & INRIA Futurs

ADRESSE: Université Paris-Sud, 91405 Orsay Cedex

MAIL: Michele.Sebag@lri.fr

URL: <http://www.lri.fr/~sebag/Stages/HEEG.pdf>

RÉSUMÉ :

Les signaux cérébraux impliqués dans les interfaces cerveaux-machine présentent plusieurs difficultés pénalisant les méthodes d'apprentissage statistique:

- L'étiquetage des données est bruité;
- Le fait que nous ne disposions pas d'une bonne distance (ou similarité) permettant de recouvrir la variété de faible dimension dans laquelle vivent les données;
- L'insuffisance des données étiquetées compte tenu de la dimensionalité élevée de leur description.

Objectif du stage

L'objectif est de proposer un codage i) induisant une pseudo-distance pertinente; ii) permettant l'appariement approché rapide entre le signal courant et les données disponibles de l'interface.

Approche

L'approche proposée s'inspire du codage binaire présenté par Norouzi et al., NIPS 2012, qui correspond à raffiner le codage LSH (locality sensitive hashing) dans un contexte supervisé. Dans cette approche, les signaux (images) sont étiquetés et chaque feature est apprise comme résultat de la discrimination d'un triplet d'exemples.

Une difficulté majeure est identifiée (outre la nature non iid des signaux disponibles dans le contexte BCI). En effet, l'apprentissage de feature ci-dessus repose sur l'existence de labels définis et variant dans un nombre important de classes. Dans le contexte BCI en revanche, le nombre de classes est faible (souvent binaire) et les labels sont connus pour être bruités.

L'extension proposée s'appuiera sur les connaissances a priori du domaine (notamment relatives à la continuité spatio-temporelle de l'activité cérébrale) pour identifier des triplets d'exemples appropriés pour la construction de features.

De très bonnes capacités algorithmiques sont nécessaires ainsi que le gout de la conceptualisation et un esprit d'observation approfondi. Des données EEG, étiquetées ou resting-state, sont disponibles.

Pointeurs

- Hamming Distance Metric Learning, Mohammad Norouzi, David Fleet, Ruslan Salakhutdinov, Neural Information Processing Systems (NIPS), 2012.
- Approximate nearest neighbors: towards removing the curse of dimensionality. P. Indyk and R. Motwani, STOC, 1998.