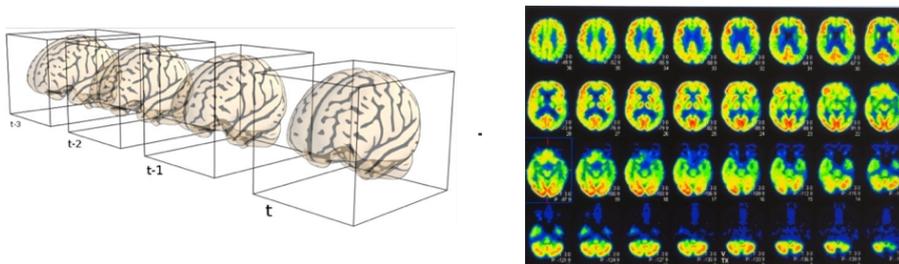


Apprentissage sur des données tensorielles de rang faible : application au clustering des données d'imagerie magnétique fonctionnelle (IRMf)

Encadrants : Ricardo Borsoi (ricardo-augusto.borsoi@cnrs.fr),
Sebastian Miron (sebastian.miron@univ-lorraine.fr), David Brie (brie5@univ-lorraine.fr)
-CRAN, UMR 7039, Nancy-



Mots clés : analyse de données, traitement du signal, apprentissage automatique, tenseurs, rang faible, clustering, variétés grassmanniennes

Contexte : Les tenseurs (tableaux à plusieurs indices) sont un outil important pour modéliser les relations complexes entre les échantillons de données multivariées naturellement présentes dans divers domaines tels que les neurosciences, les sciences de l'environnement, les médias sociaux et la météorologie. Néanmoins, le traitement conjoint des ensembles de données tensorielles pose le défi d'extraire des caractéristiques pertinentes pour la classification ou la fusion de données. Dans ce contexte, l'enjeu majeur réside dans la capacité des algorithmes à résumer les informations d'intérêt tout en préservant les caractéristiques liées à la variabilité individuelle propre à chaque jeu de données [1,2].

Objectifs du stage : Le présent stage vise à étudier l'utilisation de différentes représentations tensorielles de rang faible (Parafac, Tucker, etc.), pour réaliser des tâches de classification sur des ensembles de données tensorielles. L'objectif est de comparer les performances des méthodes de clustering naïves, qui utilisent la distance euclidienne directement dans l'espace des tenseurs, avec des approches prenant en compte la géométrie de rang faible des données tensorielles (distance sur des variétés ou sur les facteurs de la décomposition). Pour les représentations de rang faible, des métriques telles que les distances grassmanniennes seront utilisées pour la classification, afin de résoudre les ambiguïtés inhérentes aux décompositions tensorielles [3].

La principale application envisagée est l'imagerie par résonance magnétique fonctionnelle (IRMf), une méthode couramment utilisée pour étudier le cerveau humain. L'objectif de l'analyse des données IRMf est de localiser les zones du cerveau activées lors de l'exécution d'une tâche spécifique. Cela permet la création de cartes d'activation cérébrale utilisées par les neuroscientifiques pour mieux comprendre le fonctionnement du cerveau [4]. Les données IRMf présentent une structure intrinsèquement tensorielle avec des diversités spatiales et temporelles. Les approches tensorielles de clustering seront appliquées aux données IRMf de cohortes de patients pour étudier et prédire des pathologies neurologiques.

Ce stage s'inscrit dans le cadre d'un projet de recherche franco-américain, impliquant des réunions de travail avec des chercheurs de University of Maryland Baltimore County (UMBC). A la fin du stage, l'étudiant.e recruté.e aura l'opportunité de poursuivre ses travaux de recherche dans le cadre d'une thèse de doctorat en partenariat avec l'UMBC.

[1] E. Acar et al., "Structure-revealing data fusion," *BMC Bioinformatics*, vol. 15, no. 1, pp. 1-17, 2014.

[2] A. K. Smilde et al., "Common and distinct components in data fusion," *Journal of Chemometrics*, vol. 31, no. 7, p. e2900, 2017.

[3] R. A. Borsoi et al., "Coupled tensor decomposition for hyperspectral and multispectral image fusion with inter-image variability," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 702-717, 2021.

[4] M. Akhonda et al., "Disjoint subspaces for common and distinct component analysis: Application to the fusion of multi-task fMRI data," *Journal of Neuroscience Methods*, vol. 358, p. 109214, 2021.

Profil et aptitudes souhaités :

- Étudiant.e en MASTER deuxième année (niveau bac + 5)
- Connaissances traitement du signal et/ou algèbre linéaire
- Connaissances en programmation et maîtrise des logiciels de calcul numérique (e.g., Matlab)
- Un bon niveau en anglais sera un atout

Conditions du stage :

- Stage de 6 mois à pourvoir sur une période février à septembre 2024
- Rémunération : environ 550 euros /mois (la prise en charge des frais d'hébergement sur la durée du stage est envisageable)
- Le stagiaire sera basé au CRAN (Centre de Recherche en Automatique de Nancy), Vandœuvre-lès-Nancy (54)