Objective Clustering of GRBs Associated with Flux Duration

Je Qin Chooi[1], Ashley Villar[2]

1: Harvard College; 2: Harvard-Smithsonian Center for Astrophysics

Harvard-Smithsonian Center for Astrophysics

Abstract

Gamma-ray bursts (GRB) have at least two astrophysical origins: core-collapse and compact mergers.[1] Compact mergers are of interest as they are the only confirmed sources of r-process nucleosynthesis responsible for the generation of heavy elements.[2] Previous work partitions GRBs at T90 ≈ 2s, the duration containing 90% of the statistically significant flux.[3] Long GRBs are associated with core-collapse[4] while short GRBs are associated with compact mergers.[5] However, some long GRBs have no observed supernovae[6] and there is a long GRB with a compact merger progenitor.[7] Recent works classified the GRBs by their light curves using t-SNE.[8] However, t-SNE is designed for data visualization and does not rigorously reflect physical similarities. This project uses a GRB dataset[9] and employs autoencoders for dimensionality reduction and K-Means for clustering. We find that our results largely agree with the t-SNE approach, but we have a predicted a larger size for the smaller cluster that is associated with shorter T_{90}.

Introduction

R-process nucleosynthesis are only known to happen in compact merger events which are associated with short-duration GRBs. However, the discovery of compact mergers with long-duration GRBs make it relevant to identify possible characteristics of these bright events that hints at compact mergers, allowing cross-validation with gravitational wave detectors too.

Methods

A dataset of 1,502 GRBs from 2004 to 2023 from the Swift Burst Alert Telescope (BAT) are downloaded through the package ClassiPy GRBs (Garcia-Cifuentes et al. 2023).

Preprocessing procedures:

1. Apply FABADA for noise reduction
2. Limit all GRBs out of T_{100}
3. Normalize light curves (lc) by total fluence in the 15-350keV band.
4. Pad with zeros all GRBs, putting them on the same time standard basis.
5. Concatenate data from all bands and perform DFT to get the Fourier amplitude spectrum.

An autoencoder with hidden size 100 and laten size 5 is trained for 500 epochs and a learning rate of 10^{-3} for dimensionality reduction. K-means is applied with varying numbers of clusters.

Note: a 1D CNN autoencoder was trained, but it failed to generalize properly and has significantly worse performance than the plain autoencoder.

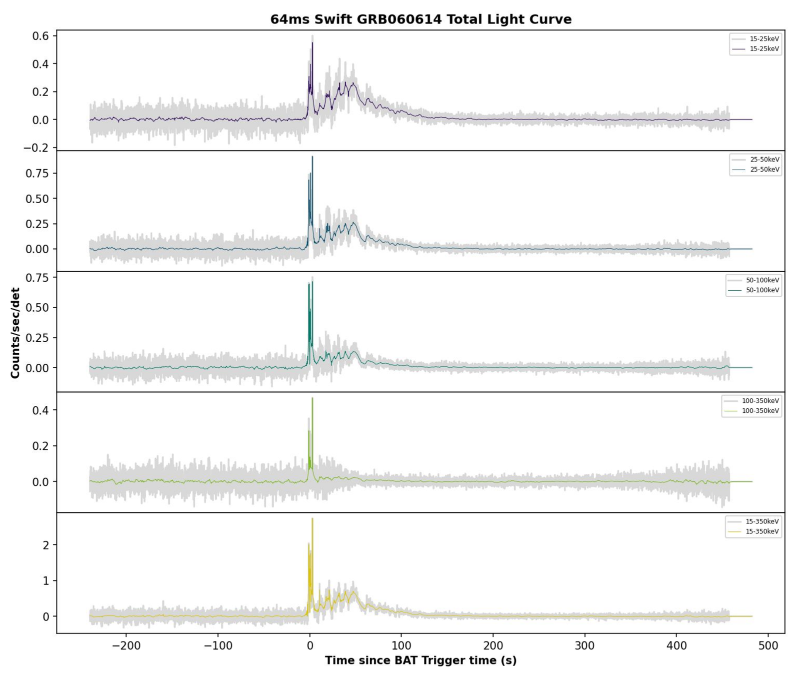

Figure 1. Noise reduction using FABADA

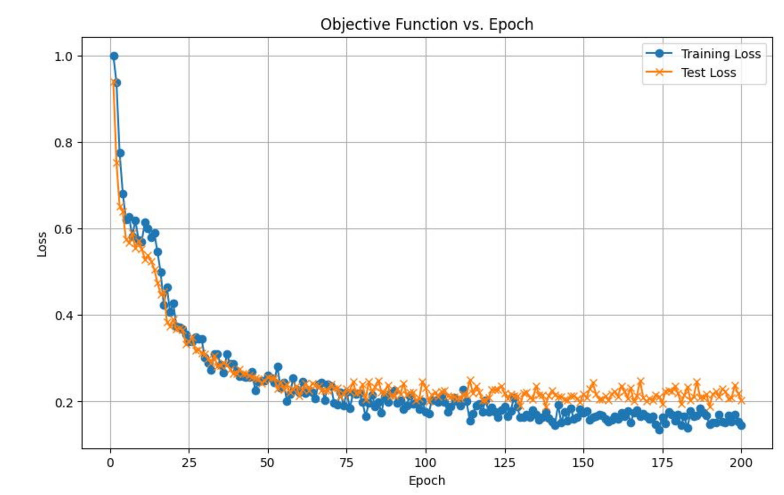

Figure 2. Training and test loss of the autoencoder on light curves

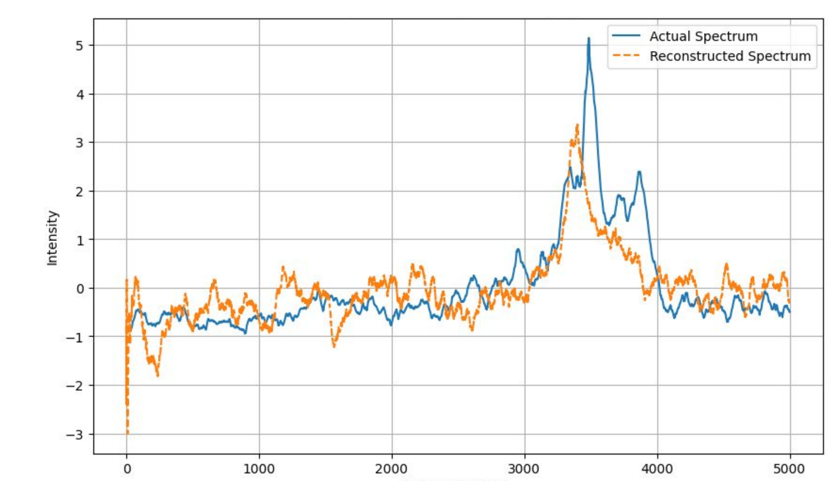

Figure 3. Example of reconstructed light curve

Results

GRBs cluster by duration of significant flux similar to previous work using t-SNE.

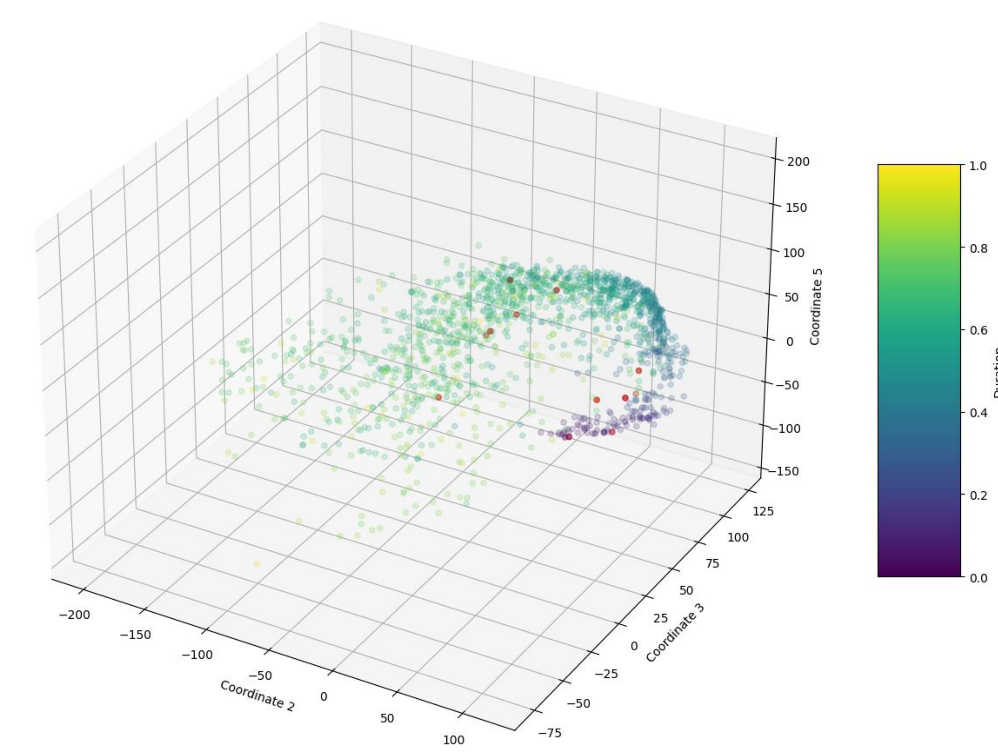

Figure 4. Spatial coordinates represent dimensionally-reduced light curve. Color represents duration of significant flux. Red are compact mergers.

Compact mergers do not seem to belong to any particular group consistently, regardless of number of clusters.

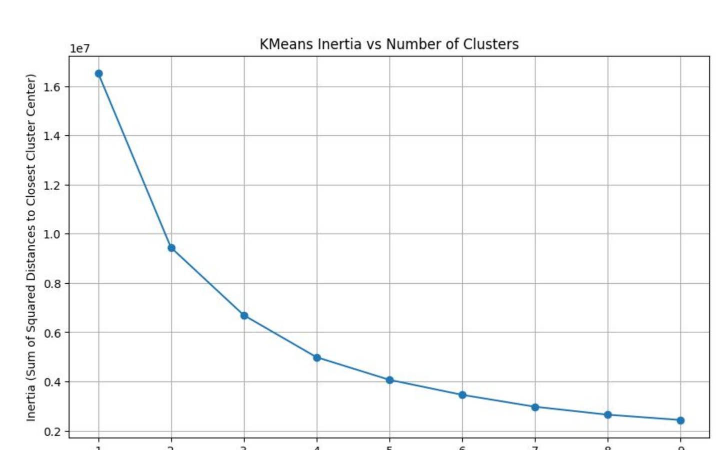

Figure 5. K-means loss by number of cluster. No "knee" observed.

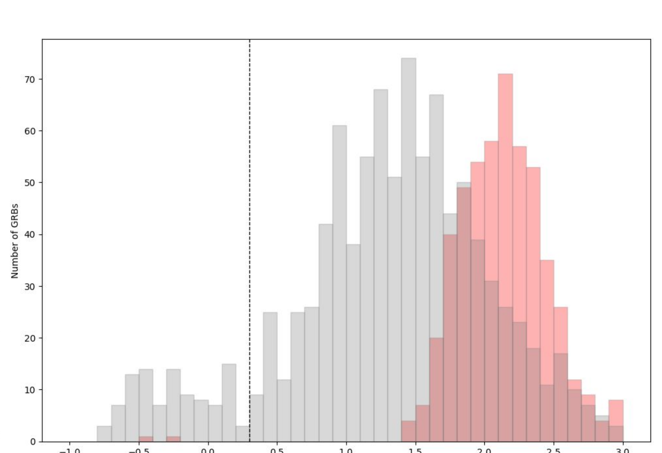

Figure 6. Distribution of duration of significant flux by clusters found with K-means with number of clusters set to 2

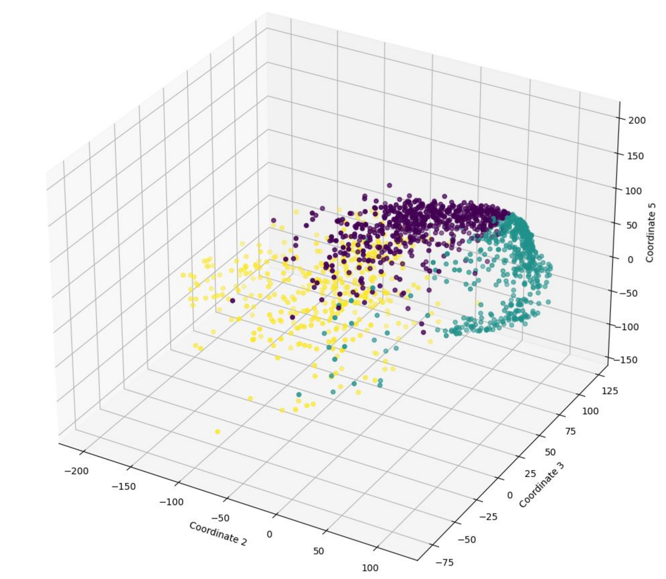

Figure 7. Dimensionally reduced GRB light curves color-coded by K-means

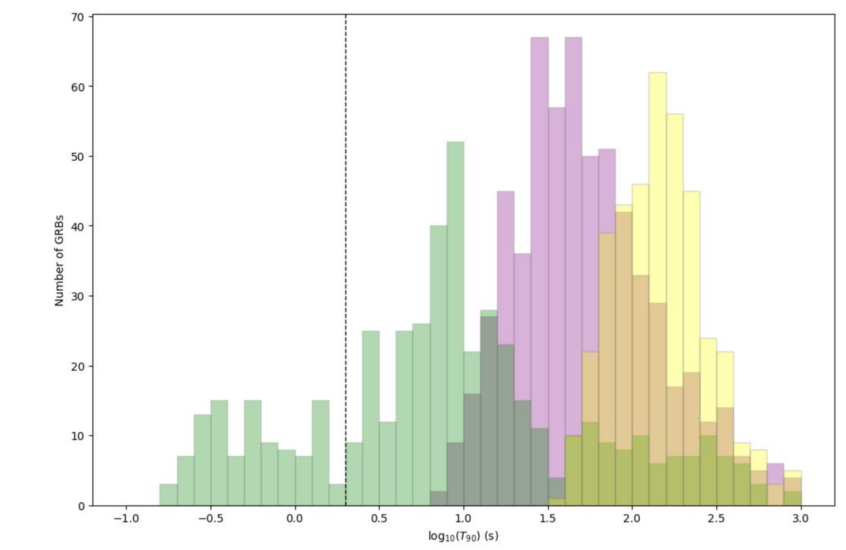

Figure 8. Distribution of duration of significant flux by clusters found with K-means with number of clusters set to 4

Discussion

Dimensionality reduction using autoencoders is done for the first time on GRB light curves, resulting a smooth visual distribution of GRBs by duration of significant flux that is similar to previous work using t-SNE. K-means cluster the GRBs to two distributions characterized by different average duration of significant flux, but the proportion of each distribution differs greatly from previous work.

Once the autoencoder is trained, it can be applied cheaply to new GRBs, unlike t-SNEs which computation has to be done offline on the entire set of GRBs. This will allow fast classification of new GRBs to initiate follow-up observations.

The current model only uses one of the five bands. Future, larger models can be trained to further capture the entirety of the dataset and reduce test loss. With lower reconstruction loss, it might be possible that a well-defined group for Extended Emission GRBs to emerge.

References & Acknowledgements

1. Kouveliotou et al., 'Identification of Two Classes of Gamma-Ray Bursts'.
2. Tanvir et al., 'The Emergence of a Lanthanide-Rich Kilonova Following the Merger of Two Neutron Stars'.
3. Paciesas et al., 'The Fourth BATSE Gamma-Ray Burst Catalog (Revised)'.
4. Stanek et al., 'Spectroscopic Discovery of the Supernova 2003dh Associated with GRB 030329*'.
5. Berger, Fong, and Chornock, 'AN R-PROCESS KILONOVA ASSOCIATED WITH THE SHORT-HARD GRB 130603B'.
6. Valle et al., 'An Enigmatic Long-Lasting γ-Ray Burst Not Accompanied by a Bright Supernova'.
7. Rastinejad et al., 'A Kilonova Following a Long-Duration Gamma-Ray Burst at 350 Mpc'.
8. Jespersen et al., 'An Unambiguous Separation of Gamma-Ray Bursts into Two Classes from Prompt Emission Alone'.
9. Garcia-Cifuentes et al., 'Identification of Extended Emission Gamma-Ray Burst Candidates Using Machine Learning'.