Data Series Analytics and Deep Learning for Gravitational Wave Glitch Detection

Themis Palpanas
Senior Member of Institut Universitaire de France (IUF)
University of Paris
themis.palpanas@u-paris.fr

Kostas Zoumpatianos
Harvard University
kostas@seas.harvard.edu

Decades of progress in experimental ingenuity were radiantly crowned with the first detection of gravity waves by the VIRGO/LIGO consortium in 2015 (Abbott et al. 2016), confirming one of the most remarkable predictions of General Relativity and probing the extreme conditions around black holes. Gravitational astronomy has thus become a powerful new tool, promising fundamental discoveries in astrophysics and basic gravity research.

Gravity wave signals are also exceedingly weak, requiring the measurement of displacements smaller than atomic nuclei over distances of kilometers. This is achieved by interfering light waves traveling along different arms of an interferometer buried underground. These highly sensitive detectors must weed out a vast host of parasitic signals caused by environmental and instrumental effects, some whose origin remains poorly understood [1].

We will explore machine learning and deep learning techniques to help mitigate two types of parasitic signals in the VIRGO/LIGO time streams. Glitches are instrumentally generated signals of poorly understood origin, and hence difficult to model a priori. Machine learning is well suited to the detection and classification of glitches because a neural network can learn the features that distinguish them from astrophysical signals through training on known events. Environmental effects (vibrations, etc.) generate the second type of parasitic signal that we will study. The gravity wave detector incorporates a myriad other sensors to monitor the environment. A possible direction would be to combine data streams from these sensors with the science data stream to look for complex correlations representing identifying features in the very large dimensional data space.

In this internship, we will study the application of time series processing techniques for VIRGO/LIGO time streams. The advantage of this approach is that it takes into account the trends that these data exhibit. Therefore, we can identify time series that follow similar trends over time and subsequently group those together in clusters, or detect abnormal behavior [2]. In this context, nearest neighbor queries are of paramount importance, since they form the basis of virtually every data mining, or other complex analysis task involving time series. In order to provide a time-efficient solution, we will use the state of the art time series indexes [3], which minimize both the pre-processing, as well as the query-answering time. Moreover, we will study various deep learning techniques suitable for the solution of these problems, as well as the interplay between the data series analysis and the deep learning techniques (e.g., the former may provide training examples for the latter).

This work will be performed within an international collaboration with Harvard University.

[Prerequisites] Experience machine learning and deep learning, excellent programming skills in C/C++, and Python.

[Internship] Apply by emailing Prof. Themis Palpanas your detailed CV (including course marks). Accepting this project will make you part of diNo (LIPADE, Paris Descartes University), an enthusiastic team working on real, challenging problems! The internship will last between 3-6 months, and is fully funded.

References

[1] Sara Bahaadini, Vahid Noroozi, Neda Rohani, Scott Coughlin, Michael Zevin, J. R. Smith, Vicky Kalogera, Aggelos K. Katsaggelos: Machine learning for Gravity Spy: Glitch classification and dataset. Inf. Sci. 444: 172-186 (2018).

[2] Michele Linardi, Yan Zhu, Themis Palpanas, Eamonn J. Keogh: VALMOD: A Suite for Easy and Exact Detection of Variable Length Motifs in Data Series. SIGMOD Conference 2018: 1757-1760

[3] Botao Peng, Panagiota Fatourou, Themis Palpanas: ParIS: The Next Destination for Fast Data Series Indexing and Query Answering. BigData 2018: 791-800