Randomized reward redistribution for HPGe waveform classification under weakly-supervised learning setup

Sonata Simonaitis-Boyd
Halıcıoğlu Data Science Institute
University of California, San Diego
La Jolla, CA
sonata@ucsd.edu

Aobo Li
Halıcıoğlu Data Science Institute, Department of Physics
University of California, San Diego
La Jolla, CA
liaobo77@ucsd.edu

Abstract

High-Purity Germanium (HPGe) detectors represent a leading technology in the search for neutrinoless double beta ($0\nu\beta\beta$) decay [1] [2], a Beyond the Standard Model (BSM) process that, if discovered, would fundamentally revise our understanding of the universe. A key analysis task in $0\nu\beta\beta$ decay experiments is the classification task of separating signals from background noise, which is traditionally approached as a supervised learning problem. However, HPGe detector data is often unlabeled, and producing ground-truth labels for every data point is an involved process. In this work, we reformulate the classification task of HPGe detector data as a weakly-supervised learning task and apply an episodic reinforcement learning (RL) algorithm with Randomized Return Decomposition (RRD) to address it. We evaluate our algorithm on real data produced by the MAJORANA DEMONSTRATOR HPGe detector experiment. With significantly fewer labels, the RL-trained weakly-supervised classifier slightly outperforms a fully-supervised classifier under the same energy cut. This work shows potential for training classifiers to reject background in future HPGe experiments like LEGEND.

1 Introduction

The Standard Model (SM) of particle physics represents physicists' current best understanding of elementary particles and fundamental symmetries. It is very successful, albeit some particles are still not very well understood, like neutrinos. Alongside being electrically neutral, interacting poorly with matter, and having incredibly small masses, neutrinos also have the potential to be Majorana particles—that is, their own antiparticles. Currently, the only feasible method for determining the Majorana status of neutrinos is the search for neutrinoless double beta ($0\nu\beta\beta$) decay.

Accurate signal/background classification is a key factor in the search for $0\nu\beta\beta$ decay in HPGe detectors. Traditionally, this is considered a supervised learning task, requiring all training data to be fully labeled. While raw detector data for $0\nu\beta\beta$ decay experiments like the MAJORANA DEMONSTRATOR [1] are plentiful, labels are very time-consuming to acquire, prompting scientists to look for alternatives to traditional supervised learning methods for data analysis. However, knowledge of $0\nu\beta\beta$ decay physics can provide a partial label, allowing us to establish a weakly-supervised learning problem (see Section 2.2). We turn to episodic reinforcement learning (RL) to solve our classification problem.

In episodic RL, the reward is generated at the end of a sequence of actions (referred to as an *episode*) rather than after each individual action. This raises the important question of how to distribute

Machine Learning and the Physical Sciences Workshop, NeurIPS 2024.

the episodic reward among the individual (state, action) pairs. For instance, how can labels like 'win' or 'loss' help determine which actions contributed most to the final outcome of each episode? Randomized Return Decomposition (RRD) [3], which builds on the foundations of two existing reward redistribution implementations [4] [5] by introducing Monte Carlo estimation into the loss function, is a new training method that scales up and outperforms baseline algorithms. This work utilizes RRD as a reward redistribution algorithm to address the weakly-supervised learning problem arising from the HPGe detector dataset.

2 Methods

2.1 MAJORANA DEMONSTRATOR data

This study is conducted upon the open data release from the MAJORANA DEMONSTRATOR (MJD) [6], a HPGe detector experiment dedicated to the search for $0\nu\beta\beta$ decay. All data points are generated from a real HPGe detector array rather than simulation. The dataset is a subset of ^{228}Th calibration data from the DEMONSTRATOR and each data point consists of a raw detector waveform, a set of accompanying analysis labels (including particle energy [7] and other discrimination cuts), the rise time (marking the start of the waveform's rising edge), and metadata. In MJD, the $0\beta\beta$ decays are considered single-site events (SSEs) which produce a single-step-like waveform because the electron decay products deposit their energy within $\sim$ 1 mm in the detector; signals with a large energy deposition region ($\sim$ 1 cm), however, will produce a multi-step-like waveform, hence multi-site event (MSE). MSEs are characteristic of photon backgrounds, which we would like to exclude from our dataset while retaining SSEs. Examples of SSE and MSE waveforms can be found in Figure 2b.

In the MJD data release, each data point comes with a label indicating whether it is a SSE or MSE. This serves as the ground-truth label for the data, with a label of 0 representing rejection by the A vs E cut (MSE) and a label of 1 representing acceptance by the A vs E cut (SSE). These labels are generated through a traditional analysis method known as A vs E [8]; designing and fine-tuning A vs E typically takes 1–2 years. In this work, we aim to train a machine learning model without relying on the ground-truth labels generated by A vs E, but ultimately use them to validate the trained model.

2.2 Weakly-supervised learning

The MAJORANA DEMONSTRATOR produces time series data, or waveforms, as its raw output, with the amplitude of each waveform being approximately proportional to the energy of the incident particle that created it. The overall behavior of the processes detected by the DEMONSTRATOR are best visualized by plotting an energy histogram of many waveforms; a typical histogram from the MJD data release is shown in the lower right of Figure 1. The histogram provides information as a collective label for the waveforms: peaks at specific energy values represent known nuclear physics interactions, indicating that the waveforms corresponding to those peaks are identified as either SSEs or MSEs. However, energy ranges that have no peaks provide no information about event type. The histogram is thus an inexact and incomplete label for a weakly-supervised learning setup, as SSE and MSE labels will only emerge after the histogram is created.

Given this setup, we designed three energy cuts on the training data at three energy ranges that typically corresponded to either SSE or MSE peaks. The ranges—1592 ± 5 keV (SSE), 2103 ± 5 keV (MSE), 866 ± 5 keV (MSE)—were chosen based on knowledge of underlying nuclear physics processes. This energy cut is applied to the first five MJD data release training sets to form our training dataset. After all cuts, 2857 waveforms remain. Post-processing, the data is separated into training and testing sets using a 70/30 split. We then used RRD for an episodic RL training algorithm to distribute the collective label among the raw detector data.

2.3 Reinforcement learning

The RRD-RL algorithm's (see Figure 1) policy is a five-layer fully-connected neural network (FCNN) consisting of a linear transformation, batch normalization, and Leaky ReLU function in each of the first four layers. We used hyperbolic tangent (tanh) as the activation function of the output layer. Chosen for its outputs within the range $[-1, 1]$, tanh has a natural midpoint at 0 that allows for a clear separation between negative and positive outputs. This is conducive to a binary classifier, as our construction of the reward function (see Algorithm 1, bottom left of Figure 1) encourages the model

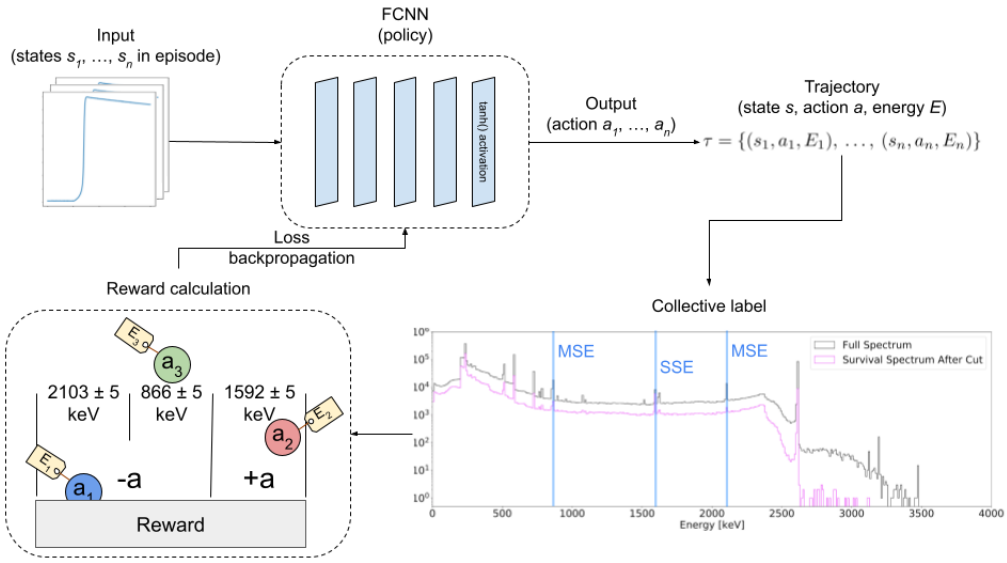


Figure 1: Schematic of the RRD-RL classifier.

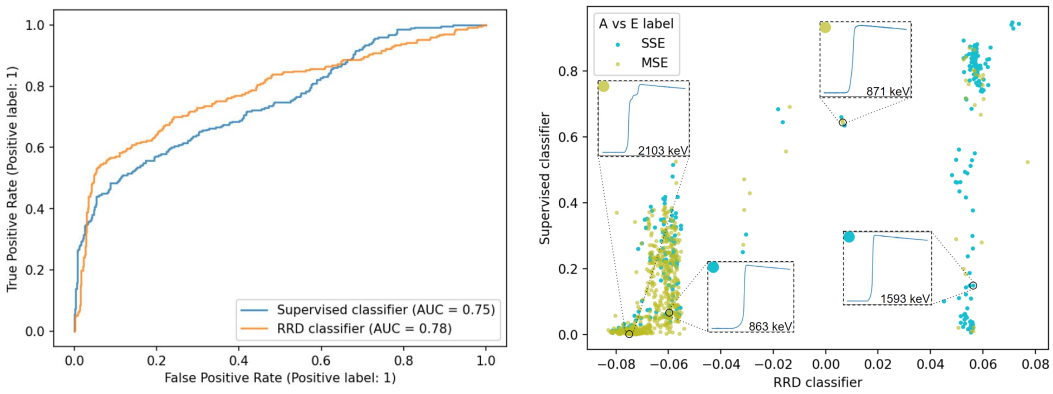
to assign negative outputs to MSE-type waveforms and positive outputs to SSE-type waveforms, producing an overall large, positive reward.

Algorithm 1 RRD Classifier Reward Function

Require: action a, energy E
Initialize:
Reward $r = 0$
Peak width w
for a in trajectory **do**
if $1592 - w < E < 1592 + w$ **then**
$r+ = a$
else if $2103 - w < E < 2103 + w$ **then**
$r- = a$
else if $866 - w < E < 866 + w$ **then**
$r- = a$
end if
end for
return r

The training algorithm and RRD loss (Eq. 1) are implemented along the lines of Ref. [3], the latter by randomly sampling subsequences $\mathcal{I}_j$ of varying lengths M times from a given trajectory of length T_j and calculating subsequence rewards $\widehat{R}_\theta$, which are then compared to the trajectory reward R_{ep}. Subsequence and trajectory rewards were both calculated using Algorithm 1. In this context, a trajectory is defined as the collection of all (state, action, energy) tuples throughout the episode (see top right of Figure 1), with one reward assigned for the entire trajectory.

$$\widehat{\mathcal{L}}_{\text{Rand-RD}}(\theta) = \frac{1}{M}\sum_{j=1}^{M}\left[\left(R_{\text{ep}}(\tau_j) - \frac{T_j}{|\mathcal{I}_j|}\sum_{t\in\mathcal{I}_j}\widehat{R}_\theta(s_{j,t}, a_{j,t})\right)^2\right] \tag{1}$$



(a) (b)

Figure 2: (a) ROC curves for the supervised classifier (blue) and the RRD classifier (orange). (b) Scatter plot of supervised classifier outputs vs. RRD classifier outputs, with SSEs in cyan and MSEs in olive. The waveforms of two SSEs and two MSEs are displayed, including event energy.

3 Results & Discussion

To demonstrate the efficiency of our approach, we also constructed a traditional supervised learning classifier with the same network structure and hyperparameters. The supervised learning classifier uses a Sigmoid activation function as the output for its compatibility with Binary Cross-Entropy loss. As discussed in Section 2.2, the supervised learning classifier requires a complete label for every data point, while the RL classifier can make use of the inexact label generated by the histogram. Both models are trained with the Adam optimizer [9].

During model evaluation, the RL-trained classifier was found to perform marginally better than the supervised classifier. The ROC curves for the models are shown in Figure 2a. The supervised classifier had an Area Under the Curve (AUC) of 0.75 for the ROC curve, slightly lower than the RRD classifier's AUC of 0.78. This indicates that the supervised classifier was marginally less accurate in its classifications compared to the RRD model. We acknowledge that this may be due to the limited amount of training data.

Figure 2b provides a qualitative comparison of the two models' classification abilities. The y-axis represents the output of the supervised learning classifier, while the x-axis shows the output of the RRD classifier for the same event. In both cases, a higher output suggests the event is classified as more SSE-like, while a lower score indicates it is classified as more MSE-like. Four of the plotted points are isolated to show their corresponding waveforms. Although our reward function relies on the assumption that certain energy ranges contain only SSEs or only MSEs, this is not true in practice—for example, the event in the lower-left is single-site, yet has an energy that would have our RL model sort it as multi-site. Even so, these pulled-out waveforms imply that the classification decision was based primarily on waveform energy—and if this is true, then the results of the ROC curve show energy to be a more reliable indicator of event type than ground-truth labels.

That said, energy is not the only metric. While the scale of outputs for the RL classifier is incredibly small, the model outputs on the whole are tightly grouped: the majority of waveforms assigned MSE-like outputs are grouped closely together on the left, while SSE-like waveforms have outputs primarily within a small range on the right. This makes the handful of waveforms assigned middling outputs stand out. The waveform in the top right is ground-truth a background event and has an MSE-like energy following the reward function, yet it is not grouped with the other background-type waveforms. This is not surprising, because while event energy is a robust indicator of event type, it is not a perfect one.

4 Conclusions

Overall, the RRD classifier is an alternative to a traditional supervised learning model in the realm of signal/background event discrimination via classification task. Its main advantage over the supervised

model is that it does not require exact or complete labels for training; instead, the RRD model relies solely on waveform energy, which can be easily derived from waveform amplitude. Following our validation study, the RRD model, trained on incomplete and inexact labels, demonstrated a classification performance comparable to that of the supervised model trained on complete and precise labels.

Our future work involves reworking the reward function to allow for a more comprehensive analysis of the data and to reduce any existing energy dependencies, as well as hyperparameter tuning, increasing the amount of training data, and investigating the RRD model output sizes.

Acknowledgments and Disclosure of Funding

This work was supported by resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The data used in this work was obtained from a public release by the MAJORANA DEMONSTRATOR experiment [6], and we thank the collaboration for making their data publicly available to the community. We also acknowledge support from the NSF HDR Institute Accelerated AI Algorithms for Data-Driven Discovery (A3D3) under Award PHY-2117997.

References

[1] I. J. Arnquist, F. T. Avignone, A. S. Barabash, C. J. Barton, P. J. Barton, K. H. Bhimani, E. Blalock, B. Bos, M. Busch, M. Buuck, T. S. Caldwell, Y-D. Chan, C. D. Christofferson, P.-H. Chu, M. L. Clark, C. Cuesta, J. A. Detwiler, Yu. Efremenko, H. Ejiri, S. R. Elliott, G. K. Giovanetti, M. P. Green, J. Gruszko, I. S. Guinn, V. E. Guiseppe, C. R. Haufe, R. Henning, D. Hervas Aguilar, E. W. Hoppe, A. Hostiuc, M. F. Kidd, I. Kim, R. T. Kouzes, T. E. Lannen V., A. Li, A. M. Lopez, J. M. López-Castaño, E. L. Martin, R. D. Martin, R. Massarczyk, S. J. Meijer, S. Mertens, T. K. Oli, G. Othman, L. S. Paudel, W. Pettus, A. W. P. Poon, D. C. Radford, A. L. Reine, K. Rielage, N. W. Ruof, D. C. Schaper, D. Tedeschi, R. L. Varner, S. Vasilyev, J. F. Wilkerson, C. Wiseman, W. Xu, C.-H. Yu, and B. X. Zhu. Final result of the majorana demonstrator's search for neutrinoless double-β decay in ^{76}Ge. *Phys. Rev. Lett.*, 130:062501, Feb 2023.

[2] M. Agostini, G. R. Araujo, A. M. Bakalyarov, M. Balata, I. Barabanov, L. Baudis, C. Bauer, E. Bellotti, S. Belogurov, A. Bettini, L. Bezrukov, V. Biancacci, D. Borowicz, E. Bossio, V. Bothe, V. Brudanin, R. Brugnera, A. Caldwell, C. Cattadori, A. Chernogorov, T. Comellato, V. D'Andrea, E. V. Demidova, N. Di Marco, E. Doroshkevich, F. Fischer, M. Fomina, A. Gangapshev, A. Garfagnini, C. Gooch, P. Grabmayr, V. Gurentsov, K. Gusev, J. Hakenmüller, S. Hemmer, R. Hiller, W. Hofmann, J. Huang, M. Hult, L. V. Inzhechik, J. Janicskó Csáthy, J. Jochum, M. Junker, V. Kazalov, Y. Kermaïdic, H. Khushbakht, T. Kihm, I. V. Kirpichnikov, A. Klimenko, R. Kneißl, K. T. Knöpfle, O. Kochetov, V. N. Kornoukhov, P. Krause, V. V. Kuzminov, M. Laubenstein, A. Lazzaro, M. Lindner, I. Lippi, A. Lubashevskiy, B. Lubsandorzhiev, G. Lutter, C. Macolino, B. Majorovits, W. Maneschg, L. Manzanillas, M. Miloradovic, R. Mingazheva, M. Misiaszek, P. Moseev, Y. Müller, I. Nemchenok, K. Panas, L. Pandola, K. Pelczar, L. Pertoldi, P. Piseri, A. Pullia, C. Ransom, L. Rauscher, S. Riboldi, N. Rumyantseva, C. Sada, F. Salamida, S. Schönert, J. Schreiner, M. Schütt, A.-K. Schütz, O. Schulz, M. Schwarz, B. Schwingenheuer, O. Selivanenko, E. Shevchik, M. Shirchenko, L. Shtembari, H. Simgen, A. Smolnikov, D. Stukov, A. A. Vasenko, A. Veresnikova, C. Vignoli, K. von Sturm, T. Wester, C. Wiesinger, M. Wojcik, E. Yanovich, B. Zatschler, I. Zhitnikov, S. V. Zhukov, D. Zinatulina, A. Zschocke, A. J. Zsigmond, K. Zuber, and G. Zuzel. Final results of gerda on the search for neutrinoless double-β decay. *Phys. Rev. Lett.*, 125:252502, Dec 2020.

[3] Zhizhou Ren, Ruihan Guo, Yuan Zhou, and Jian Peng. Learning Long-Term Reward Redistribution via Randomized Return Decomposition. *arXiv e-prints*, page arXiv:2111.13485, November 2021.

[4] Yonathan Efroni, Nadav Merlis, and Shie Mannor. Reinforcement learning with trajectory feedback. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(8):7288–7295, May 2021.

[5] Tanmay Gangwani, Yuan Zhou, and Jian Peng. Learning guidance rewards with trajectory-space smoothing, 2020.

[6] I. J. Arnquist et al. Majorana Demonstrator Data Release for AI/ML Applications, 2023.

[7] I. J. Arnquist et al. Charge trapping correction and energy performance of the majorana demonstrator. *Physical Review C*, 107(4), April 2023.

[8] S. I. Alvis et al. Multisite event discrimination for the majorana demonstrator. *Physical Review C*, 99(6), June 2019.

[9] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.