

ClassipyGRB: Machine Learning-Based Classification and Visualization of Gamma Ray Bursts using t-SNE

Keneth Garcia-Cifuentes^{1*}, Rosa L. Becerra^{1*}, and Fabio De Colle¹

¹ Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, Apartado Postal 70-543, 04510 CDMX, México ¶ Corresponding author * These authors contributed equally.

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Dan Foreman-Mackey](#) ↗

Reviewers:

- [@wkerzendorf](#)
- [@dfm](#)

Submitted: 22 May 2023

Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

Gamma-ray burst (GRBs) are the brightest events in the universe. Since decades, astrophysicists have known about their cosmological nature. Every year, spacecraft missions such as Fermi and SWIFT detect hundred of them. In spite of this large sample, these phenomena show a complex taxonomy in the first seconds after their appearance, which makes it very difficult to find resemblance between them using conventional techniques.

It is known that GRBs are originated by the death of a massive star or from the merger of two compact objects. The typical classification is based on their duration ([Kouveliotou et al., 1993](#)). Nevertheless, events such as GRB 211211A ([Yang et al., 2022](#)), whose duration of about 50 seconds lies in the group of long GRBs, has challenged this categorization by the evidence of features related with the short GRB population (the kilonova emission and the properties of its host galaxy). Therefore, a classification based only on their gamma-ray duration is not completely reliable to determine the progenitors of such events.

Motivated by this problem, ([Jespersen et al., 2020](#)) and ([Steinhardt et al., 2023](#)) carried out analysis of GRB lightcurves by using the t-SNE algorithm, showing that Swift/BAT GRBs database, consisting of light curves in four energy bands (15-25 keV, 25-50 keV, 50-100 keV, 100-350 keV), clusters into two groups corresponding with the typical long/short classification. However, in this case, this classification is based on the information provided by their gamma-ray emission light curves.

ClassiPyGRB is a Python 3 package to download, process, visualize and classify GRBs database from the [Swift/BAT Instrument](#) (up to July 2022). It is distributed over the GNU General Public License Version 2 (1991). We also included a noise-reduction and an interpolation tools for achieving a deeper analysis of the data.

Statement of need

ClassipyGRB was designed to be used by astronomers whose scientific research is focused on GRBs. This module provides interactive visualizations of the light curves of GRBs, and the similarities they share. ClassipyGRB allows the comparison in a few seconds with other events in order to find resembling GRBs in high-frequencies through the proximity between them.

This method has been successfully applied for the correct identification of candidates for GRBs with extended emission that have not been previously identified by other groups, saving time and human effort ([Garcia-Cifuentes et al., 2023](#)). We locate extended emission GRBs reported by various authors under different criteria in our t-SNE maps and discuss its possible identification using only the gamma-ray data provided by the automatic pipeline of Swift/BAT.

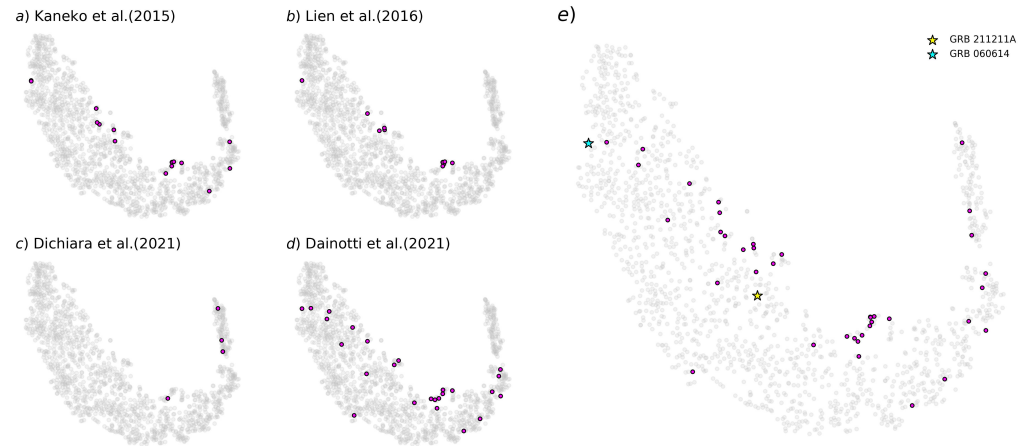


Figure 1: t-SNE visualization map obtained for the noise-reduced dataset binned at 64 ms with $pp = 30$. GRBs colored in magenta are classified as Extended Emission by previous works. Image taken from (Garcia-Cifuentes et al., 2023)

Moreover, ClassipyGRB has been use to find in few seconds, similar GRBs with some specific feature, such as a bright component (Angulo Valdez et al.(2023), in prep).

Methodology and Structure of ClassiPyGRB

ClassipyGRB mainly consists of three parts:

- 1) Retrieval and visualization of data from Swift/BAT: We implement an easy and fast code to download and plot an existing GRB post-processed data. There is the possibility to modify the time resolution of the datafiles (2ms, 8ms, 16ms, 64ms, 256ms, 1s and 10s) and analyze the data by selecting only some of the energy bands.

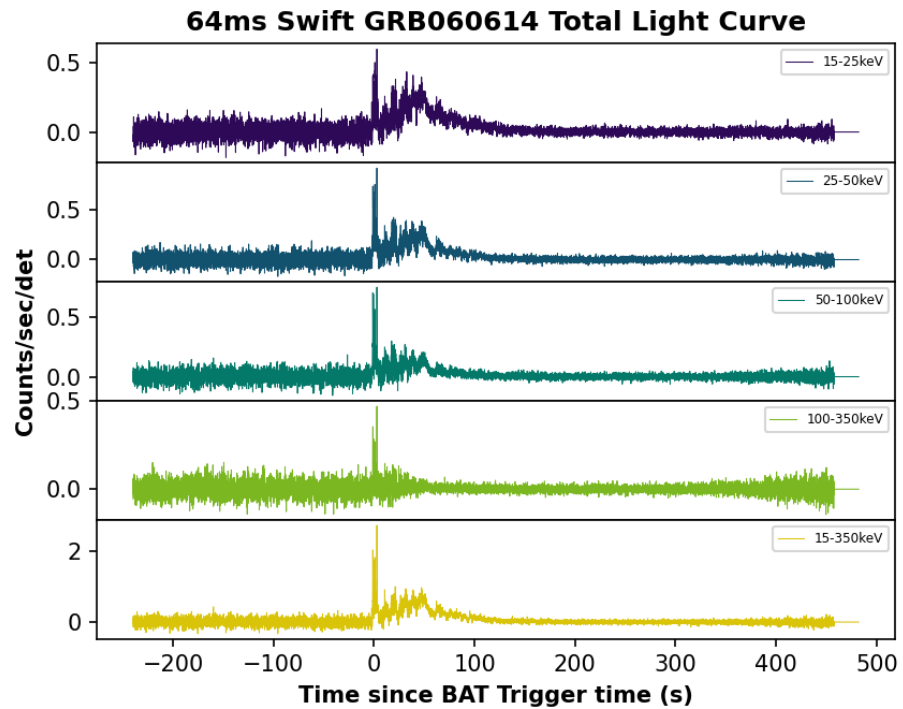


Figure 2: Light curve of GRB 060614A. Image taken from (Garcia-Cifuentes et al., 2023)

2) Data processing. ClassipyGRB is able to:

- constrain the light curves by their duration T_{100} , T_{90} or T_{50} .
- normalise the flux.
- standardise the temporal interval of all the sample (by zero-padding).
- improve the signal/noise (S/N) ratio applying the non-parametric noise reduction technique called FABADA (Sanchez-Alarcon & Ascasibar Sequeiros, 2022) to each band for every single light curve.
- interpolate the flux between two specific times.

3) Visualization and plot of t-SNE maps

ClassipyGRB produces 2D visualization maps colored by the duration of GRBs. It includes:

- Intuitive graphic interface.
- It is possible to add either of the two features to the t-SNE maps or to visualize the raw data.
- Manipulation of the t-SNE parameters.
- Visualization of the light curves from the events with and without noise-reduction.
- There is the possibility of working only on the desired bands of Swift/BAT.
- Specific events can be searched for and highlighted on the display.

Moreover, any plot created with ClassipyGRB can be customized by the user.

Note: Algorithms such as t-SNE visualization maps are very sensitive to any change in the perplexity and learning rate parameters. Therefore, as is the case when using any of these visualization techniques derived from machine learning, care must be taken in the correct interpretation of the data.

This repository requires Python 3.8 or high, and the packages from the [requeriments.txt](#) file. Other packages will be required optionally in Documentation (i.e., Jupyter).

Acknowledgements

KSGC acknowledges support from CONACyT fellowship. RLB acknowledges support from CONACyT postdoctoral fellowship.

References

- Garcia-Cifuentes, K., Becerra, R. L., De Colle, F., Cabrera, J. I., & del Burgo, C. (2023). Identification of Extended Emission Gamma-Ray-Bursts Candidates using Machine Learning. *arXiv e-Prints*, arXiv:2304.08666. <https://doi.org/10.48550/arXiv.2304.08666>
- Jespersen, C. K., Severin, J. B., Steinhardt, C. L., Vinther, J., Fynbo, J. P. U., Selsing, J., & Watson, D. (2020). An Unambiguous Separation of Gamma-Ray Bursts into Two Classes from Prompt Emission Alone. *Astrophysical Journal Letters*, 896(2), L20. <https://doi.org/10.3847/2041-8213/ab964d>
- Kouveliotou, C., Meegan, C. A., Fishman, G. J., Bhat, N. P., Briggs, M. S., Koshut, T. M., Paciesas, W. S., & Pendleton, G. N. (1993). Identification of Two Classes of Gamma-Ray Bursts. *Astrophysical Journal Letters*, 413, L101. <https://doi.org/10.1086/186969>
- Sanchez-Alarcon, P. M., & Ascasibar Sequeiros, Y. (2022). Fully Adaptive Bayesian Algorithm for Data Analysis, FABADA. *arXiv e-Prints*, arXiv:2201.05145. <https://doi.org/10.48550/arXiv.2201.05145>
- Steinhardt, C. L., Mann, W. J., Rusakov, V., & Jespersen, C. K. (2023). Classification of BATSE, Swift, and Fermi Gamma-Ray Bursts from Prompt Emission Alone. *The Astrophysical Journal*, 945(1), 67. <https://doi.org/10.3847/1538-4357/acb999>
- Yang, J., Ai, S., Zhang, B.-B., Zhang, B., Liu, Z.-K., Wang, X. I., Yang, Y.-H., Yin, Y.-H., Li, Y., & Lü, H.-J. (2022). A long-duration gamma-ray burst with a peculiar origin. *Nature*, 612(7939), 232–235. <https://doi.org/10.1038/s41586-022-05403-8>