Application of artificial neural network to search for gravitational-wave signals associated with short gamma-ray bursts

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2015 Class. Quantum Grav. 32 245002
(http://iopscience.iop.org/0264-9381/32/24/245002)

View the table of contents for this issue, or go to the journal homepage for more

Download details:

IP Address: 194.94.224.254
This content was downloaded on 05/02/2016 at 08:37

Please note that terms and conditions apply.
Application of artificial neural network to search for gravitational-wave signals associated with short gamma-ray bursts

Kyungmin Kim, Ian W Harry, Kari A Hodge, Young-Min Kim, Chang-Hwan Lee, Hyun Kyu Lee, John J Oh, Sang Hoon Oh and Edwin J Son

1 Department of Physics, Hanyang University, Seoul 133-791, Korea
2 Department of Physics, Syracuse University, Syracuse, NY 13244, USA
3 Department of Physics, California Institute of Technology, Pasadena, CA 91125, USA
4 Department of Physics, Pusan National University, Busan 609-735, Korea
5 Division of Computational Sciences in Mathematics, National Institute for Mathematical Sciences, Daejeon 305-811, Korea

E-mail: kmkim82@hanyang.ac.kr

Received 3 March 2015, revised 24 August 2015
Accepted for publication 2 October 2015
Published 24 November 2015

Abstract

We apply a machine learning algorithm, the artificial neural network, to the search for gravitational-wave signals associated with short gamma-ray bursts (GRBs). The multi-dimensional samples consisting of data corresponding to the statistical and physical quantities from the coherent search pipeline are fed into the artificial neural network to distinguish simulated gravitational-wave signals from background noise artifacts. Our result shows that the data classification efficiency at a fixed false alarm probability (FAP) is improved by the artificial neural network in comparison to the conventional detection statistic. Specifically, the distance at 50% detection probability at a fixed false positive rate is increased about 8%–14% for the considered waveform models. We also evaluate a few seconds of the gravitational-wave data segment using the trained networks and obtain the FAP. We suggest that the artificial neural network can be a complementary method to the conventional detection statistic for identifying gravitational-wave signals related to the short GRBs.

Keywords: gravitational-waves, short gamma-ray bursts, artificial neural networks

(Some figures may appear in colour only in the online journal)
1. Introduction

Gamma-ray bursts (GRBs) are the most energetic electromagnetic events among various observable astronomical phenomena in the Universe. GRBs are very frequent events, as we observe them more or less once a day in isotropic spatial distribution (Meegan et al 1992). In general, all observed GRBs can be classified in two categories, long and short, by a characteristic duration time, $T_{90}$, of the gamma-ray flashes. If the $T_{90}$ of a GRB is longer or shorter than $7 \times 2$ s, it is classified as a long or short GRB, respectively. The most plausible scenario for the progenitor of long GRBs is a stellar collapse of a massive star to observe them more or less once a day in isotropic spatial distribution observable astronomical phenomena in the Universe. GRBs are very frequent events, as we observe them more or less once a day in isotropic spatial distribution (Meegan et al 1992). In general, all observed GRBs can be classified in two categories, long and short, by a characteristic duration time, $T_{90}$, of the gamma-ray flashes. If the $T_{90}$ of a GRB is longer or shorter than $7 \times 2$ s, it is classified as a long or short GRB, respectively. The most plausible scenario for the progenitor of long GRBs is a stellar collapse of a massive star to (i) a black hole with a forming accretion disk or (ii) a highly magnetized neutron star (Ott 2009). In contrast, for short GRBs, it is believed that the inspiral merging process of a compact binary system composed of at least one neutron star such as a binary neutron star (BNS) or a neutron star–black hole binary (NS-BH) is the most viable progenitor model (Berger 2014, Shibata and Taniguchi 2008).

The progenitor of short GRBs has been considered as one of the most promising sources of gravitational-waves (GWs) (Abadie et al 2012b) that can be detected by the ground-based GW detectors such as the Laser Interferometer Gravitational-wave Observatory (LIGO) (Abbott et al 2009b) in the US and Virgo (Acernese et al 2008) in Italy. Indeed, the LIGO scientific collaboration and the Virgo collaboration have conducted searches for GW signals from compact binary coalescences (CBCs) related to short GRBs (or CBC-GRB searches in short) with the data taken from two recent joint runs (Abadie et al 2010b, 2012a, Aasi et al 2014). The data of the first joint search (Abadie et al 2010b) were taken from the fifth LIGO science run (S5) and the first Virgo science run (VSR1) and the data of the second search (Abadie et al 2012a) were taken from the sixth LIGO science run (S6) and the second and third Virgo science runs (VSR2 and VSR3). On the other hand, the third joint search (Aasi et al 2014) used the data obtained from both joint runs, S5/VSR1 and S6/VSR2,3, together. For the sky location and event time of a short GRB, the first and second searches used the gamma-ray coordinates network (GCN) alert and the third search used the information observed by the satellites of the interplanetary network (IPN). In order to try to observe relevant GW signals, the search pipelines for those previous searches use a matched filtering method (Owen and Sathyaprakash 1999) with template waveforms (Cokelaer 2007).

From the matched filtering, the search pipeline finds events which are highly correlated to the template waveforms by calculating the SNR. When events in more than one detector within the network have an SNR that exceeds a predetermined search threshold, the events are recorded as a trigger. A trigger is characterized by several statistical quantities, for example, the SNRs in individual detectors, combined SNR statistics, values of various signal consistency tests, and other variants of them which are used for other consistency tests (refer to Harry and Fairhurst 2011 for details). Thus, a trigger is characterized by a multi-dimensional vector of values.

Numerous studies (Chen and Holz 2013, Clark et al 2015, Siellez et al 2014, Singer et al 2014) estimate that the expected detection rate of joint GW-GRB events observed by advanced LIGO and advanced Virgo will be around 0.1–1 per year. Furthermore, the maximum distance reach of advanced LIGO, $z \sim 0.1$, is approximately the distance to the closest

---

6 It is defined by the time taken to accumulate 90% of the burst fluence starting at 5% of fluence level.
7 This fiducial time has been determined from the statistical distribution of observed duration times for BATSE sample. Note that $T_{90} = 0.8$ s is used to distinguish between collapsar and non-collapsar progenitors for Swift sample (Bromberg et al 2013).
8 http://gcn.gsfc.nasa.gov
9 https://heasarc.gsfc.nasa.gov/docs/heasarc/missions/ipn.html
short GRB with known redshift. Therefore, as the era of advanced LIGO and Virgo is imminent, we can expect that even a small improvement in the distance reach of a search could make the difference between a tentative association and the confident detection of a GW signal from a short GRB. Also, since the GW signals due to binary mergers are well modelled, once we detect a GW signal from a short GRB with the advanced detectors, it will allow the advent of the GW astronomy.

In the recent CBC-GRB searches, two different ranking methods, a likelihood ratio (Abadie et al 2010b) and a detection statistic (Harry and Fairhurst 2011) were calculated to classify whether a trigger is more likely to be caused by either a real GW signal radiated from an expected astrophysical source or a noise artifact originated by non-Gaussian and non-stationary noises coming from instruments and/or environments. The classification was done by the estimation of the false alarm probability (FAP) of a trigger based on the value of its detection statistic. The FAP of the triggers related to short GRBs were consistent with background; no evidence of GW signals in the GW data was found related to the considered short GRBs.

Over recent decades, various machine learning algorithms (MLAs) such as artificial neural networks (ANN) (Hecht-Nielsen 1989, Hastie et al 2009), random forests of bagged decision trees (Breiman 1996, 2001), and support vector machines (Cortes and Vapnik 1995, Cristianini and Shawe-Taylor 2000) have been developed and evolved to analyze multi-dimensional data efficiently. Application of MLAs to many problems including several GW related searches (Cannon 2008, Biswas et al 2013, Adams et al 2013, Rampone et al 2013) have shown good classification performances for their nonlinear multi-dimensional data, which provides a complementary way of making a decision. Among various MLAs, we specifically consider ANN which is used in Biswas et al (2013) for this study on the feasibility of applying MLA to the CBC-GRB search. At first, we investigate whether ANN can be as a new way of ranking a trigger for the CBC-GRB search. Then, we explore the possibility of the ANN as a potential method of improving the classification efficiency which is related to the improvement in the confidence level of a detection for a GW signal associated with a short GRB.

This paper is organized as follows. In section 2, we summarize the data preparation. In section 3, we introduce the basics of the ANN employed in this work, summarize the methodology of our investigation, and discuss the classification performance test. In section 4, we present the results of our application of ANNs in terms of the detection sensitivity as a function of distance and the evaluation of unknown triggers. Finally, we summarize the results and discuss the future prospects of this approach in the era of the advanced LIGO and advanced Virgo detectors in section 5.

### 2. Data preparation

#### 2.1. Gravitational-wave data

We focus on data from the fifth LIGO science run (S5) and the first Virgo science run (VSR1). The S5 data has been taken from the two LIGO detectors (H1 and H2) with 4 km and 2 km arms, respectively, in Hanford, Washington and a LIGO detector (L1) with 4 km arms, in Livingston, Louisiana. The VSR1 data has been taken from the 3 km arm Virgo detector (V1) at Cascina, Italy. Among the 22 short GRBs observed when at least two of the LIGO and

---

10 https://advancedligo.mit.edu
11 https://wwwcascina.virgo.infn.it/advirgo
Virgo instruments were operating during the first joint search (S5/VSR1), we select two GRBs, GRB070714B (Barbier et al. 2007a, Graham et al. 2007, Racusin et al. 2007, Berger 2014) and GRB070923 (Barbier et al. 2007b, Stroh et al. 2007, Berger2014), which were observed by Swift satellite (Gehrels et al. 2004) and have corresponding GW data in three detectors, H1, L1, and V1 available, coincidentally. The availability of using GW data is determined by the requirements of both stable operation at the event time of a GRB and provision of sufficient data (40 min in minimum) for the estimation of background distribution. We summarize the observation information of GRB070714B and GRB070923, including the event time and the sky location, in table 1, and the characteristics of corresponding GW detectors in table 2.

<table>
<thead>
<tr>
<th>Observation</th>
<th>GRB</th>
<th>UTC Time</th>
<th>RA</th>
<th>DEC</th>
<th>Duration, $T_{90}$ (sec)</th>
<th>Redshift</th>
</tr>
</thead>
<tbody>
<tr>
<td>070714B</td>
<td>14-07-2007</td>
<td>04:59:29</td>
<td>57.85°</td>
<td>28.29°</td>
<td>2.0</td>
<td>0.923</td>
</tr>
<tr>
<td>070923</td>
<td>23-09-2007</td>
<td>19:15:23</td>
<td>184.62°</td>
<td>-38.29°</td>
<td>0.05</td>
<td>N/A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GRB</th>
<th>H1</th>
<th>L1</th>
<th>V1</th>
<th>H1</th>
<th>L1</th>
<th>V1</th>
</tr>
</thead>
<tbody>
<tr>
<td>070714B</td>
<td>(−0.25, −0.07)</td>
<td>(0.27, 0.26)</td>
<td>(−0.83, −0.03)</td>
<td>0.26</td>
<td>0.37</td>
<td>0.83</td>
</tr>
<tr>
<td>070923</td>
<td>(0.15, −0.28)</td>
<td>(−0.13, 0.37)</td>
<td>(0.56, 0.40)</td>
<td>0.32</td>
<td>0.39</td>
<td>0.69</td>
</tr>
</tbody>
</table>

In the previous search (Abadie et al. 2010b), the authors concluded that they could find no evidence of GW signals related to these GRBs, GRB070714B and GRB070923; the estimated FAPs for the selected GRBs were consistent with the noise hypothesis. The exclusion distances to the potential progenitors of a GRB with 90% confidence level were computed—3.2 Mpc to a BNS progenitor and 5.1 Mpc to a NS-BH progenitor for GRB070714B, and 5.1 Mpc to a BNS progenitor and 7.9 Mpc to a NS-BH progenitor for GRB070923.

With the given GRBs’ observation information, we run the coherent search pipeline (Harry and Fairhurst 2011), which has been used for a recent CBC-GRB search (Abadie et al. 2012a), on the data from the sixth LIGO science run (S6) and the second and third Virgo science runs (VSR2 and VSR3). As a first step, the pipeline coherently combines the data from the three operational detectors, H1, L1, and V1. Next, the pipeline divides the combined
data into several partial segments such as the on-source and off-source segments as shown in figure 1 (see Abadie et al 2010b and Abadie et al 2012a for details of the segmentation). The on-source segment is \((-5,+1)\) seconds around the event times of selected GRBs. The off-source segment is 1944 seconds of data around the on-source segment that does not overlap with the on-source segment. The off-source segment is divided to 324 background bins in order to estimate FAP in the standard search so that the lowest FAP is \(1/324\) (see section 4.2 for further details). Note that, in section 3, we estimate the false positive rate (FPR) instead of FAP by using the number of triggers when testing ANN performance since the main interest of the section is to evaluate the classification performance of trained ANN (see section 3 for further details). In order to avoid contamination between on-source triggers and off-source triggers, there is a buffer segment between them. In section 4.2, we use trained ANN to classify unknown triggers in the buffer segment.

The coherent search pipeline performs matched filtering (Owen and Sathyaprakash 1999) on the data with the template bank of GW waveforms (Cokelaer 2007). The products of matched filtering are called as triggers: the on-source triggers denote triggers that may contain a potential GW event candidate and the off-source triggers are believed to be noise artifacts originated by instrumental and/or environmental noises. The off-source triggers are used for the background estimation of the on-source triggers. These on- and off-source triggers are characterized by several statistical quantities and physical parameters such as SNR, various signal consistency tests, and recovered value of component masses of a binary system. In specific, the statistical quantities are used to filter out useless triggers with following conditions (Harry and Fairhurst 2011): (i) \(\rho_{\text{coh}} > 6\), (ii) \(\rho_{N} < 5.25\), (iii) single detector SNR > 4, and (iv) \(\rho_{\text{new}} > 6\) where \(\rho_{\text{coh}}\) denotes the coherent SNR and \(\rho_{N}\) indicates the null SNR which is defined by

\[
\rho_{N}^2 = \rho_{\text{coh}}^2 - \rho_{\text{coinc}}^2.
\]

Here \(\rho_{\text{coinc}}\) is a quadratic sum of SNRs of triggers observed from two or more detectors in time coincidence.

In this case, the lowest possible FPR is a reciprocal of the number of off-source triggers (see equation (13)).
In our case, the index, \( i = \{ H1, L1, \text{and} V1 \} \) since we consider GW data observed from three detectors. Note that a GW signal matching the template waveform has no contribution to the null SNR. On the other hand, a noise transient that is incoherent across the data may give a large coherent SNR, but it is likely to also give a large null SNR. Therefore, requiring a small null SNR is effective at distinguishing incoherent noise transients from real GW signals.

Since the definition of the null SNR makes use of the template waveform, GWs which do not match the template can contribute to the null SNR (Harry and Fairhurst 2011). \( \rho_{\text{new}} \) is defined as

\[
\rho_{\text{coinc}}^2 = \sum_i \rho_i^2.
\]

In our case, the index, \( i = \{ H1, L1, \text{and} V1 \} \) since we consider GW data observed from three detectors. Note that a GW signal matching the template waveform has no contribution to the null SNR. On the other hand, a noise transient that is incoherent across the data may give a large coherent SNR, but it is likely to also give a large null SNR. Therefore, requiring a small null SNR is effective at distinguishing incoherent noise transients from real GW signals. Since the definition of the null SNR makes use of the template waveform, GWs which do not match the template can contribute to the null SNR (Harry and Fairhurst 2011). \( \rho_{\text{new}} \) is defined as

\[
\rho_{\text{new}} = \begin{cases} 
\rho_{\text{coh}} & \chi^2 \leq p, \\
\frac{\rho_{\text{coh}}}{\chi^2/p} & \chi^2 > p,
\end{cases}
\]

where \( p \) is the number of frequency bins used in the \( \chi^2 \) test. Then, for those remaining triggers, the detection statistic defined by

\[
\rho_{\text{DetStat}} = \begin{cases} 
\rho_{\text{new}} & \rho_N \leq 3.5, \\
\rho_{\text{new}}/(\rho_N - 2.5) & 3.5 < \rho_N < 5.25, \\
0 & \text{otherwise.}
\end{cases}
\]

With the detection statistic, coherent search estimates FAPs of remaining triggers by comparing the trigger’s statistic to the cumulative distribution of the statistic for off-source triggers. Then, we determine whether an on-source trigger is a GW signal or not based on the

### Table 3

Brief description of the input variables we consider. One can find more detailed descriptions and forms of listed features in Harry and Fairhurst (2011).

<table>
<thead>
<tr>
<th>Input variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single detector’s SNR, ( \rho_{\text{IFO}} )</td>
<td>Signal-to-noise ratio (SNR) value obtained from each of the detectors’ data where IFO = H1, L1, or V1. In this work, we have ( \rho_{H1}, \rho_{L1}, \text{and} \rho_{V1} ).</td>
</tr>
<tr>
<td>Coherent SNR, ( \rho_{\text{coh}} )</td>
<td>Coherent combination of single detector’s SNRs</td>
</tr>
<tr>
<td>Coherent ( \chi^2 )-test value</td>
<td>Mitigating non-Gaussian noise contribution by testing the differences between template waveforms and instrumental/environmental noise</td>
</tr>
<tr>
<td>New SNR, ( \rho_{\text{new}} )</td>
<td>Filtered ( \rho_{\text{coh}} ) by checking whether the ( \chi^2 )-test value is larger or smaller than the number of degrees-of-freedom of ( \chi^2 ) statistic</td>
</tr>
<tr>
<td>Coherent bank ( \chi^2 )-test value</td>
<td>Testing the consistency of the observed ( \rho_{\text{coh}} ) over different template waveforms in the template bank at the time of signal candidate trigger</td>
</tr>
<tr>
<td>Coherent auto-correlation ( \chi^2 )-test value</td>
<td>Testing the consistency of the observed ( \rho_{\text{coh}} ) over SNR time series around the trigger in the template bank at the time of signal candidate trigger</td>
</tr>
<tr>
<td>Masses</td>
<td>Component masses of a binary system, ( m_1 ) (NS or BH) and ( m_2 ) (NS)</td>
</tr>
</tbody>
</table>
estimated FAP. On the other hand, for signal consistency test, the coherent search conducts \( \chi^2 \)-test. In specific, coherent \( \chi^2 \)-test, which is a sum-squared measure of the difference between the data and the template waveform, is used to eliminate or down-weight triggers with large \( \chi^2 \) values and, consequently, mitigate the effect of non-Gaussian background noise. For more details about statistical quantities and signal consistency tests, one can refer to table 3 and Harry and Fairhurst (2011).

The coherent search pipeline also adds simulated waveforms\(^{13}\) to the off-source segment\(^{14}\) and this procedure is referred to software injections. Then the search pipeline repeats the matched filtering on the software injected waveforms. The simulated waveforms for the software injection are generated by the spinning TaylorT4 waveform (Boyle et al. 2007) code in the LIGO Algorithm Library\(^{15}\) with correction terms up to the 3.5 post-Newtonian order. In the generation of those simulated waveforms, several physical parameters such as the masses and spins of component objects (NS or BH), and the distance to the expected binary system need to be chosen: the ranges of masses are chosen to be 1–3 \( M_\odot \) for NS and 2–25 \( M_\odot \) for BH and the ranges of spin magnitudes are chosen to be \([0, 0.4]\) and \([0, 0.98]\) for NS and BH, respectively. Also, for the simulated waveforms, the inclination angle is considered such that the inclination angle to be distributed within the four possible sizes, \(10^\circ, 30^\circ, 45^\circ, \) and \(90^\circ\), which are the half-opening angle of cone-shaped outflow of a short GRB. The details of choosing physical parameters and their distributions are discussed in appendix A. From this procedure, we get software injection triggers which are separated into two categories. If there is a trigger found within 100 ms of the time of simulation, we call that trigger as a found injection trigger (Predoi 2012). While, if not, it becomes a missed injection trigger.

2.2. Training and evaluation samples

We configure two different sample sets, training and evaluation samples for the ANN. For the performance test, which is done by evaluating known samples in the off-source segment using the trained ANN, of section 3, we take the found injection triggers and off-source triggers as the sets of signal samples, \( X_S \), and background samples, \( X_B \), respectively. With this notation, we define similar notation \( x^S \) and \( x^B \) to denote each sample of \( X_S \) and \( X_B \), respectively, such as

\[
X_S = \left\{ x_S^l; l = 1, 2, \ldots, N_S \right\},
\]

\[
X_B = \left\{ x_B^m; m = 1, 2, \ldots, N_B \right\},
\]

where \( N_S \) and \( N_B \) correspond the total number of the signal samples and the background samples, respectively. However, we find that the GPS times of some found injection triggers are also found in off-source triggers. So, we discard those found injection triggers from the configuration of signal samples. Note that, in the preparation of samples, the missed injection triggers are discarded because they are improper for the purpose of training ANN.

\(^{13}\) The template waveforms also can be called as simulated waveforms. However, the template waveforms are used only for the matched filtering. Thus, we use the term of simulated waveform only for the software injection to prevent confusion.

\(^{14}\) The on-source segment is not used in order to avoid any contaminations in the potential candidate GW event in the on-source segment.

\(^{15}\) https://www.lsc-group.phys.uwm.edu/daswg/projects/lal/nightly/docs/html/.

Class. Quantum Grav. 32 (2015) 245002 K Kim et al
For the configuration of the multi-dimensional sample data, we take the 8 statistical and 2 physical quantities from the CBC-GRB search pipeline as input variables, since they are considered important in discriminating signals from noise and in characterizing the GW sources: three single detectors’ SNRs ($\rho_{H1}$, $\rho_{L1}$, and $\rho_{V1}$), coherent SNR ($\rho_{coh}$), coherent $\chi^2$-test value, new SNR ($\rho_{new}$) of the coincidence search, coherent bank $\chi^2$-test.
value, coherent auto-correlation $\chi^2$-test value, and recovered values of component masses, $m_1$ (NS for BNS model or BH for NS-BH model) and $m_2$ (NS) for each trigger. The selected input variables are tabulated in table 3 with brief descriptions (For more details, see Harry and Fairhurst 2011). We plot the scatter plots of input variables in figure 2. From the figure, one can see that the distributions of statistical quantities have similar shapes between the NS-BH and the BNS model. The mass-related distributions for the signal samples are different, as expected given the injected ranges on masses for NS-BH versus BNS systems. However, one can see that the distributions of masses for BNS model covers a bit wider range than the mass ranges of the injection because the points in the scatter plots are obtained by the recovered values by the search pipeline. Even though, we see that many of mass values are distributed within the range of injection. The background samples are exactly the same for both NS-BH and BNS models because we use the same bank of template waveforms for both models. Figure 2 shows that it is almost impossible to find simple and appropriate border line between the distributions of signal samples and background samples.

3. Classification performance test using artificial neural network

3.1. Artificial neural network

The ANN (Hecht-Nielsen 1989, Hastie et al 2009) is a widely used machine learning algorithm based on mimicking the biological neural system, which has been designed for artificial intelligence. This algorithm can be simplified with some mathematical models with data-driven input and output samples.

The implemented mathematical model includes nodes, a network topology, and learning rules adopted to a specific data processing task. Nodes are described by their number of inputs/outputs and the connection weights associated with each input and output. The network topology is closely related to the connections between the nodes. The learning rules represent how the connection weights are optimized. A node can be activated if the summed value of input nodes exceeds its threshold value.

Among various models of ANNs with different topologies, the multi-layered perceptron model (Rosenblatt 1961) is widely used for its efficient classification. The model is composed of input and output layers as well as a few hidden layers in between. We present a simple schematic example of a network topology in figure 3. For the nodes $x_i$ in the input layer, the input variables of an input sample are used. Those nodes in the input layer are connected to the nodes $y_j$ in the adjacent hidden layer with different connection weights $w_{ij}$. The value of $w_{ij}$ is given randomly at the initial stage and updated by the improved resilient back propagation (iRPROP) algorithm (appendix B) by repeating iterations. Then the value of $y_j$ is determined by an activation function $f$ as

$$y_j = f(N_j),$$

and $N_j$ is defined by a linear combination of the $x_i$ and $b$ such as

$$N_j = \sum_{i=1}^{i} x_i \cdot w_{ij} + b \cdot w_{0j},$$

where $b$ denotes the bias node that makes the node $y_j$ activated and $w_{0j}$ indicates the connection weight between the bias node in the input layer and the node $y_j$. The activation function can be chosen in various options such as sigmoid, piecewise linear, step, and gaussian functions. Here, we choose the sigmoid function.
which is the activation function chosen in Biswas et al (2013). From equation (9), one can see that the property of this sigmoid function is determined by the steepness \( s \). Also, if many hidden layers are given in a network topology, similar processes are repeated until the connections will converge in the output layer. Then, the output node, \( z_k \), is determined by

\[
    z_k = f\left( N_k \right) = \left( 1 + e^{-2sN_k} \right)^{-1},
\]

where \( N_k \) is the input to the output layer, which is the sum of the weighted inputs from the hidden layer.

In each layer, the learning algorithm finds the optimal connection weights between nodes. We particularly use the iRPROP algorithm (Igel and Hüskens 2000), which minimizes the error between the output and the goal values: 1 for signal samples and 0 for background samples in our case. We choose the fast artificial neural network (FANN) library package\(^\text{16}\) for machine learning and find optimal connection weights by controlling parameters such as the number of layers, the number of nodes in each layer, and so on, which are given in the library.

We construct a simple network topology with one input layer, one hidden layer, and one output layer. For our feature space of 10 input variables, we let 10 nodes be placed in the input layer and put the same number of nodes in the hidden layer. In the output layer, we have only 1 output node. Bias nodes in the input and the hidden layers are systematically placed by the FANN library and the steepness \( s \) in the activation function equation (9) is set to be 0.5 for simplicity.

\(^{16}\) To see the detailed features of this library and get the source, visit http://leenissen.dk/fann/wp/
When the network topology and other training parameters are fixed, we can train the ANN with the training samples which have preassigned classes either 1 for a signal sample or 0 for a background sample. The goal of the training process is to find the most optimal set of connection weights such that the error between the values assigned by the ANN and the target values of samples is minimized. In this work, the error is represented by the mean-squared-error (MSE) that is defined as

\[ \text{MSE} = \frac{1}{N} \sum_{k=1}^{N} (z_k^t - z_k^o)^2, \]  

where \( N \) is the total number of samples and \( z_k^t \) and \( z_k^o \) are the target value of the output and the observed value of a sample, respectively, of a sample. In this work, \( z_k^t \) is the class (1 or 0) of a sample. Meanwhile, \( z_k^o \) is determined by equation (10) and it has a value between 0 and 1. We terminate the training process when MSE stops decreasing and reaches a plateau with small oscillation around it as the iteration evolves.

Samples of known and unknown class can be evaluated deterministically by the trained ANN. The final \( z_k \) value obtained after this evaluation process corresponds to the prediction of whether a sample is a signal or not. Hereafter, we call the final \( z_k \) as the rank, \( r \), of a sample.

3.2. Results

We introduce the receiver operating characteristic (ROC) curve in order to interpret the result. The ROC curve is obtained by calculating the efficiency and the FPR which are defined as

\[ \text{Efficiency}(R) \equiv \frac{N_S(R)}{N_S}, \]  
\[ \text{FPR}(R) \equiv \frac{N_B(R)}{N_B}, \]

where \( R \) denotes a threshold chosen among the ranks of evaluation samples. \( N_S(R) \) and \( N_B(R) \) in the numerators of equations (12) and (13) are defined as

\[ N_S(R) \equiv \left\{ x_l^S(r); r \geq R, l = 1, 2, \ldots, N_S \right\}, \]  
\[ N_B(R) \equiv \left\{ x_m^B(r); r \geq R, m = 1, 2, \ldots, N_B \right\}, \]

respectively, that is, the number of evaluated signal samples and background samples with scored ranks, \( r \), exceeding a criterion rank value, \( R \) among all \( x_l^S(r) \) and \( x_m^B(r) \).

When we compute equations (12) and (13), we only have \( \sim 2,000 \) signal samples and \( \sim 7,000 \) background samples for the denominator \( N_S \) and \( N_B \), respectively. Therefore, if the numerator \( N_B(R) \) is 1, the minimum value of FPR becomes \( 10^{-4} \) due to the number of background samples.

In order to fairly evaluate our classification efficiency, we split the full set of signal and background samples prepared in the previous section into training and evaluation samples via a round-robin process (see appendix C for details). Then we train the ANN with the training samples and test the classification performance with the evaluation samples. The training process ends with the MSEs reaching 0.08 and 0.06 for NS-BH model and BNS model, respectively. We find that each of the trained ANNs result in different ranks on some of evaluation sample depending on the randomly given initial connection weights. Therefore, we repeat the training process 100 times in order to see distribution of rank and to obtain a representative statistical quantity for a given evaluation sample. The method for obtaining
A representative statistical quantity will be discussed later. From the repeated trials, it is shown that the resulted MSEs are similar to the values stated above for all trials of both NS-BH and BNS cases. Then we evaluate the evaluation samples with each trained classifier. In figure 4, we present histogram and normalized, inverted cumulative histogram of ranks scored on evaluation samples by using the result of evaluation trial \#1 out of 100 trials. One can see that the scored ranks of both signal and background samples are widely spread between 0 and 1. This result shows that the ANN does not classify the given data clearly. On the other hand, with the cumulative histogram, we can easily expect the shape of ROC curve which will be discussed in below.

We show the ROC curve of each trial in figure 5. The ROC curves are given in the figure 5 by calculating equations (12) and (13) by varying $R$ in a range from $R_{\text{min}}$ with $N_b(R_{\text{min}}) = N_b$ to $R_{\text{max}}$ with $N_b(R_{\text{max}}) = 1$. In particular, if $N_b(R) = N_b$, equation (13) becomes 1 and this special point corresponds to the point at the upper-right corner of the ROC curves. On the other hand, $N_b(R) = 1$ corresponds to the leftmost point, i.e., the minimum FPR of each curve. This variation in $R$ is consistently applied to both $N_s(R)$ and $N_b(R)$. As one can see from the ROC curves in figure 5, when we repeat the training and evaluation processes many times (100 trials in our case), we find that there are large variations in the efficiency below FPR of $\sim 10^{-3}$ of each run: We find that when (i) the maximum rank of all background samples is greater than that of all signal samples or (ii) the number of signal samples exceeding the maximum rank of background samples is small, the efficiency at the minimum FPR is significantly decreased. On the other hand, when (i) the maximum rank of background samples is smaller than the maximum rank of signal samples and (ii) there are many signal samples exceeding the maximum rank of background samples, the efficiency can be increased at the minimum FPR.

Simple statistics such as mean and standard deviation can hardly represent either the distribution of efficiencies at the minimum FPR or the distributions of ranks for the given samples as shown in figure 6. Therefore, in this work, we adopt the maximum likelihood ratio (MLR) which is shown to be an optimal method in obtaining a representative quantity for GW data (Biswas et al 2012, 2013).
The MLR for an \( n \)th sample can be calculated by

\[
\lambda(n) = \max_{\alpha} \left\{ \int_{r_n^{(\alpha)}}^{1} P_1(r_n^{(\alpha)}) \, dr_n^{(\alpha)} - \int_{0}^{r_n^{(\alpha)}} P_0(r_n^{(\alpha)}) \, dr_n^{(\alpha)} \right\} = \max_{\alpha} \left\{ \frac{P_1(r_n^{(\alpha)})}{P_0(r_n^{(\alpha)})} \right\}
\]

(16)

among \( r_n \)'s of \( \alpha \)-trials where \( \alpha \) denotes the trial (\( \alpha = 1, 2, \ldots, 100 \)) and \( r_n^{(\alpha)} \) indicates the rank of \( n \)-th sample in a trial \( \alpha \). As one can see from equation (16), we find the maximum value of the ratio between the probability of correctly finding a true signal \( (P_1) \) and the probability of finding a false signal \( (P_0) \). Thus, in our case, equation (16) can be rewritten as

\[
\lambda(n) = \max_{\alpha} \left\{ \frac{\text{Efficiency}(r_n^{(\alpha)})}{\text{FPR}(r_n^{(\alpha)})} \right\}.
\]

(17)
In figure 7, we plot the distribution of MLRs. From this figure, one can see that the background samples have finite values of MLR. However, for signal samples, about half of the samples have similar values of MLR to the values of background samples and the rest of them are separated from the range of finite MLRs and take an infinite value. This tendency is consistently shown in all considered cases. The infinity MLR of signal sample can be easily derived from equation (16), i.e., if there are no background samples exceeding $r_0$, then $\text{FPR}(r_0)$ becomes 0 and it leads $\lambda(r_0)$ to be infinity. Also, when we compare this figure 7 to the histogram of ranks in figure 4, we see that the classification efficiency with the MLR is enhanced by the fraction of clearly separated signal samples.

With the MLR, the efficiency and FPR of equations (12) and (13) are changed to

\[
\text{Efficiency}(\Lambda) \equiv \frac{N_S(\Lambda)}{N_S},
\]

\[
\text{FPR}(\Lambda) \equiv \frac{N_B(\Lambda)}{N_B},
\]

Figure 7. The histogram of the maximum likelihood ratios of the signal and background samples. The blue bars and red bars denote the signal samples and the background samples, respectively. One can see that about half of signal samples are clearly separated from the distributions of background samples for all considered cases.
where

\[ N_S(\lambda) \equiv \left\{ x^S_\lambda; \lambda \geq \lambda, l = 1, 2, \ldots, N_S \right\}, \]

\[ N_B(\lambda) \equiv \left\{ x^B_\lambda; \lambda \geq \lambda, m = 1, 2, \ldots, N_B \right\}. \]

The ROC curves with equations (18) and (19) are drawn in figure 8 by varying \( \lambda \) from \( \lambda = \lambda^B_{\text{min}} \) to \( \lambda = \lambda^B_{\text{max}} \). From this figure, one can see that the ROC curves with the MLR are more or less similar to the maximum efficiency calculated via the rank given in figure 5. Also, we see that MLR-aided ANN’s performances are improved by 8%–12% compared to the detection statistic as expected from the histogram of figure 7. Therefore, we conclude that the MLR-aided ANN can improve the classification performance.

Meanwhile, when we compare NS-BH model and BNS model of each GRB data, we find that the classification efficiencies of both MLR-aided ANN and the detection statistic for the BNS model are better than the efficiencies for the NS-BH model. In order to find reason of this difference, we look the scatter plots of \( m_1 \) versus \( \ell_{\text{coh}} \), \( m_1 \) versus \( \chi^2_{\text{coh}} \), and/or \( m_1 \) versus \( m_2 \) in figure 2. From these scatter plots of NS-BH model, one can see that the \( m_1 \) parameters of signal and background samples are distributed almost in the same region. On the other
hand, those scatter plots of the BNS model show that the $m_1$ parameter of signal samples has a rather squeezed distribution: the distribution of the $m_1$ parameter for signal samples is related to the setup on the mass parameters for the waveform models; whereas, for background samples, the distribution of $m_1$ is caused by the fact that background noises are predominant at low frequencies and so have better overlaps with high-mass templates than with low-mass templates in the generation of off-source triggers. Therefore, it is expected that we can obtain better performance when a sophisticated classification method, such as the MLR-aided ANN or the detection statistic, is applied to samples which have visibly distinguishable input variables, e.g., $m_1$ for BNS model in this work.

### 3.3. Test for an important input variable

In the previous section, we see that if the distributions of signal and background samples are visibly distinguishable through an input variable, we expect to obtain better efficiency with the MLR-aided ANN at a fixed FPR. So, in order to study more on this expectation, we perform a test called the importance test via the sum of fully connected weights, $W$, for 10 input variables. $W$ is defined as

$$W \equiv \sum_j (w_{ij} w_{jk}).$$

For calculation of $W$, we use the connection weights which are determined in the training process. As an example of this test, we present the top two variables and its connection.

**Figure 9.** Top two input variables obtained by the importance test for training trial #1.
weights for training trial #1 of each waveform model for GRB070714B data in figure 9. We see that $\chi^2_{\text{coh}}$ is the most significant variable for both NSBH and BNS models because it’s significance is top in 4 round-robin sets and 5 round-robin sets for NSBH and BNS, respectively. On the other hand, second top significant variable is $\rho_{\text{coh}}$ for NSBH model and $m_1$, as expected, for BNS model. However we also find that the most significant variables are different in other trials, e.g., for trial #2 of NSBH, $\chi^2_{\text{coh}}$ is top in 1 round-robin set (set 3) and $\rho_{\text{coh}}$ is top in 3 round-robin sets (set 1, 4, and 9) and, for trial #2 of BNS, $\chi^2_{\text{coh}}$ is top in 3 round-robin sets (set 2, 4, and 6) and $m_1$ is top in 2 round-robin sets (set 7 and 9). Therefore, we conclude that taking all considerable input variables is better than introducing a specific input variable selection (IVS) rule for CBC-GRB data for generality.

4. Applications

4.1. Detection sensitivity on distance

So far, we have demonstrated the applicability of ANNs to classifying the samples generated by running the CBC-GRB coherent search pipeline on the selected data segments and have demonstrated the ANN’s classification performance on the data classification via plotting the ROC curves.

In this section, we demonstrate the results of performance test in terms of an astrophysical observable, i.e., the distance. It is possible to estimate the detection sensitivity as a function of the distance since we have set the distribution of possible distance range for the simulated signals.

Firstly, we define the number of found injection samples that is exceeding a criterion MLR, $\Lambda_T$, in a $l$th distance bin $[D_l, D_{l+1})$ such as

$$N_{\text{inj}}^{\text{found}}(\Lambda_T)|_{[D_l, D_{l+1})} = \{\chi_{\text{inj}}^{\text{found}}(\lambda) ; \lambda \geq \Lambda_T\}|_{[D_l, D_{l+1})}. \tag{23}$$

Here, we set $\Lambda_T$ to give 0.1% of FPR. With equation (23), we define the fraction of correctly classified signal samples (or, simply, the detection probability), $P$, for each distance bin, $[D_l, D_{l+1})$, such as

$$P(\Lambda_T)|_{[D_l, D_{l+1})} \equiv \frac{N_{\text{inj}}^{\text{found}}(\Lambda_T)|_{[D_l, D_{l+1})}}{N_{\text{inj}}^{\text{tot}}(\Lambda_T)|_{[D_l, D_{l+1})}}. \tag{24}$$

where the denominator denotes the total number of signal samples in a given distance bin and it is defined as

$$N_{\text{inj}}^{\text{tot}}(\Lambda_T)|_{[D_l, D_{l+1})} \equiv N_{\text{inj}}^{\text{found}}(\Lambda_T)|_{[D_l, D_{l+1})} + N_{\text{inj}}^{\text{missed}}(\Lambda_T)|_{[D_l, D_{l+1})}. \tag{25}$$

In order to get the appropriate numbers of both terms, $N_{\text{inj}}^{\text{found}}(\Lambda_T)$ and $N_{\text{inj}}^{\text{missed}}(\Lambda_T)$ in the right-hand side of equation (25), we use the found injection triggers and missed injection triggers, respectively. We can see that if there are no missed triggers in a distance bin, namely, $N_{\text{inj}}^{\text{missed}}$ equals zero, then

$$N_{\text{inj}}^{\text{tot}}(\Lambda_T) = N_{\text{inj}}^{\text{found}}(\Lambda_T) \tag{26}$$

is easily derived from equation (25) and the right-hand side of equation (24) becomes 1. With the calculated detection probability, $P(\Lambda_T)$, we draw the detection probability versus distances to the progenitor, as presented in figure 10 for the given distance bins. One can see that the signal samples injected in the close distance range ($\leq 10$ Mpc) are mostly found injection events and the number of missed injection events gradually increases as the distance
becomes larger. We also draw error bars assuming a binomial distribution with 1-σ confidence interval for each plotting point.

From figure 10, we clearly see that the ANN is a more efficient tool in detecting short GRBs at large distances for a fixed FPR. The detectable distances at 50% and 90% of

Table 4. Distances where 50% and 90% of detection probability, \( P(\Lambda_T) \), obtained with the conventional detection statistic and the MLR-aided ANN. As in figure 10, we only consider injection triggers which have FPRs equal to or smaller than 0.1%.

<table>
<thead>
<tr>
<th>Data</th>
<th>Waveform type</th>
<th>Distance at 50% of ( P(\Lambda_T) )</th>
<th>Distance at 90% of ( P(\Lambda_T) )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>DetStat</td>
<td>ANNs</td>
</tr>
<tr>
<td>GRB070714B</td>
<td>NS-BH</td>
<td>35.0 Mpc</td>
<td>40.0 Mpc</td>
</tr>
<tr>
<td></td>
<td>BNS</td>
<td>18.0 Mpc</td>
<td>20.4 Mpc</td>
</tr>
<tr>
<td>GRB070923</td>
<td>NS-BH</td>
<td>41.7 Mpc</td>
<td>45.9 Mpc</td>
</tr>
<tr>
<td></td>
<td>BNS</td>
<td>21.3 Mpc</td>
<td>22.9 Mpc</td>
</tr>
</tbody>
</table>
detection probability are summarized in table 4. The distances in this table mean that, if a short GRB event occurs within this distance, the probability of detecting the GW signal associated with the GRB event is 50% or 90% at least. Note that the distances at 50% of detection probability can be used in the estimation of detection rates such as done by Abbott et al. (2007). On the other hand, the distances at 90% of detection probability summarized in table 4 are not the exclusion distance to the co-progenitor of a GW signal and a short GRB event. For the calculation of the exclusion distance, we need to consider not only errors in calibration of GW data but also general populations for possible GW signals: the standard CBC-GRB search considers 10°, 30°, 45°, and 90° for the inclination angle in the generation of simulated waveforms. However we don’t take account the errors in calibration of used GW data and have restricted the population to be limited by the inclination angle which is set to be less than 10°. Thus, it is hard to directly compare our results to the exclusion distances given in the result of the previous search (Abadie et al. 2010b).

4.2. Evaluation of unknown triggers

We would like to extend this work to the classification of unknown triggers. For this purpose, we use some triggers in a randomly chosen 6 seconds block of the buffer segment17 to mimic an on-source segment18 and evaluate them with the trained ANNs in the previous section. We believe that no triggers in the buffer segment are associated with any GRB.

The coherent search pipeline finds 37 triggers and 27 triggers, respectively, in each of the selected blocks in buffer segments of GRB070714B and GRB070923. We evaluate these triggers 100 times with the 100 sets of different connection weights as done in the previous section and then calculate the MLR with the ranks. The distribution of calculated MLRs of the unknown triggers are plotted in figure 11. When we compare this figure with figure 7, it looks like all unknown triggers, even the loudest triggers, can be seen as either signal or background samples for both NS-BH and BNS model cases. From this fact we conclude that discrimination of a signal from background based on the value of the loudest MLR by comparing it with the MLR values of background triggers is inefficient.

Meanwhile, the loudest triggers of the NS-BH and BNS models are different (7th and 35th triggers, respectively, among 37 triggers in the GRB070714B buffer segment and 7th and 15th triggers, respectively, among 27 triggers in the GRB070923 buffer segment). We examine these triggers more precisely by estimating the FAPs as drawn in figure 12. In the calculation of FAP, we use equation (19) with a reduced entry for the numerator, \( N_B(L) \), by taking the loudest off-source trigger of each 6 s block in the off-source segment as done in Abadie et al. (2010b)19. Here, we pick the loudest off-source trigger of each 6 s block based on the value of MLR. From this selection, the number of reduced background samples \( N_B' \) becomes 324 and equation (21) changes to

\[
N_B'(N) = \left\{ X^B_q(N) ; X \geq N, q = 1, 2, \ldots, 324 \right\}.
\]  

17 The buffer segment is adopted to prevent biasing our background estimation due to a potential loud signal in the on-source segment (Fotopoulos 2010). Therefore the buffer segment is placed between the on- and off-source segment.

18 In the real CBC-GRB search, we do not know whether the on-source triggers are real GW signals or noise artifacts. Thus we intend to mock up the situation.

19 In Abadie et al. (2010b), the authors were interested in the existence of GW signal within the 6 s on-source segment via examining the most significant trigger. Thus, they divided the off-source segment into 6 s long 324 trials to estimate the distribution off background due to the accidental coincidences of noise triggers.
We draw the FAPs of background samples as grey lines in figure 12 using equation (19) by varying from $N = \lambda^{\min}_{B} \text{ to } N = \lambda^{\max}_{B}$. From figure 12, we see that the FAPs of the loudest unknown triggers are placed on the line of FAPs of the background samples. If they were real GW signals, the loudest triggers should be placed out of the line of FAPs of the background samples, i.e., it should be placed in more left and upper area of the minimum FAP\textsuperscript{20}. This result shows that those loudest unknown triggers are less significant than the loudest background sample and are not GW signals. Also, the calculated FAPs are 0.20 (NS-BH) and 0.04 (BNS) for the loudest triggers in the GRB070714B buffer segment and 0.27 (NS-BH) and 0.69 (BNS) for the loudest triggers in the GRB070923 buffer segment. Therefore, we conclude that we find no significant trigger in the selected 6 seconds block. Also, we suggest a simple criterion that if the estimated FAP of the loudest unknown trigger is

\textsuperscript{20} For this case, we need more precise follow-up analysis, e.g., testing correlation with known background events.
smaller than 0.1% (or 3-σ in the normal distribution, equivalently) of FAP, the trigger is likely a signal candidate.

5. Summary and discussion

In this work, we discuss the improvement of the search performance for GW candidate events related to short GRBs by using ANN algorithm compared to a conventional detection statistic. With this demonstration, we aim to increase the search sensitivity on GWs associated with short GRBs.

We use the GW data obtained by the LIGO and Virgo detectors during S5 and VSR1 and take short GRBs, GRB070714B and GRB070923 as test samples. By using the coherent CBC-GRB search pipeline, on-source, off-source, and software injection triggers are generated. For the generation of the software injection triggers, we consider both NS-BH and BNS binaries as the progenitor of a GRB for the determination of component masses. We set the distributions of the distances to the progenitors to have different ranges depending on the detectors’ responses at the given event time and/or on the type of binary system. We train
ANNS with taking found injection triggers as signal samples and off-source triggers as background samples. Then, we evaluate test samples with the trained ANNs. Each sample for both training and evaluation is characterized by the 8 statistical quantities and 2 physical quantities listed in table 3.

It appears that the performance on data classification varies significantly between the trained ANNs. We find that this variation results from the statistical variance in the ANN algorithm because ANN uses a randomly distributed initial input configuration. Therefore, we need to mitigate the variation in order to get a reliable interpretation on our results. For the mitigation of the statistical variance, there are several possible ways of reducing the variance, for example, calculating ensemble average of ranks, taking median value, and computing MLR. Among those, we adopt MLR because the other two methods are not suitable to the data as shown in figure 6. Moreover, the MLR method has already been shown to be an optimal method in obtaining a representative quantity for GW data in Biswas et al (2012) and Biswas et al (2013). With this prescription, we resolve the variance in ranks and the classification performance by combining the results from 100 trials.

As a result of the performance test, we see that the background samples have finite values of MLR from the distribution of MLRs given in figure 7. However, for signal samples, about half of samples have similar values of MLR to the values of background samples and the rest of them are separated from the range of finite MLRs and take an infinite value. This tendency is consistently shown in all considered cases. When we compare this figure 7 to the histogram of ranks in figure 4, we see that the classification efficiency with the MLR is enhanced by the fraction of clearly separated signal samples.

Also, we examine the improvement in the classification performance by comparing the ROC curves, instead of directly comparing the ranks of two ranking methods, because there is no rule to connect two different ranks, scored independently by the conventional detection statistic or by the ANN. When we look the ROC curves, we find that the ROC curve obtained by MLR is more or less similar to the maximum efficiency calculated via the ANN’s rank given in figure 5. From the comparison of the ROC curves between the one obtained by using the conventional detection statistic and the other one obtained by using the ANN, we see that the data classification efficiency at the minimum FPR is improved by 8%–12% as expected from the histogram of figure 7. Therefore, we conclude that the MLR-aided ANN can improve the classification performance.

Meanwhile, when we compare the NS-BH model and BNS model of each GRB data, we find that the classification efficiency of the BNS model is better than that of the NS-BH model. To understand this difference, we look the scatter plots of $m_1$ in figure 2. From the scatter plots of the NS-BH model, one can see that the $m_1$ parameters of signal and background samples are distributed almost in the same region. On the other hand, the scatter plots of the BNS model show that the $m_1$ parameters of signal samples have different distribution, that is, rather squeezed distribution compared to the distribution of background samples: the distribution of $m_1$ parameter for signal samples is related to the setup on the mass parameters for the waveform models; whereas, for background samples, the distribution of $m_1$ is caused by the fact that background noises are predominant at low frequencies and so have better overlaps with high-mass templates than with low-mass templates in the generation of off-source triggers. Therefore, it is expected that we can obtain better performance when a sophisticated classification method, such as the MLR-aided ANN or the detection statistic, is applied to samples which have visibly distinguishable input variables, e.g., $m_1$ for BNS model in this work.

We extend the results of the MLR-aided ANN’s classification performance to the estimation of the sensitivity of analysis as a function of the distance which denote the upper
bound of the distance where we can observe a GW event within the given distance. In this work, it is possible to estimate this quantity since we have set the distribution of possible distance range for the simulated signals. For this estimation, we assume that mass parameters are marginalized. From the results presented in figure 10 and table 4, we find that the distance at 50% detection probability can be increased by $\sim 7.9\% - 14.3\%$ and the distance at 90% detection probability can be increased by $\sim 5.7\% - 16.3\%$ compared to the distance calculated with the conventional detection statistic. This means that if a GW event occurs at a distance greater than the distance estimated by the conventional detection statistic and smaller than the distance estimated by the MLR-aided ANN, we can identify that event with MLR-aided ANN. Therefore, it is shown that the estimated sensitivity of analysis obtained by using the MLR-aided ANN’s classification results allows us to observe more events occurring at a farther distance than the conventional method.

We apply our analysis to the data in the buffer segment of the two GRBs. From the evaluation of triggers in the buffer segment, we see that the loudest triggers are placed on the line of FAPs of the background samples as shown in figure 12. If they were real GW signals, the loudest triggers should be placed out of the line of FAPs of the background samples, i.e., it should be placed in more left and upper area of the minimum FAP. Therefore, this result shows that those loudest triggers are less significant than the loudest background sample and they cannot be GW signals. The estimated FAPs are 0.20 (NS-BH) and 0.04 (BNS) for the loudest triggers in the GRB070714B buffer segment and 0.27 (NS-BH) and 0.69 (BNS) for the loudest triggers in the GRB070923 buffer segment. Therefore, it can be concluded that we find no significant event with the analysis with MLR-aided ANN on the buffer triggers. Also, we suggest a simple criterion that if the estimated FAP of the loudest unknown trigger is smaller than 0.1% (or more probable than 3-$\sigma$ in the normal distribution) of FAP, the trigger is likely a signal candidate.

Throughout this work, we investigate the feasibility of application of ANN to CBC-GRB search as a new ranking method and we find that it can improve the search performance and the detectable distance. Therefore, we would suggest that the artificial neural network can be a complementary method to the conventional detection statistic for identifying gravitational-wave signals related to the short GRBs.

We however have a limit obviously in arguing the robustness and the consistency of this approach since we have tested only four-cases (two GW data $\times$ two binary models). This fact means that we need to test more data. As a possible way of obtaining more test data, we may consider to use GW data for other targeted GRBs summarized in (Abadie et al 2010b, 2012a, Aasi et al 2014). Or, alternatively, we may choose random sky locations and random times to generate triggers, instead of restricting our focus to known GRBs’ event times and sky locations, because we know that GRBs are isotropically distributed in the sky. So, in a future work, we will test more data to examine the robustness of this approach and will discuss the general characteristic of this analysis method.

Acknowledgments

We thank the LIGO Scientific Collaboration and the Virgo Collaboration for the use of the data. We are also grateful for computational resources provided by the Leonard E Parker Center for Gravitation, Cosmology and Astrophysics at University of Wisconsin-Milwaukee (NSF-0923409). The authors would like to thank S Bose, K Cannon, T Dent, C Hanna, H M Lee, C Kim, and R Vaulin for helpful comments and useful discussions. KK would like to specially thank J Burguet-Castell, A Dietz, and N Fotopoulos for suggesting the initial
motivation of this work. KK, YMK, CHL, HKL, JJO, SHO, and EJS were supported in part by the Global Research Network program of the National Research Foundation (NRF) funded by the Ministry of Science, ICT, and Future Planning of Korea (MSIP) (NRF-2011-220-C00029). KK, YMK, CHL, HKL, JJO, SHO, and EJS were also supported in part by the Global Science experimental Data hub Center (GSDC) at KISTI. The work of YMK and CHL was also supported by the NRF funded by the MSIP (NRF-2015R1A2A2A01004238).

Appendix A. Physical parameters of simulated waveforms for software injection

In this section, we need to corroborate some physical parameters such as mass, distance, and spin that are used in the generation of simulated waveforms. These parameters represent various types of GW progenitors and the property of waveform itself. For the mass parameter, we consider two appropriate binary systems such as BNS and NS-BH as the sources of simulated signals based on the predicted event rates (Abadie et al 2010a). For the neutron stars in the BNS system, their masses are given in range of \(1–3M_\odot\) with mean mass, \(m_{\text{NS}} = 1.4 M_\odot\) and standard deviation, \(\sigma_{\text{NS}} = 0.2 M_\odot\) with the assumption of the Gaussian distribution. Similarly, for the NS-BH system, the ranges of component masses are set as \(m_{\text{NS}} = 1–3M_\odot\) (mean mass), for neutron star and \(\sigma_{\text{NS}} = 0.4 M_\odot\) and \(m_{\text{BH}} = 2–25 M_\odot\) (with \(m_{\text{BH}} = 10 M_\odot\) and \(\sigma_{\text{BH}} = 6.0 M_\odot\) ) for black hole under the assumption of the Gaussian distribution too (for details, see Abadie et al 2012a).

One can expect from section 2.1 that the used waveforms for this software injection contain spin effect on the contrary to the template waveforms for the matched filtering process. With observed information and models (Hessels et al 2006, Mandel and O’Sullivan et al 2006, Dietz2011) we assume possible ranges of spin magnitudes for NS and BH as \([0,0.4]\) and \([0,0.98]\), respectively. Therefore, we set possible values of them to be uniform in given ranges with random orientation.

When we take the spinning waveform into account, the related important parameter for describing a binary system is the inclination angle which describes the size of angle between the direction of the total angular momentum and the line of sight to the observer because the strength of a GW to a detector can be varied depending on the angle. Many observations and models suggest that a GRB is generated by cone-shaped outflow from a CBC system (Burrows et al 2006, Grupe et al 2006, Dietz 2011). From these, it is supposed that when the inclination angle is placed within the cone it is possible to find a GW signal directly related to a short GRB. Thus, the pipeline uses four different half-opening angles, \(10^\circ, 30^\circ, 45^\circ, \) and \(90^\circ\) as possible sizes of cone and it allows the inclination angle to be distributed within the cone (for more details refer to Abadie et al 2012a). We only take \(10^\circ\) for simplicity.

The range of distances to the progenitors are differently determined by the type of progenitor and applied with taking the detectors’ sensitivities on given GRB’s sky location into account too. To estimate the sensitivity, first, we need to calculate the antenna response, \(F\) (Allen et al 2012):

\[
F = \left( F_+^2 + F_\times^2 \right)^{1/2},
\]

where \(F_+\) and \(F_\times\) are the antenna factors of a detector for the ‘+’- and ‘×’-polarizations of a GW signal, respectively (for more details see Allen et al 2012). These \(F_+\) and \(F_\times\) represent the amount of sensitivity on each polarization of the incident GW signal. A GRB’s event time and its sky location play an important role in the determination of the values of \(F_+\) and \(F_\times\). The determined values of \(F_+\) and \(F_\times\) of each detector for the selected GRB are summarized in table 2. In particular, if the value of \(F\) is 1, one finds that the detector is at mostly optimal
location for the putative GW source at that location. On the other hand, 0 value means that it is impossible to see any GW signals with that detector. When we obtain the antenna response $\mathcal{F}$ of each detector, we can calculate the volume-weighted average distance range $D_{\text{IFO}}$ by simply multiplying $\mathcal{F}$ to the horizon distance $D_h$ (Abbott et al 2009a, Abadie et al 2010c).

$$D_{\text{IFO}} \sim \mathcal{F}_{\text{IFO}} \times D_h$$

where IFO of the superscripts stands for interferometric observatories, H1, L1, or V1. The horizon distances $D_h$ and volume-weighted average distances $\bar{D}_{\text{IFO}}$ of relevant detectors for GRB070714B and GRB070923 are listed in table A1. From the values of $\bar{D}_{\text{IFO}}$, we can read the upper limit of reachable distance of each detector. That is, if a GW event is occurred within the distance limit at the given sky location, we may detect that event. Among the listed $\bar{D}_{\text{IFO}}$ in table A1, in particular, we choose the second largest value as the decisive distance for the GRB070714B as done in (Abadie et al 2012a). Resultantly, applied distance ranges for the software injection are 2–30 Mpc for BNS system and 2–72.5 Mpc for NS-BH case. We also suppose that the sources of GWs are distributed in uniform within the ranges.

### Appendix B. Learning procedure in iRPROP algorithm

The *resilient backpropagation* (RPROP) algorithm proposed by Riedmiller and Braun (1993) is one of the best performing first-order learning methods for neural networks. In order to speed up the learning speed of the RPROP algorithm, *improved RPROP* (iRPROP)\textsuperscript{21} algorithm is introduced in (Igel and Hüsken 2000). In this section, we present how the iRPROP algorithm updates the connection weights.

We use following notations: $w_{ij}$ denote the weight in a neural network from a node $j$ to a node $i$ and $E$ be an arbitrary error measure (in our case, MSE) that is differentiable with respect to the weight. Superscripts indicate the learning epoch (iteration).

In the iRPROP algorithm, the direction of each weight update is based on the sign of the partial derivative $\partial E / \partial w_{ij}$. A step-size, i.e., the update amount of a weight, is adapted for each weight individually.

One iteration of the iRPROP algorithm can be divided into two parts. The first part is the adjustment of the step-sizes. For each weight, $w_{ij}$, an individual step-size, $\Delta_{ij}$, is adjusted

\textsuperscript{21} The iRPROP algorithm of this chapter corresponds to iRPROP$^+$ in (Igel and Hüsken 2000) and the FANN library package imposes it only. So, we here use iRPROP instead of using iRPROP$^+$. 

---

**Table A1.** The horizon distance and volume-weighted average distance range of each detector. All tabulated values are given in unit of Mpc. For the horizon distances of H1 and L1, we take the maximum value of given range in (Abbott et al 2009a). Meanwhile, we read the value of mode from figure 1 of (Abadie et al 2010c) for the horizon distance of V1. Among the values of volume-weighted average distance ranges, asterisks indicate the decisive distances.

<table>
<thead>
<tr>
<th>Data</th>
<th>Horizon distance, $D_h$ (Mpc)</th>
<th>Volume-weighted average distance, $\bar{D}_{\text{IFO}}$ (Mpc)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H1</td>
<td>L1</td>
</tr>
<tr>
<td>GRB070714B</td>
<td>35.0</td>
<td>35.0</td>
</tr>
<tr>
<td>GRB070923</td>
<td>35.0</td>
<td>35.0</td>
</tr>
</tbody>
</table>
using the following rule:

\[
\Delta_{ij}^{(t)} = \begin{cases} 
\eta^- \Delta_{ij}^{(t-1)} & \text{if } \frac{\partial E}{\partial w_{ij}} \cdot \frac{\partial E^{(t-1)}}{\partial w_{ij}} > 0, \\
\eta^+ \Delta_{ij}^{(t-1)} & \text{if } \frac{\partial E}{\partial w_{ij}} \cdot \frac{\partial E^{(t-1)}}{\partial w_{ij}} < 0, \\
\Delta_{ij}^{(t-1)} & \text{otherwise},
\end{cases}
\]

(B.1)

where \(\eta^-\) and \(\eta^+\) are input parameters should be given to satisfy \(0 < \eta^- < 1 < \eta^+\). If the partial derivative, \(\frac{\partial E}{\partial w_{ij}}\), possesses the same sign for consecutive steps, the step-size is increased, whereas if it changes sign, the step-size is decreased.

The step-sizes are bounded by another input parameters, \(\Delta_{\text{min}}\) and \(\Delta_{\text{max}}\). So, we can rewrite equation (B.1) as following:

\[
\Delta_{ij}^{(t)} = \begin{cases} 
\min \left( \eta^+ \Delta_{ij}^{(t)}, \Delta_{\text{max}} \right) & \text{if } \frac{\partial E}{\partial w_{ij}} \cdot \frac{\partial E^{(t-1)}}{\partial w_{ij}} > 0, \\
\max \left( \eta^- \Delta_{ij}^{(t)}, \Delta_{\text{min}} \right) & \text{if } \frac{\partial E}{\partial w_{ij}} \cdot \frac{\partial E^{(t-1)}}{\partial w_{ij}} < 0.
\end{cases}
\]

(B.2)

Note that setting the stored derivative to zero avoids an update of the learning rate in the next iteration because the otherwise branch in equation (B.1) becomes active.

After adjusting the step-sizes at an iteration, \(t\), the second part of the algorithm, i.e., the update of the weights of next step, \(w_{ij}^{(t+1)}\), is determined by

\[
w_{ij}^{(t+1)} = w_{ij}^{(t)} \cdot \text{sign} \left( \frac{\partial E}{\partial w_{ij}} \cdot \Delta_{ij}^{(t)} \right),
\]

(B.3)

where the sign operator returns +1 if its argument is positive, −1 if the argument is negative, and 0 otherwise.

**Appendix C. Round-robin process**

In the preparation of input samples, we conduct a pre-process, called *round-robin process*, on the sample data before performing training process. The purpose of this pre-process is to mitigate the possibility of overestimation (or, equivalently, overtraining) which may occur by inadequate training with small or limited numbers of sample data. In order to successfully implement this pre-process and reduce the rate of overtraining, we prepare \(M\) sets of roundrobined sample data by evenly dividing the whole samples of \(X_S\) and \(X_B\) into \(M\) different sets of \(X_S^k\) and \(X_B^k\), where \(k = 1, 2, \ldots, M\), in the same manner: a set to be consistently composed of one-tenth of total signal samples and one-tenth of total background samples. When it is done, there are no overlaps between the samples in one set and the samples in other sets. Then, to make a pair of training and evaluation sets, we let one of total round-robined sets to be an evaluation set and rest of them be the pairing training set. Then, with the definition in equations (5) and (6), we can get the first pair of an evaluation set \(E^{1st}\) and a training set \(T^{1st}\) as follows:

\[
E^{1st} = X_S^1 \cup X_B^1 = \left\{ x_S^i; i = 1, 2, \ldots, N_S^1 \right\} \cup \left\{ x_B^j; j = 1, 2, \ldots, N_B^1 \right\},
\]

(C.1)
where $N'_S = N_S/M$ and $N'_B = N_B/M$. Then, with the same manner, other $M - 1$ pairs can be configured as

$$
E^{2nd} = X^2_S \cup X^2_B = \left\{ x^2_i; i = N'_S + 1, \ldots, 2N'_S \right\} \\
\cup \left\{ x^2_m; m = N'_B + 1, \ldots, 2N'_B \right\},
$$

Then, with the same manner, other $M - 1$ pairs can be configured as

$$
T^{2nd} = X^2_S \cup X^2_B \cup \cdots \cup X^M_S \cup X^M_B = \left\{ x^2_i; i = 1, \ldots, N'_S \right\} \\
\cup \left\{ x^2_m; m = 1, \ldots, N'_B \right\}
$$

Note that, from the above relations, one can easily see that overlaps in configured samples between different training sets are allowed while there are no overlaps between each of the evaluation sets. By repeating similar pairing for other sets we now have $M$ pairs of $[T^{k}, E^{k}]$ which have different evaluation samples than each other. In this work, we prepare 10 pairs of round-robined samples with $M = 10$.

Firstly we train ANN with the samples in a training set $T^{k}$ and then evaluate $x^k_i$s and $x^k_m$s in a paired evaluation set $E^{k}$. In the training process, ANN recursively finds an optimal connection weight between each node by the iRPROP algorithm. When the training process is finished, the resulting connection weights are saved for the evaluation process. Then, evaluation samples $x^k_i$ and $x^k_m$ get their ranks based on the trained result. Finally, from the evaluation, ANN scores a rank between 0 and 1 for each evaluation sample.

References

Abadie J et al 2010a Class. Quantum Grav. 27 173001
Abbott B et al 2007 Class. Quantum Grav. 24 5343–70
Acernese F et al 2008 Class. Quantum Grav. 25 184001
Barbier L et al 2007a GCN Circ. 6623 1
Barbier L et al 2007b GCN Circ. 6821 1
Boyle M et al 2007 Phys. Rev. D 76 124038
Breiman L 1996 Mach. Learn. 24 123–40
Breiman L 2001 Mach. Learn. 45 5–32 and references therein
Cannon K C 2008 Class. Quantum Grav. 25 105024
Cokelaer T 2007 Phys. Rev. D 76 102004
Dietz A 2011 Astron. Astrophys. 529 A97
Graham J F et al 2007 GCN Circ. 6836 1
Hessels J W T et al 2006 Science 311 1901
Mandel I and O’Shaughnessy R 2010 Class. Quantum Grav. 27 114007
Meegan C A et al 1992 Nature 355 143
Ott C D 2009 Class. Quantum Grav. 26 063001
Pascanu R, Dauphin Y N, Ganguli S and Bengio Y 2014
Predoi V 2012 Gravitational waves and short gamma ray bursts PhD Thesis Cardiff University
Racusin J L et al 2007 GCN Circ. 6620 1
Rampone S, Pierro V, Troiano L and Pinto I M 2013 Int. J. Mod. Phys. C 24 1350084
Stroh M C et al 2007 GCN Circ. 6818 1