Utilizing Machine Learning to Search for LIGO Sources

Sarah McCarthy[1]

June 19, 2018

[1]Grinnell College,
1115 8th Ave. Grinnell, IA 50112

1. Introduction

In 1916, Albert Einstein predicted the existence of gravitational waves, ripples created by accelerating masses that would propogate through spacetime [1,2]. However, he believed they would be too small for human detection. On September 14, 2015 at 09:50:45 UTC, nearly one century after Einstein's prediction, both detectors of the Laser Interferometric Gravitational-wave Observatory (LIGO) simultaneously observed a transient gravitational-wave signal [3]. This was the first direct detection of gravitational waves, as well as the first detection of a binary black hole merger. The two LIGO obeservatories are in Livingston, LA and Hanford, WA. Each instrument is a dual-recycled Michelson interferometer with 4 km arms [4]. LIGO's discoveries were made possible by a factor of 10 sensitivity improvement in the frequency regime around 100 Hz.

One of the challenges for LIGO is differentiating between signal and noise. Transient noise arises as short $O(seconds)$ glitches in the data that can mimic true transient astrophysical gravitational wave signals including binary black hole mergers. One powerful method to identify signals with waveforms that are well predicted by Einstein's relativity, including neutron star and black hole binaries, is matched filtering, which calculates the cross-correlation between modeled templates and the noisy gravitational wave detector data. The PyBC pipeline, which identified all LIGO discoveries to date, employs matched filtering to calculate the signal-to-noise ratio (SNR) between all modeled templates considered in a gravitational wave search. To help make the pipeline more robust to noise, PyCBC uses a χ^2 test to downweight the SNR of events where the data does not match the modeled template well. Any times were the re-weighted SNR is above threshold are saved as "triggers" [5]. Any given trigger will have multiple associated modeled templates with non-zero re-weighted SNR, each with individual re-weighted SNR tends to be densely clustered in total mass and effective spin.

One of the outputs of this process is a set of parameters, such as total mass, effective spin, and maximum re-weighted SNR [6]. These parameters can be plotted for a given stretch of data. For gravitational wave signals, the plots typically have a compact area of high maximum re-weighted SNR, with a relatively low re-weighted SNR over other values of total mass and effective spin. If the event is noise, typically maximum re-weighted SNR is not well localized in these template parameters.

Our approach to improving the PyCBC pipeline performance by limiting the impact of transient noise is to use a convolutional neural network and image classification. A neural network is a biology-inspired computer program in which a computer learns a specified task from a series of provided 'training' examples. Neural networks have successfully been used to classify images of LIGO data in the past [7]. We will build an image classifier that takes as input a plot representing the distribution of templates associated with a trigger time in total mass and effective spin. We expect that the much more well-localized appearance of true signals in this parameter space will serve as a powerful distinguishing feature for our machine learning image classifier.

2. Objectives

The aim of my summer research project is twofold:
1) Create a convolutional neural network that will differentiate signal and noise in LIGO data
2) Test this algorithm's performance on increasingly large data sets

3. Approach

I will first design and build a simple convolutional neural network (CNN) algorithm that can intake images of the total mass and effective spin distribution of PyCBC triggers and output some likelihood that the trigger belongs to the 'signal' class or the 'noise' class. Next I will need to develop a training set to train the CNN to make accurate classifications. I will inject a series of simulated gravitational wave signals into data from Advanced LIGO's second observing run and flag each of these as part of the 'signal' class. For the glitch class, I will use glitch examples identified by Gravity Spy.

After I have trained the CNN with known examples of both the signal and glitch classes, I will test it with a test data set. To ensure this data is indepdent from the training set I will use a different period of Advanced LIGO data, and I will follow the same method as above to inject signal examples and identify glitch examples. I will evaluate the performance of the CNN by producing a confusion matrix for the test data which will calculate the fraction of misclassified PyCBC triggers for each class. I anticipate tuning the CNN based on these results.

At this point, I will be ready to expand my training set and data set to an extended set of O2 data. Then, I will produce and analyze my results. All simulations will be run from my personal laptop using Caltech's LDAS computing cluster for computational power.

4. Project Schedule

I will follow the timeline below:

Weeks 1-2: Learn LIGO software and computing clusters, assemble basic code
Weeks 3-5: Tune algorithm using isolated test cases
Weeks 6-8: Run on extended data set
Weeks 9-10: Combine results, prepare final report and presentation

Acknowledgements

I would like to thank Dr. TJ Massinger and Dr. Jess McIver for their guidance and support throughout the proposal process.

References

1. A. Einstein, Sitzungsber. K. Preuss. Akad. Wiss. 1, 688 (1916).
2. A. Einstein, Sitzungsber. K. Preuss. Akad. Wiss. 1, 154 (1918).
3. B.P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration). "Observation of Gravitational Waves from a Binary Black Hole Merger." *Phys. Rev. Lett. 116. 061102* (2016)
4. LIGO Scientific Collaboration. "Advanced Ligo." *arXiv:1411.4547* (2014)
5. B.P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration). "Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914." *arXiv:1602.03844* (2016)
6. B P Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration). "Effects of data quality vetoes on a search for compact binary coalescences in Advanced LIGO's first observing run." *Class. Quantum Grav. 35 065010* (2018)
7. M Zevin et al. "Gravity Spy: integrating advanced LIGO detector characterization, machine learning, and citizen science." *Classical and Quantum Gravity, Volume 34, Number 6.*(2017)