Ph.D. thesis in signal processing & machine learning (2024–2027)
CRAN (UL – France) / MLSP Lab (UMBC – USA)

Tensor decomposition methods for data fusion: Application to multi-subject fMRI data

Context: Given the recent explosion in the amount of data from multiple modalities, data fusion has been growing in importance for multiple applications [1]. A fundamental problem is fusing heterogeneous datasets containing dataset-specific information [4], such as in multi-task functional magnetic resonance imaging (fMRI) data [2] or multimodal image fusion [3]. Moreover, many datasets of interest have three or more dimensions, such as spatial and temporal, which poses a challenge to classical data processing methods.

Challenges: With the growing emphasis of data-driven methods approaches based on matrix and tensor decompositions, which are directly interpretable, the field is transforming. However, the theoretical study of more flexible factorizations is still in its infancy. The uniqueness of coupled tensor decompositions was only investigated recently [7]. More flexible decompositions have been proposed [8], but existing uniqueness results are still incipient [3]. Despite considerable advances in machine leaning methods for data fusion and clustering in recent years, current methods are limited in their capacity to jointly account for information that is common and specific to different subgroups of datasets. Addressing such challenges is of fundamental importance when developing neuroimaging data analysis methods.

Research program: The Ph.D. candidate will develop new flexible matrix/tensor decomposition methods that can unravel the common and specific information of the different datasets. This can lead to discovering subgroups of datasets which share common features. A key aspect of the develop methods will be their interpretability and the theoretical guarantees, for which the use of tensor decompositions provides an adequate mathematical framework [7]. This will involve important methodological challenges, such as the development of adequate metrics to measure the similarity between the different multi-dimensional datasets (such as distances in Riemannian manifolds), solving large-scale optimization problems, and linking clustering and tensor decompositions into an unified unsupervised learning framework. The developed methods will be applied for analyzing multi-subject fMRI data for personalized medicine, finding subject-specific features that are predictive of mental disorders or which can characterize subtypes of diseases and risk groups for addictive behavior [2,5]. Of particular interest is the longitudinal ABCD study [6], which collects fMRI and non-neuroimaging data (e.g., cognitive scores, substance use) of the same subjects over time.

Supervision and environment: The candidate will be jointly supervised by Prof. Sebastian Miron, Dr. Ricardo Borsoi, and Prof. David Brie, members of the Multidimensional Signal Processing (SiMul) team (`https://cran-simul.github.io/`), CRAN Laboratory, University of Lorraine, France, and by Prof. Tülay Adali, director of the Machine Learning for Signal Processing (MLSP) Laboratory (`https://mlsp.umbc.edu/`), University of Maryland Baltimore County (UMBC), USA. He/She will be primarily based in the CRAN Laboratory, University of Lorraine, in Vandoeuvre-lès-Nancy, France, with the possibility for research visits to the MLSP lab in Baltimore, USA.

Salary and Duration: Approximately 2100 euros/month. The thesis has a duration of 36 months, starting in 2024.

Expected profile: Master's degree or equivalent, with experience in one or more of the following topics: data analysis, signal processing, machine learning, applied mathematics. Good communication skills in English (written and oral). Candidates should send their application to sebastian.miron@univ-lorraine.fr, ricardo.borsoi@univ-lorraine.fr, david.brie@univ-lorraine.fr, adali@umbc.edu, including an academic CV and a motivation letter (1 page max.) explaining their research interests and their motivation for this position.

References

[1] D. Lahat et al., “Multimodal data fusion: an overview of methods, challenges, and prospects,” Proceedings of the IEEE, vol. 103, no. 9, pp. 1449–1477, 2015.

[2] 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.

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

[5] E. S. Finn et al., “Functional connectome fingerprinting: identifying individuals using pat- terns of brain connectivity,” Nature Neuroscience, vol. 18, no. 11, pp. 1664-1671, 2015.

[6] B. J. Casey et al., “The adolescent brain cognitive development (ABCD) study: imaging acquisition across 21 sites,” Developmental Cognitive Neuroscience, vol. 32, pp. 43-54, 2018.

[7] M. Sørensen and L. D. De Lathauwer, “Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank-$(L_{r,n}, L_{r,n}, 1)$ terms-part I: Uniqueness,” SIAM Journal on Matrix Analysis and Applications, vol. 36, no. 2, pp. 496-522, 2015.

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