IFP Energies nouvelles, internship: Analysis and prediction of transform performance for clustering and classification (Spring/Summer 2018)

1 Topic

Sparse-domain data (signal/image) processing for clustering, classification and learning: basis/frame influence and selection for *Functional Data Analysis* and *Scattering networks* with experimental datasets

2 City and country

Rueil-Malmaison (Paris suburbs), France

3 Contact names and mails

- Control, signal and systems department: Jean Charléty, Aurélie Pirayre
- Advisor: Laurent Duval; laurent.duval[ad]ifpen.fr
- Head of the department: Yann Creff; yann.creff[ad]ifpen.fr

4 General presentation of the topic

We wish to study large datasets of experimental data (e.g. spectrum or physico-chemical signals, microscopy and subsurface images) toward clustering, classification and learning. When data satisfy regularity properties, they often admit sparse or compressible representations in a judicious transformed domain: a few transformed coefficients provide accurate data approximation. Such representations, like multiscale or wavelet transforms, are beneficial to subsequent processing, and they form the core of novel methodologies. The two recent processing methodologies, under study here, are primarily:

- *Functional Data Analysis (FDA) or Regression* [RS05, RHG09];
- *Scattering networks (SN) or transforms* [Mal12, BM13, Mal16].

Due to the variety of such transforms [JDCP11, AAG14], without prior knowledge, it is not evident to find the most suitable representation for a given set of data. The aim of this subject is to investigate potential relations between transform properties and data sparsity in the one hand, and classification/clustering performance on the other hand, especially with respect to the robustness to shifts/translations or noise in data features.

5 Objective of the internship

The first objective is to develop a framework to allow the use of different sparsifying transformations (bases or frames of wavelets and multiscale transformations) at the input of reference FDA and SN [Mal16] algorithms. This will permit to evaluate the latter on a variety of experimental datasets, with the aim of choosing the most appropriate, both in terms of performance and of usability, since the redundancy in transformations may hinder their application to large datasets. A particular interest could be laid on complex-like transformations, that may improve either the sparsification or "invariance properties" [BM13, BG17] in the transformed data. Their importance has been underlined recently for deep convolutional networks [TBC$^+$16, TBS$^+$17].

Then, starting from real data, the trainee will develop realistic models reproducing the expected behaviors in the data, for instance related to shifts or noise.

Finally, the relative clustering/classification performances will be assessed with respect to different transformation choices [DBMdH17], and their impact on both realistic models and real data. A particular interest could be laid on either transform properties (redundancy, frame bounds, asymptotic properties) or the resulting data multiscale statistics or sparsity [HR09, Wal17, DBMdH17, CL17].

6 Expected ability of the student

Second/third year engineering school and/or master of science with strong skills and curiosity in signal/image processing, statistics, machine learning, applied mathematics. Applicants should provide a resume and a motivation letter emphasizing prior knowledge related to the subject.

Information updates

`http://www.laurent-duval.eu/lcd-2018-intern-transform-classification.html`

References

[AAG14] S. T. Ali, J.-P. Antoine, and J.-P. Gazeau. *Multidimensional Wavelets and Generalizations*, pages 411–456. Theoretical and Mathematical Physics. Springer-Verlag, Aug. 2014.

[BG17] Randall Balestriero and Hervé Glotin. Linear time complexity deep Fourier scattering network and extension to nonlinear invariants. In *PREPRINT*, 2017. Under review.

[BM13] J. Bruna and S. Mallat. Invariant scattering convolution networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(8):1872–1886, Aug. 2013.

[CL17] Wojciech Czaja and Weilin Li. Analysis of time-frequency scattering transforms. *Appl. Comput. Harmon. Analysis*, 2017.

[DBMdH17] I. Dokmanić, J. Bruna, S. Mallat, and M. de Hoop. Invariant multiscale statistics for inverse problems. In *Proc. Signal Processing with Adaptive Sparse Structured Representations (SPARS) workshop*, Lisbon, Portugal, Jun. 5-8, 2017.

[HR09] N. Hurley and S. Rickard. Comparing measures of sparsity. *IEEE Trans. Inform. Theory*, 55(10):4723–4741, Oct. 2009.

[JDCP11] L. Jacques, L. Duval, C. Chaux, and G. Peyré. A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity. *Signal Process.*, 91(12):2699–2730, Dec. 2011.

[Mal12] S. Mallat. Group invariant scattering. *Comm. Pure Appl. Math.*, 65(10):1331–1398, Oct. 2012.

[Mal16] S. Mallat. Understanding deep convolutional networks. *Phil. Trans. R. Soc. A*, 374(2065):20150203, mar 2016.

[RHG09] J. O. Ramsay, G. Hooker, and S. Graves. *Functional Data Analysis with R and MATLAB*. 2009.

[RS05] J. O. Ramsay and B. W. Silverman. *Functional data analysis*. Springer Series in Statistics. Springer, 2nd edition, 2005.

[TBC+16] M. Tygert, J. Bruna, S. Chintala, Y. LeCun, S. Piantino, and .r Szlam. A mathematical motivation for complex-valued convolutional networks. *Neural Comput.*, 28(5):815–825, May 2016.

[TBS+17] C. Trabelsi, O. Bilaniuk, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal. Deep complex networks. *PREPRINT*, 2017.

[Wal17] I. Waldspurger. Exponential decay of scattering coefficients. In *Proc. Int. Conf. Sampling Theory Appl.*, Tallin, Estonia, Jul. 3-7, 2017. IEEE.