Postdoc/Ph.D. position in signal processing & machine learning (2024–2027)
CRAN (UL – France)

Statistical and Tensor Methods for Spatiotemporal Heterogeneous Data Analysis

Context: Information gathered across both spatial and temporal dimensions to characterize a phenomenon at specific locations and time points is commonly referred to as spatiotemporal data. Numerous real-world processes inherently exhibit spatiotemporal characteristics, spanning domains such as neuroscience, epidemiology, climate science and pollution monitoring.

The development of efficient data analysis tools capable of extracting pertinent features from spatiotemporal datasets is crucial for tasks including classification and regression across a wide range of applications. These tools must account for the high-dimensional and potentially heterogeneous nature of the data, encompassing not only spatial and temporal dimensions, but also additional modalities such as, e.g., longitudinal acquisitions.

For instance, in neuroscience applications, the data often presents various neuroimaging modalities (such as fMRI, MRI, EEG) as well as non-neuroimaging data such as demographic information (e.g., age, gender) of behavioral factors (e.g., substance use) [6]. Therefore, robust methodologies need to be developed to effectively handle such complex and diverse datasets and extract meaningful insights [1].

Challenges: Developing feature extraction methods that can effectively and jointly handle data from diverse modalities poses a significant challenge. A particular difficulty is to devise flexible models which are directly interpretable, readily providing insight into the relationships that are learned from the data [2]. This is a critical aspect in fields including medical applications.

Since spatiotemporal datasets are naturally represented as matrices and higher-order tensors, low-rank decompositions have become an invaluable for extracting meaningful representations. These methods not only offer a strong mathematical foundation for understanding their underlying mechanisms but also facilitate the direct interpretability of the learned representations [4]. However, unlike approaches based on statistical models, most matrix and tensor decomposition methods are fundamentally algebraic and do not directly exploit statistical properties of the data. Moreover, they do not inherently provide a means to quantify the uncertainty associated with the results, which is another crucial requirement in health-related applications.

Thus, developing representation learning in the form of decomposition methods that leverage the strengths of both statistical and algebraic methods is an important task. Specific challenges include handling heterogeneous (e.g., both continuous and categorical) data types, while ensuring interpretability, computational efficiency, and robust theoretical guarantees (such as reproducibility, identifiability, and consistency) [3].

Research program: The candidate will develop flexible representations learning and data analysis tools specifically designed to handle heterogeneous spatiotemporal data. These methods should effectively utilize both algebraic (matrix/tensor) and statistical frameworks to generate results that are not only interpretable but also backed by statistical guarantees. A key focus will be exploring the connections and interactions between the tensor decomposition framework (e.g., PARAFAC2 [5]) and the statistical source separation framework (e.g., Independent Vector Analysis [7]).

The developed approaches must efficiently accommodate various types of data, including spatiotemporal datasets (fMRI acquisitions from longitudinal studies with multiple subjects) and categorical data (socioeconomic status, substance use). They should be capable of extracting meaningful features which can be used for tasks such as clustering and prediction.

The proposed methods will be applied to personalized medicine focusing on the adolescent brain cognitive development (ABDC) dataset [6], with the aim to unveil the latent structure of adolescent brain development and to elucidate the interplay between neuroimaging data (e.g., fMRI) and cognitive/socioeconomic factors as well as their temporal evolution.

Supervision and environment: Depending on the candidate's profile, this research work will be carried out as part of a (2-3 years) post-doctorate or a Ph.D. thesis starting in 2024. 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. This research will be conducted in collaboration with Prof. Tülay Adali, head of the Machine Learning for Signal Processing (MLSP) Laboratory (`https://mlsp.umbc.edu/`), University of Maryland Baltimore County (UMBC), USA. He/She will be 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 funding: The future researcher will be funded by the NSF-ANR grant AGDAM (ANR-23-CE94-0001). The salary is approximately 2100 euros per month for the doctoral position, and approximately 3000 euros per month (depending on the research experience) for a post-doctoral position.

Expected profile:
- For the Ph.D. position: Master degree or equivalent, with experience in one or more of the following topics: data analysis, signal processing, machine learning, applied mathematics.
- For the post-doctoral position: Ph.D. degree in electrical engineering, applied mathematics or related fields.
Strong mathematical background
- 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`, 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] J. W. Murdoch et al. "Interpretable machine learning: definitions, methods, and applications." arXiv preprint arXiv:1901.04592, 2019.

[3] T. Adali et al. "Reproducibility in Matrix and Tensor Decompositions: Focus on model match, interpretability, and uniqueness." IEEE Signal Processing Magazine, vol. 39, no. 4, pp. 8-24, 2022.

[4] S. Miron et al. "Tensor methods for multisensor signal processing." IET signal processing, vol. 14, no. 10, pp.693-709, 2020.

[5] H. Kiers et al. "PARAFAC2—Part I. A direct fitting algorithm for the PARAFAC2 model." Journal of Chemometrics: A Journal of the Chemometrics Society, vol. 13., no. 3-4, pp. 275-294, 1999.

[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] T. Adali et al.,"Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging," IEEE Signal Processing Magazine, vol. 31, no. 3, pp. 18–33, 2014.