

Proposition de stage M2 de 6 mois

Classification géométrique robuste d'électroencéphalogrammes : une approche semi-paramétrique

1 Contexte scientifique

L'électroencéphalographie (EEG) est une modalité de neuroimagerie non-invasive qui consiste à enregistrer l'activité électrique du cerveau à l'aide d'électrodes placées sur le scalp. Son faible coût, sa simplicité et sa bonne résolution temporelle ont fait la popularité de cette technique. Elle est notamment primordiale pour les interfaces cerveau-ordinateur (*brain computer interfaces*; BCI), où le sujet interagit avec un ordinateur grâce à ses signaux cérébraux. Les BCI peuvent par exemple être employées pour le contrôle d'exosquelettes [Kal+16] ou la respiration artificielle [Che+18a]. Cependant, l'EEG souffre d'un faible rapport signal à bruit (*signal to noise ratio*; SNR) et d'une résolution spatiale limitée. En effet, les signaux électriques du cerveau se mélangent en traversant les tissus cérébraux, le crâne et le scalp. De plus, les électrodes enregistrent également des perturbations environnementales (e.g., réseau électrique) et biologiques (e.g., mouvements oculaires, battements de coeur).

Le traitement, la classification et le partitionnement de signaux EEG a attiré beaucoup d'attention ces dernières années et reste un des défis clés de la recherche à ce jour. L'état de l'art s'est principalement concentré sur l'emploi de matrices de covariance empiriques (*sample covariance matrices*; SCM) des signaux et les méthodes développées se sont tournées vers l'exploitation de leur géométrie riemannienne intrinsèque. Par exemple, pour la classification, le classifieur de distance minimal à la moyenne (*minimum distance to mean*; MDM) se base sur le calcul des centres de masses de chacune des classes puis détermine la classe d'une SCM inconnue en choisissant celle qui correspond à la moyenne la plus proche [Bar+11; Lot+18]. D'autres classifieurs commencent par projeter l'ensemble des matrices de covariance dans l'espace tangent d'un point de référence et un classifieur euclidien usuel peut ensuite être employé.

Bien que les méthodes géométriques se soient montrées efficaces, les méthodes de classification et de partitionnement pour l'EEG souffrent encore de limitations importantes. En particulier, la plupart des méthodes exploitent des SCM et se reposent sur une hypothèse de gaussianité des signaux. Cependant, du fait de leur nature biologique, les signaux EEG ont une grande variabilité, contiennent systématiquement des données aberrantes et sont souvent limités en quantité. De ce fait, on peut s'attendre à ce que la distribution des données acquises soit en fait à queue lourde et les méthodes existantes peuvent bénéficier de l'exploitation de la théorie des statistiques robustes; voir e.g. [Mar+19]. Dans l'état actuel, les méthodes existantes sont très dépendantes de prétraitements spécifiques et complexes; voir e.g. [Che+18b]. En conséquence, ces méthodes demandent des connaissances multidisciplinaires poussées pour être employées correctement et sont donc réservées à un nombre limité de spécialistes. Développer des méthodes avec un nombre limité et automatisés de prétraitements est donc un enjeu important.

2 Objectifs du stage

Du point de vue méthodologique, l'objectif principal de ce stage est de développer des méthodes de classification et de partitionnement des signaux EEG qui exploitent la géométrie riemannienne et les statistiques robustes. Au lieu de considérer que les observations suivent la distribution gaussienne multivariée centrée, nous supposons qu'elles obéissent à une distribution plus générale dans la famille des lois elliptiques centrées [Oll+12]. En pratique, on ne connaît pas la loi exacte que les données suivent et, pour obtenir une matrice de covariance robuste, la solution est d'employer un M -estimateur comme celui de Tyler [Ty187]. Une première approche consiste donc à adapter et étendre les méthodes existantes afin d'exploiter de tels estimateurs pour l'EEG. Malheureusement, d'un point de vue théorique, les M -estimateurs ont le désavantage de ne pas être statistiquement efficaces. Pour dépasser cette limitation, nous pouvons exploiter une approche semi-paramétrique [Bic+93]. Dans ce contexte, la

famille des R -estimateurs a prouvé sa capacité à réconcilier les concepts de robustesse et d'efficacité (semi-paramétrique) [HOP06; FRP20]. Nous exploiterons donc ces estimateurs pour développer des méthodes de classification spécifiques au contexte de l'EEG. En particulier, nous pourrions partir des travaux de [HP02] sur l'extension de la distance de Mahalanobis dans le cas semi-paramétrique pour obtenir un nouveau classifieur, tout en gardant en tête l'importance de la géométrie dans les performances des classifieurs en EEG.

Les différentes méthodes développées dans le cadre du stage feront l'objet d'une validation sur des scénarios EEG avec l'utilisation de données réelles et simulées. L'un des objectifs du projet consiste à proposer des méthodes de traitement et de classification utilisables par les praticiens de la communauté EEG. En ce sens, les données choisies pour faire cette étape de validation seront sélectionnées au sein de ceux librement disponibles dans la librairie MOABB [JB18]. De plus, un effort particulier sera mis en oeuvre pour rendre accessible les méthodes développées dans le cadre du stage en intégrant celles-ci dans la librairie de traitement déjà existante. Cela garantira également la reproductibilité des résultats obtenus avec la réalisation de benchmarks systématiques se reposant sur des pipelines de traitement du signal. En effet, la librairie MOABB adhère aux principes FAIR¹ plébiscités par la communauté scientifique dans le développement de logiciels à des fins scientifiques.

3 Encadrement

Le stage se déroulera au sein du laboratoire L2S, situé dans les locaux de CentraleSupélec sur le plateau de Saclay. La durée sera de 6 mois.

L'encadrement sera effectué en collaboration avec le LISTIC, à Annecy. Le stagiaire sera donc encadré par :

- Florent Bouchard, Chargé de recherche au CNRS, L2S
- Stefano Fortunati, Enseignant chercheur à l'IPSA, L2S
- Ammar Mian, Maître de conférences à l'Université Savoie Mont Blanc, LISTIC

4 Profil et démarche de candidature

Compétences

- Master recherche ou ingénieur avec une expérience en apprentissage statistique
- Intérêt pour les mathématiques appliquées (statistiques, algèbre linéaire)
- Compétences : Statistiques, apprentissage automatique, programmation python

Candidater

Envoyer un mail d'intérêt avec le CV aux adresses suivantes :

- florent.bouchard@centralesupelec.fr
- stefano.fortunati@centralesupelec.fr
- ammar.mian@univ-smb.fr

Références

- [Bar+11] A. BARACHANT et al. "Multiclass brain-computer interface classification by Riemannian geometry". In : *IEEE Transactions on Biomedical Engineering* 59.4 (2011), p. 920-928.
- [Bic+93] P.J. BICKEL et al. *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, 1993.
- [Che+18a] S. CHEVALLIER et al. "Brain-Machine Interface for Mechanical Ventilation Using Respiratory-Related Evoked Potential". In : *International Conference on Artificial Neural Networks (ICANN)*. Rhodes, Greece, 2018.

1. Findability, Accessibility, Interoperability, and Reuse

- [Che+18b] S. CHEVALLIER et al. “Riemannian classification for SSVEP based BCI : offline versus online implementations”. In : *BCI Handbook : Technological and Theoretical Advances*. CRC Press, 2018.
- [FRP20] S. FORTUNATI, A. RENAUX et F. PASCAL. “Robust Semiparametric Efficient Estimators in Complex Elliptically Symmetric Distributions”. In : *IEEE Transactions on Signal Processing* 68 (2020), p. 5003-5015.
- [HOP06] M. HALLIN, H. OJA et D. PAINDAVEINE. “Semiparametrically Efficient Rank-Based Inference for Shape II. Optimal R -Estimation of Shape”. In : *The Annals of Statistics* 34.6 (2006), p. 2757-2789.
- [HP02] M. HALLIN et D. PAINDAVEINE. “Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks”. In : *The Annals of Statistics* 30.4 (2002), p. 1103-1133.
- [JB18] V. JAYARAM et A. BARACHANT. “MOABB : trustworthy algorithm benchmarking for BCIs”. In : *Journal of neural engineering* 15.6 (2018), p. 066011.
- [Kal+16] E.K. KALUNGA et al. “Online SSVEP-based BCI using Riemannian geometry”. In : *Neurocomputing* 191 (2016), p. 55-68.
- [Lot+18] F. LOTTE et al. “A review of classification algorithms for EEG-based brain–computer interfaces : a 10 year update”. In : *Journal of neural engineering* 15.3 (2018).
- [Mar+19] R. A. MARONNA et al. *Robust statistics : theory and methods (with R)*. John Wiley & Sons, 2019.
- [Oll+12] E. OLLILA et al. “Complex elliptically symmetric distributions : Survey, new results and applications”. In : *IEEE Transactions on Signal Processing* 60.11 (2012), p. 5597-5625.
- [Tyl87] D. E. TYLER. “A distribution-free M-estimator of multivariate scatter”. In : *The Annals of Statistics* (1987), p. 234-251.