Binary Black Hole Automated Identification by Agglomerative Clustering based on Gravitational Waves

Subhrangshu Adhikary¹, Saikat Banerjee²

¹Department of Computer Science and Engineering, Dr. B.C. Roy Engineering College, Durgapur-713206, West Bengal, India
²Department of Mechanical Engineering, CubicX, Kolkata-700070, West Bengal, India
E-mail: subhrangshu.adhikary@spiraldevs.com

Abstract. The General Theory of Relativity, proposed by Albert Einstein theoretically predicted that very large accelerating mass creates ripples in spacetime which is the strongest for merging binary black hole system and the ripples can travel billions of light-years and these ripples are called Gravitational Waves. By the time these waves reach Earth, they become very faint and can't be detected with regular methods. For this, LIGO has created specialized detectors based on the laser interference principle to detect strains caused by gravitational waves in e-19 scale. GW190521 is a gravitational wave event recorded on 21 May 2019 at 03:02:29 UTC and caused by the merger of two black holes of 85M⊙ and 66 M⊙ whose progenitor was the largest ever recorded. Throughout literature, very few amounts of autonomous black hole identification models have been made because of limited data availability. This experiment proposes methods for autonomous identification of black holes by using an unsupervised machine learning algorithm called Agglomerative Clustering with very little data to train which can adapt quickly to gravitational wave events. The model could be easily deployed near laser interferometric observatories for autonomous black hole identification with minimal effort.

Keywords: Binary Black hole, Gravitational Wave, GW190521, Unsupervised Machine Learning, Clustering

1. Introduction

The "General Theory of Relativity" proposed by Sir Albert Einstein in 1915 predicts the existence of black holes as well gravitational waves [1]. The interferometer records of September 14th, 2015 were very close with Einstein's theoretical predictions. This was a remarkable moment in science while this invention opened a new age of gravitational wave astronomy. Black holes are one of the most enigmatic astronomical substances found in the universe. Theory and simulation of black holes allow us to improve our understanding of gravity. A black hole is a type of singularity which is a location in spacetime where the gravity and density tend to be of infinite scale causing it to pull the spacetime continuum surrounding it in such a manner that even light fails to escape [2-4]. The theory, predicts that massive bodies like black holes or neutron stars capable of accelerating can create ripples within spacetimes called gravitational waves. Specifically, gravitational forces compress the distances between the bodies. A gravitational effect on a test mass ring is to alternately stretch it in one of the directions and compress in the other.

\[
\frac{1}{2} \frac{d^2 \xi^i}{dt^2} \frac{d^2 \xi^j}{dt^2} \Omega_{ij}(\ell - i j \hat{c})
\]  

(1)
Einstein was one of the first to calculate the energy loss caused by the emission by gravitational signals from a binary system. His famous quadrupole equation (1) gives us the formula for gravitational strain from a gravitational wave where $Q_{ij}$ denotes the quadrupole tensions at the source.

1.1. Detection of Gravitational Waves by LIGO

Einstein observed that gravitational wave from realistic astrophysical sources is so weak that they cannot be detected. There was also some doubt about the existence of a gravitational wave. With the development of theories for understanding the behaviour of massive objects, for example, black holes or neutron stars and other astrophysical observations like binary pulsar systems, it is clear that gravitational waves do exist. This all changed when Laser Interferometer Gravitational-wave Observatory (LIGO) which is a physical instrument that senses spacetime undulations detected gravitational waves on September 14, 2015 which was possibly caused by two colliding black holes estimated to be around 1.3 billion light-years far from earth. LIGO's discovery is one of the greatest scientific achievements in human history. Although gravitational waves can be very destructive and violent, they become quadrillion times weaker by the time they reach Earth. The amount of space-time vibrating because of gravitational waves that were first detected by LIGO was 1000 times smaller than an atom's nucleus. LIGO was created to detect strains on this scale [5-7].

1.2. GW190521 Event Recorded by LIGO

GW190521 is a gravitational wave that was recorded by LIGO and Virgo observatories on 21 May 2019 at 03:02:29 UTC was a landmark discovery because the mass of the progenitor black hole was the highest ever recorded to date. The event was a result of the collision of two black holes having mass 85 $M_\odot$ and 66 $M_\odot$ respectively and they were estimated to be around 17 billion light-years apart from the earth. Astronomers observed a flash of light in June 2020 that could be linked to GW190521. Zwicky Transient Facility reported the detection of a transient optical supply in the vicinity of the GW190521 trigger. However, the merger was still uncertain due to the uncertainty in the sky position spread across hundreds of square degrees and it was later confirmed with subsequent observations. If the two events are related, the event is believed to be the first-ever detection of electromagnetic sources concerning the merger between the two black holes. Lights are usually not emitted by black hole mergers unless there are some Hawkin's Radiation detected. Scientists around the world suggest that this phenomenon could be understood by the fact that the collision of two smaller black holes produces a newly formed intermediate-mass black hole, disrupting the accretion disk material and emitting a flare at around 200 km/s through its accretion disk. This explains that the flare is likely to occur again after 1.6 years from the collision when the intermediate-mass black hole encounters the accretion disk and accordingly the observations of the merger event was finally confirmed by LIGO and Virgo observatories [8-10].

Advanced LIGO and Virgo now have a new vision of understanding the Universe by detecting gravitational waves (GWs). The LIGO and Virgo Scientific Collaboration (LVC), have reported that GWs have been successfully detected from 10 binary black hole mergers (BH) and a binary star moving in a spiral. O3 was the third and final observing run. It was started on April 1, 2019, however, this was suspended on March 27, 2019. Numerous alerts about possible GW detections were received by the astronomical society, with three confirmed discoveries. The discovery of GW150914 and the subsequent events have revealed that a large population of binary BHs have a combined total mass between 19 $M_\odot$ to 84 $M_\odot$. Their component masses range from 8-50 $M_\odot$. The signals consistent for heavier BHs have been reported in it was concluded that less than 1% of primary BH masses are greater than 45 $M_\odot$ in parametrized population modelling. Stars with a range of helium centre mass between 32-64 $M_\odot$ and 65 $M_\odot$ will be affected by pulsational pairing instability. Stars with helium central mass between 64-135 $M_\odot$ would also be vulnerable to pair-instability. Stars with helium content in the range 64-135 $M_\odot$ would not leave any compact remnants, while stars with 135 $M_\odot$ helium mass are believed to immediately collapsing to intermediate mass (IMBHs). LVC O1 to O2
observations indicate that there is no heavy BH formation via pair-instability supernovae (PISN). For GW190521, there is a high chance that the mass for the heavier binary component will be within the PISN Mass Gap. BHs having mass below the PISN gap are likely to form in the dense stellar systems or active galactic nuclear nuclei disks. They might be formed through a sequence of hierarchical collision of smaller BHs or the direct collapse of a stellar fusion of an evolved and main sequence companion. The search for quasicircular binary colliders revealed GW190521, but no evidence is found to support such a departure from the signal model [11-13].

1.3. Machine Learning and its applications in black hole studies
Unsupervised learning is a popular subset of machine learning which is used to detect abnormalities or irregularities from a dataset without the difference being priorly trained to the model. Agglomerative Clustering is a type of unsupervised learning which is a form of hierarchical clustering which groups the dataset in a bottom-up approach. The algorithm initially tends to detect all the points to be a separate cluster and then the cluster subsets are merged one after another based on their properties to form \( n \) large cluster where \( n \) is the desired number of clusters to be formed. [14-16]

1.4. Motivation for the experiment
After an extensive literature review, we have found some gaps in studies to implement machine learning algorithms for the study of astronomical events such as black holes. Several works have been made by studying the accretion disk and images of shadow of black holes with limited resolution telescope with high frequencies with convolution neural network to extract many features of a black hole. Gravitational waves have been analysed with Bayesian neural networks for multiple parameter estimations. Machine learning has also been widely used to filter noises from LIGO data helping it to extract the important features. Efforts have also been made to classify machine learning in a supervised fashion however due to the limited amount of GW events ever recorded, it makes it difficult to reliably predict black holes with this approach [17-20]. There are no significant studies performed to cluster the irregular LIGO data which corresponds to a black hole reliably without prior training. Therefore to overcome this gap, the article describes methods to use Agglomerative Clustering to detect abnormalities in LIGO recordings and cluster the black holes whenever detected without prior training.
Figure 1. The binary black hole system causes ripples in spacetime called gravitational waves which travels through space and when reaching the earth, LIGO records the strain caused by the event which has been analysed with Agglomerative Clustering algorithm for automated detection for such events.

The work at a glance is represented by figure 1. The merger of the black holes created the ripples in spacetime which travels a long distance reaching earth which is tracked by LIGO. This data is then processed to make it suitable for unsupervised learning and then finally used the Agglomerative Clustering method to detect the black holes without prior training.

2. Data & Methods
Revolving binary black holes creates ripples in spacetime which propagates through long distances across space. The ripples cause temporary compression and expansion in any object that comes into its path. This causes a series of strains that can be detected with properties of interference of light. Laser interferometers are built by a source of light that gets split, travels through two light carrying channels perpendicular to each other which gets reflected and interfere at the end and the resultant wave is detected by the detector. But when there's an incoming gravitational wave, the space around these channels compresses or expands altering the length of the channels making the light travel from source to destination in the different time periods and ultimately deviating the interference from 0 to some detectable intensity which is ultimately detected by a specialized detector at the conjunction of the
light channels. However, in practice, ideal 0 is very less likely to be found because of noises and therefore it is necessary to filter the data from noises to isolate gravitational waves which are discussed later in the text. [21-23]

2.1. GW190521 event and its recorded data

Figure 2. The strain graph for the gravitational wave event GW190521 which were detected by LIGO Hanford, LIGO Livingston and Virgo detectors

The observation of GW190521 event which was initially known as S190521g was a resultant wave cause due to the merger between two black holes of 85 and 66 solar masses approximately 17 billion light-years apart detected on 21 May 2019 at 03:02:29 UTC by LIGO and Virgo detectors. The merger resulted in an intermediate-mass black hole which was equivalent to 142 times the mass of the sun and the black holes causing this merger are estimated to be the largest progenitor masses ever observed. The data of this event was first published publicly by LIGO and Virgo for research on 2nd September 2020 which have been further investigated in this experiment [24]. The dataset used for the experiment consists of a strain graph of 4KHz for 32 seconds between which the event has been recorded in three detectors namely LIGO Hanford (H1), LIGO Livingston (L1) and Virgo (V1). The experiment has been conducted on approximately 128000x3 data points. The changes in the intensity of the interference from different locations can be calibrated to fine-tune the location of the black hole. Figure 2 shows the representation of the data recorded by the three different detectors. The x-axis here represents the time sequence in milliseconds and the y-axis represents the strain reading. The time slice shown in the graph does not correspond to the part of any gravitational waves but represents the general reading which shows the presence of noise in the detector. When there's a gravitational wave received at the detector, the strain increases significantly from the regular noise range which in this experiment have been clustered for autonomous detection of binary black holes. [25-29]
2.2. Agglomerative Clustering for Binary Black Hole Detection

Agglomerative Clustering technique is useful for isolating $n$ number of clusters of data having contrasting features. It is preferred for the experiment over other popular unsupervised machine learning algorithms like K-Means Clustering, Gaussian Mixture Model and Affinity Propagation this is because of the speed of execution. K-Means clustering is a very effective algorithm as it works by calculating the centroid for the two clusters but it does so by comparing distances from each point to estimate the centroid making it a very time-consuming model. The Gaussian Mixture model is fast, however, it has been found to often have lower accuracy in forming the cluster. Affinity propagation is a technique that doesn't need to explicitly define the exact number of clusters to form and it parses through each point and measures its properties to estimate the cluster. This has the property to discover strains that correspond to a black hole even without prior training. Although this is quite accurate and can adapt to large variations, this has a limitation in speed due to its exponential time complexity it takes too long to train a large dataset that has not been considered for the experiment. The agglomerative clustering model developed in the experiment consists of two clusters that have been trained explicitly on the data of binary black holes to identify data corresponding to such events. To measure the distances between the points, euclidean distances have been used which works as differences of points in cartesian coordinates. While forming the clusters, the ward linkage method has been used to minimize the variance between the distances of observation sets while merging subclusters to a major cluster. [30-32]

3. Results & Discussion

The ripples in spacetime caused by binary black hole collision or revolution around each other travel large distances and similarly such ripples called gravitational waves reached earth known as GW190521 and recorded by the observatories. The observatories should have an ideal 0 strain as recorded by the interference detector but in practice, there are some noises recorded by the detectors. The noises spread across a regular range of strain which varies according to the detector. The LIGO Hanford H1 normal strain ranges from around -2x1e-19 to 2x1e-19 range where the maximum strain ranges within proximity of 0. The LIGO Livingston (L1) observatory strain data normally stays within a range of roughly around -7x1e-19 to 7x1e-19 which mostly accumulate toward 0 ranging from roughly around -2.5x1e-19 to 2.5x1e-19. Likewise, the Virgo (V1) detector range lies around -3x1e-19 to 3x1e-19. The majority of the V1 data accumulates around 0 ranging from around -1.5x1e-19 to 1.5x1e-19. These are the general range of observatory readings and figure 3 represents the graph of the strain of one detector with respect to the strain of other detectors. During a gravitational wave event, the strain increases which are detected by the observatories. During the GW190521 event, it could be noticed that the recording from all three locations have significant deviations from their general strain ranges which ensures the presence of gravitational waves and further investigations reveals the nature of the source. It can be noticed that the H1 strain peaked up to -3x1e-19 and 3x1e-19 making a clear distinction of binary black hole event. The V1 strain during the event peaked up to -6x1e-19 to 6x1e-19 which is another clear distinction of the event. Similarly, L1 strain had peaked up to -7.5x1e-19 to 7.5x1e-19 during the event. This gives a statistical reference to form the cluster to isolate binary black holes gravitational waves from regular noises.
Figure 3. The Strain vs Strain graph for different locations during GW190521 event indicating black holes being the peak of strain

To detect the deviations from regular strain readings of the detectors, agglomerative clustering plays an important role here. However, the strain data, in general, oscillates from negative to positive range and also the portions from the readings which explicitly corresponds to a binary black hole are much closer to the regular readings of the detector which are likely to degrade the detection performance of the agglomerative clustering algorithm. Therefore to solve both the issues, the strain data have been squared which amplified the event and shifting all the strain data in a positive phase making it easier for the algorithm to detect the pattern. After that the model has been trained to specifically detect the ranges which correspond to a gravitational wave event, particularly for GW190521, however, the detection mechanism can be used for any other gravitational wave event. The agglomerative clustering algorithm has successfully detected the peaks in the strain graph which explicitly corresponds to a gravitational wave caused by a binary black hole. The algorithm has clustered the H1 strain to be a GW event for squared strains above 0.42x1e-37 and all recordings below that range are marked as regular detector data. Similarly, for the L1 strain, the squared strain roughly above 2.8x1e-37 corresponds to a GW event. The squared V1 strain approximately above 2x1e-37 confirms the presence of a binary black hole event. Anything below that approximate range generally falls within the regular detector reading range and need not necessarily be a black hole. However, in order to detect a very faint gravitational wave that cannot be detected with this technique, the sensitivity of the detector requires to be increased or a more accurate noise filtering process is required which can be covered in future experiments.

Figure 4. Detection of Black Holes with Agglomerative Clustering from processed LIGO data of GW190521 event and their relationship between three different locations namely LIGO Hanford (H1), LIGO Livingston (L1) and Virgo (V1)
4. Conclusion

Accelerating large masses causes the surrounding spacetime to create ripples in spacetime. When the source is very large, the ripples travel through spacetime through very large distances called Gravitational Waves. The waves this large to travel billions of light-years is usually caused by a binary black hole system revolving around each other or their merger and neutron stars. These waves are however very faint to be detected normally and therefore specialized detectors such as LIGO are used. The paper discusses the event known as GW190521 which is a merger of two black holes whose progenitor was the largest ever recorded.

The paper presents a method to process the data making it suitable for unsupervised learning and using the algorithm called Agglomerative Clustering for autonomous identification of such gravitational waves isolating from noises to indicate the collision of a binary black hole system. The paper presents the relation of such events and their recordings for data collected from LIGO Hanford, LIGO Livingston and LIGO Virgo and how the detection results differ for the same black hole event detected with the algorithm for the three locations.

The model could be easily deployed without prior training explicitly with black hole data and can be trained with only existing noise data and whenever the gravitational waves are encountered, the algorithm would automatically detect the abnormality and confirm whether it is a black hole. The model could further be developed to address the nature of the source of the gravitational waves.

Acknowledgement

This experiment has been conducted as a segment of the Gyanam Project launched by a joint venture between CubicX and Spiraldevs Automation Industries Pvt. Ltd. West Bengal, India. The work is dedicated to sweet memories of Bisky.

References

[1] Abbott BP, Abbott R, Abbott TD, Abraham S, Acernese F, Ackley K, et al. Tests of general relativity with the binary black hole signals from the LIGO-Virgo catalog GWTC-1. Phys Rev D. 2019 Nov;100:104036. Available from: https://link.aps.org/doi/10.1103/PhysRevD.100.104036.
[2] Venumadhav T, Zackay B, Roulet J, Dai L, Zaldarriaga M. New binary black hole mergers in the second observing run of Advanced LIGO and Advanced Virgo. Phys Rev D. 2020 Apr;101:083030. Available from: https://link.aps.org/doi/10.1103/PhysRevD.101.083030.
[3] Zackay B, Venumadhav T, Dai L, Roulet J, Zaldarriaga M. Highly spinning and aligned binary black hole merger in the Advanced LIGO first observing run. Phys Rev D. 2019 Jul;100:023007. Available from: https://link.aps.org/doi/10.1103/PhysRevD.100.023007.
[4] Stoica OC. The geometry of singularities and the black hole information paradox. In: Journal of Physics: Conference Series. vol. 626. IOP Publishing; 2015. p. 012028.
[5] Liu C, Shao L, Zhao J, Gao Y. Multiband observation of LIGO/Virgo binary black hole mergers in the gravitational-wave transient catalog GWTC-1. Monthly Notices of the Royal Astronomical Society. 202006; 496(1): 182–196. Available from: https://doi.org/10.1093/mnras/staa1512.
[6] McIsaac C, Keitel D, Collett T, Harry I, Mozzon S, Edy O, et al. Search for strongly lensed counterpart images of binary black hole mergers in the first two LIGO observing runs. Phys Rev D. 2020 Oct;102:084031. Available from: https://link.aps.org/doi/10.1103/PhysRevD.102.084031.
[7] Belczynski K, Klencki J, Fields C, Olejak A, Berti E, Meynet G, et al. Evolutionary roads leading to low effective spins, high black hole masses, and O1/O2 rates for LIGO/Virgo binary black holes. Astronomy& Astrophysics. 2020;636:A104.
[8] Hannuksela OA, Haris K, Ng KKY, Kumar S, Mehta AK, Keitel D, et al. Search for Gravitational Lensing Signatures in LIGO-Virgo Binary Black Hole Events. The Astrophysical Journal. 2019 mar;874(1):L2. Available from: https://doi.org/10.3847/2041-8213/ab6c0f.
[9] Roulet J, Zaldarriaga M. Constraints on binary black hole populations from LIGO–Virgo detections. Monthly Notices of the Royal Astronomical Society. 2019 01;484(3):4216–4229. Available from: https://doi.org/10.1093/mnras/stz226.
[10] Abbott BP, Abbott R, Abbott TD, Abraham S, Acernese F, Ackley K, et al. Binary Black Hole Population Properties Inferred from the First and Second Observing Runs of Advanced LIGO and Advanced Virgo. The Astrophysical Journal. 2019 sep;882(2):L24. Available from: https://doi.org/10.3847/2041-8213/ab3800.
[11] Roulet J, Venumadhav T, Zackay B, Dai L, Zaldarriaga M. Binary black hole mergers from LIGO/Virgo O1and O2: Population inference combining confident and marginal events. Phys Rev D. 2020 Dec;102:123022. Available from: https://link.aps.org/doi/10.1103/PhysRevD.102.123022.
[12] Nitz AH. Distinguishing short duration noise transients in LIGO data to improve the PyCBC search for gravitational waves from high mass binary black hole mergers. Classical and Quantum Gravity. 2018 Jan;35(3):035016. Available from: https://doi.org/10.1088/1361-6382/aaad3d.

[13] Callister T, Fishbach M, Holz DE, Farr WM. Shouts and Murmurs: Combining Individual Gravitational-wave Sources with the Stochastic Background to Measure the History of Binary Black Hole Mergers. The Astrophysical Journal. 2020 Jun;896(2):L32. Available from: https://doi.org/10.3847/2041-8213/abf743.

[14] Murtagh F, Legendre P. Ward’s hierarchical agglomerative clustering method: which algorithms implement Ward’s criterion? Journal of classification. 2014; 31(3): 274–295.

[15] Wong KW, Franciolini G, De Luca V, Baibhav V, Berti E, Pani P, et al. Constraining the primordial black hole scenario with Bayesian inference and machine learning: the GWTC-2 gravitational wave catalog. Physical Review D. 2021;103(2):023026.

[16] Krastev PG, Gill K, Villar VA, Berger E. Detection and parameter estimation of gravitational waves from binary neutron-star mergers in real LIGO data using deep learning. Physics Letters B. 2021;815:136161. Available from: https://www.sciencedirect.com/science/article/pii/S0370269321001015.

[17] WeiW, Huerta EA. Gravitational wave denoising of binary black hole mergers with deep learning. Physics Letters B. 2020; 800: 135081. Available from: https://www.sciencedirect.com/science/article/pii/S037026931930832.

[18] Perna R, Chruslinska M, Corsi A, Belczynski K. Binary black hole mergers within the LIGO horizon: statistical properties and prospects for detecting electromagnetic counterparts. Monthly Notices of the Royal Astronomical Society. 2018 Oct; 487(1): 4228–4240. Available from: https://doi.org/10.1093/mnras/sty814.

[19] Truong TDH, Quach MA, Nguyen HH, Nguyen UT, Dong DH, Pham VNT. Investigation of the dependence of multiple recollision on laser wavelength in the non sequential double ionization process. Journal of Physics: Conference Series. 2021 May;1932(1):012004. Available from: https://doi.org/10.1088/1742-6596/1932/1/012004.

[20] Barausse E, Bellovary J, Berti E, Holley-Bockelmann K, Farris B, Sathyaprakash B, et al. Massive Black hole Science with eLISA. In: Journal of Physics: Conference Series. vol. 610. IOP Publishing; 2015. p. 012001.

[21] Lower ME, Thrane E, Lasky PD, Smith R. Measuring eccentricity in binary black hole in spirals with gravitational waves. Phys Rev D. 2018 Oct; 98:083028. Available from: https://link.aps.org/doi/10.1103/PhysRevD.98.083028.

[22] Fishbach M, Holz DE. Picky Partners: The Pairing of Component Masses in Binary Black Hole Mergers. The Astrophysical Journal. 2020 Mar;891(1):L27. Available from: https://doi.org/10.3847/2041-8213/abf747.

[23] Hotokezaka K, Piran T. Implications of the Low Binary Black Hole Aligned Spins Observed by LIGO. The Astrophysical Journal. 2017 Jun;842(2):111. Available from: https://doi.org/10.3847/1538-4357/aaaf61.

[24] Gravitational Wave Open Science Center GW190521 Event; 2020. (Accessed on 06/28/2021).https://www.gw-openscience.org/eventapi/html/O3DiscoveryPapers/GW190521/v2/.

[25] Bavera SS, Fragos T, Qin Y, Zapartas E, Neijssel CJ, Mandel I, et al. The origin of spin in binary black holes: Predicting the distributions of the main observables of Advanced LIGO. Astronomy & Astrophysics. 2020; 635:A97.

[26] Abbott R, Abbott T, Abraham S, Acernese F, Ackley K, Adams C, et al. GW190521: A binary black hole merger with a total mass of 150 M. Physical review letters. 2020; 125(10):101102.

[27] Sakstein J, Croon D, McDermott SD, Straight MC, Baxter EJ. Beyond the Standard Model Explanations of GW190521. Phys Rev Lett. 2020 Dec;125:261105. Available from: https://link.aps.org/doi/10.1103/PhysRevLett.125.261105.

[28] Dai X. The Black Hole Paradoxes and Possible Solutions. In: Journal of Physics: Conference Series. vol. 1634. IOP Publishing; 2020. p. 012088.