Commun. Comput. Phys.
doi: 10.4208/cicp.OA-2018-0092

Vol. **25**, No. 4, pp. 963-987
April 2019

Finding the Origin of Noise Transients in LIGO Data with Machine Learning

Marco Cavaglià[1,*], Kai Staats[2] and Teerth Gill[2]

[1] *Department of Physics and Astronomy, The University of Mississippi, University MS 38677-1848, USA.*
[2] *Department of Physics and Astronomy, Embry-Riddle University, Prescott AZ 86301, USA.*

Received 4 April 2018; Accepted (in revised version) 24 July 2018

Abstract. Quality improvement of interferometric data collected by gravitational-wave detectors such as Advanced LIGO and Virgo is mission critical for the success of gravitational-wave astrophysics. Gravitational-wave detectors are sensitive to a variety of disturbances of non-astrophysical origin with characteristic frequencies in the instrument band of sensitivity. Removing non-astrophysical artifacts that corrupt the data stream is crucial for increasing the number and statistical significance of gravitational-wave detections and enabling refined astrophysical interpretations of the data. Machine learning has proved to be a powerful tool for analysis of massive quantities of complex data in astronomy and related fields of study. We present two machine learning methods, based on random forest and genetic programming algorithms, that can be used to determine the origin of non-astrophysical transients in the LIGO detectors. We use two classes of transients with known instrumental origin that were identified during the first observing run of Advanced LIGO to show that the algorithms can successfully identify the origin of non-astrophysical transients in real interferometric data and thus assist in the mitigation of instrumental and environmental disturbances in gravitational-wave searches. While the datasets described in this paper are specific to LIGO, and the exact procedures employed were unique to the same, the random forest and genetic programming code bases and means by which they were applied as a dual machine learning approach are completely portable to any number of instruments in which noise is believed to be generated through mechanical couplings, the source of which is not yet discovered.

AMS subject classifications: 68T05, 83-04, 83C35, 68T20

Key words: Machine learning, gravitational waves, noise mitigation.

*Corresponding author. *Email addresses:* `marco.cavaglia@ligo.org` (M. Cavaglià), `kai.staats@ligo.org` (K. Staats), `teerth.gill@ligo.org` (T. Gill)

1 Introduction

On February 11$^{\text{th}}$, 2016, scientists from the Laser Interferometer Gravitational-wave Observatory (LIGO) [1] Scientific Collaboration (LSC) and the European Virgo Collaboration [2] announced the first direct detection of gravitational waves from a coalescing pair of two stellar-mass black holes [3]. Detection of the GW150914 gravitational-wave signal, recorded at the LIGO sites in the morning of September 14$^{\text{th}}$ 2015, marks the beginning of a new observational era in astrophysics. Strong, dynamical relativistic gravitational fields can now be used to map the dark universe and probe fundamental physics. One hundred years after the formulation of general relativity, fifty years after the pioneering work of Joseph Weber, and several decades after the foundation of LIGO, gravitational-wave astrophysics is a reality.

The next decade will see this new branch of scientific research expand to a mature field [4]. Since GW150914, another four gravitational-wave detections from binary black hole systems [5–8] and a detection from a binary neutron star system [9] were recorded in the data stream of the Advanced LIGO and Virgo interferometers. More varied detections are anticipated in future LIGO and Virgo observation runs [10–12], spurring a plethora of astrophysical and theoretical investigations. KAGRA [13] and LIGO-India [14] will join the international network, enormously improving localization of astrophysical sources and the network duty cycle. Commissioning activities will strive to bring the instruments to design sensitivity. Instrumental R&D will focus on the design and realization of the next generation of gravitational-wave interferometric detectors on Earth [15] and in space [16]. All these activities will be crucial for the growth of gravitational-wave astrophysics from a sensational news item to a full-grown scientific method to explore our universe.

The measured rate of gravitational-wave detections in the first observing run (O1) and second observing run (O2) of Advanced LIGO and Virgo implies that the international network of interferometers is poised to detect a significant number of gravitational-wave events in the coming years. The third Advanced LIGO-Virgo observing run (O3) is scheduled for early 2019. As the gravitational-wave detector network reaches a stage that supports rates of detections of astrophysical gravitational-wave sources as high as $\sim$1 day^{-1}, a fast and accurate assessment of data quality will be critical.

The Advanced LIGO and Virgo detectors are sensitive to a variety of disturbances of non-astrophysical origin with characteristic frequencies in the instrument band of sensitivity [17, 18]. Noise transients of instrumental or environmental origin increase the false alarm rate of searches for gravitational-wave bursts and compact binary coalescences as well as affect measurements of these signals. The most remarkable example of the effect of a noise transient on a gravitational-wave signal is undoubtedly the glitch that occurred in the LIGO-Livingston detector in coincidence with the binary neutron star merger detection [9] and had to be carefully modeled and subtracted from the data to accurately determine the properties of the signal. Noise in the frequency domain affects searches for long-lived transients, continuous waves and stochastic background. Removing non-astrophysical artifacts from the data and improving the background of LIGO's searches is