Gamma-Ray Burst Classification: New Insights from Mining Pulse Data

Stanley McAfee
Department of Physics and Astronomy
College of Charleston

Advisor: Jon Hakkila
Department of Physics and Astronomy, College of Charleston

(Dated: April 19, 2017)

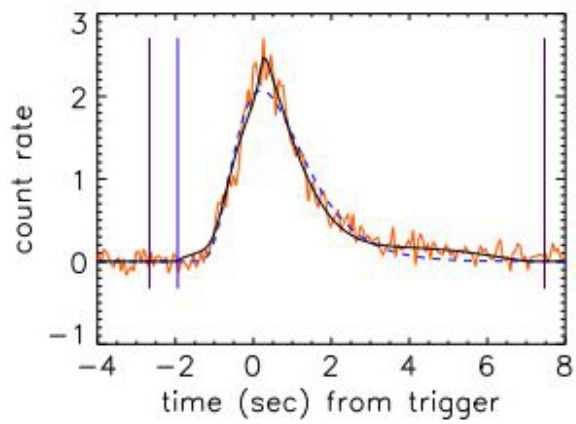
Despite being the most energetic electromagnetic explosions in the universe, gamma-ray bursts (GRBs) are still poorly understood. The literature recognizes two potentially different types of GRB progenitors, although statistical data suggest the existence of three GRB classes. Reliable inference of GRB physics depends on the identificaton of appropriate classification attributes, as well as on the statistical classification techniques used. It has recently been shown that pulses are the basic unit of GRB emission. We will use new data describing GRB pulse characteristics, in conjunction with data mining tools, to provide a more reliable gamma-ray burst classification system, and therefore provide additional constraints on GRB physics.

I. INTRODUCTION

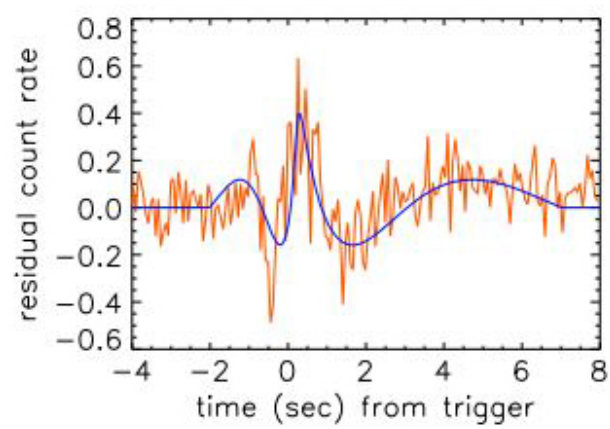
Gamma-ray bursts (GRBs) are brief emissions of high-energy photons with durations ranging from 10^{-1} to 10^{3} seconds. The initial flash of γ-rays is known as the *prompt emission*, distinct from the lower-energy afterglow that persists following the burst [1]. The *burst* itself is the totality of the original emission episode, which begins with a count rate above the background level and ends when the count rate has returned to the background level. As the GRB log-normal duration distribution is roughly bimodal, with a separation at $T_{90} \approx 2$s, a common definition of burst duration is T_{90}, the amount of time it takes to detect 90% of the prompt emission [2].

However, this classification scheme is an oversimplification. Several statistical analyses of GRB data have reliably found *three* classes of bursts: Short, Long, and Intermediate. Short bursts have shorter durations, lower fluences, and harder spectra than Long bursts. Intermediate bursts have intermediate durations, intermediate fluences, and soft spectra [3][4]. Most bursts with $T_{90} < 2$s are Short. While Short and Long bursts are understood to be distinct populations, with physical significance, this is not presently true for Intermediate bursts, which may result from the effects of systematic instrumental biases acting on Long bursts. In any case, these issues highlight the fact that GRB classification is a hard problem [5].

A *pulse* is the fundamental component of burst prompt emission. All bursts, regardless of class, share an underlying pulse structure in their prompt emission. Pulses are what are being fit in GRB data, as illustrated in Figure (1a). The pulses themselves share many common characteristics [6][7]. In general, pulses take longer to decay than to rise, evolve from hard to soft energies, and last longer at low energies than they do at high energies. Short-duration pulses have higher intensities and more symmetric rise-fall times than long-duration pulses [8].



(a) The pulse structure common to GRBs. Vertical lines indicate a deviation from the background determined by the pulse fitting algorithm.



(b) The characteristic residual structure of pulse fits.

FIG. 1: GRB 80413B, a representative single-pulsed GRB. See Figure 11 from Hakkila et al. 2015. The solid lines are the fits, and the orange line is the light curve.

Pulse properties such as duration, hardness, and fluence are correlated between pulses in both Long and Short bursts, and the properties of the longest pulses in Short bursts are similar to the properties of shortest pulses in Long bursts, suggesting that the process that produces them might be the same [7]. Additionally, GRB pulses exhibit triple-peaked structures with spectra that re-harden at or prior to their peaks [8]. Recent evidence indicates that complex, highly variable structures in GRB light curves are also a variation on pulse behavior [9]. Though these observations place constraints on GRB physics, the mechanism that produces pulses remains unknown.

In the 1990s, the Burst and Transient Source Experiment (BATSE) on NASA's Compton Gamma-Ray Observatory observed many GRBs. Though it has since been de-orbited, BATSE offers several advantages over modern instruments in the study of GRB prompt emission, namely its large surface area and moderate energy response [7]. Recent work has been done at the College of Charleston to fit pulses to bursts in the BATSE archive, creating a large catalogue of pulses for analysis.

II. GOALS

Classifying events or objects requires determining the essential properties that uniquely define them. In the case of GRBs, it is not always clear which properties provide the greatest insights into their classification. To this end, data mining tools have been applied to GRB data in the past, and have been successful in identifying pertinent burst properties such as T_{90} duration, fluence, and spectral hardness [5][11][12].

Data mining tools require appropriate *attributes* in order to obtain robust classification results. Given their success in classifying *bursts*, it is possible that data mining tools can provide better classification with the use of *pulse* properties. Are these properties similar for pulses in different classes of bursts? Do all pulses behave in a similar way, or are there different classes of pulses as well? Answering these questions, and seeing how similar or dissimilar they are to the results for bursts, will aid in constructing a more robust classification scheme for GRBs.

It must be noted that while data mining tools are highly useful ways to find related GRBs, they are little more than *tools*, and must be applied correctly to achieve a desired result. Critically, they do not necessarily guarantee the significance, physical or otherwise, of the clusters they discover. Therefore, as there is no "correct" way to classify a dataset, another question this work will seek to address is how these tools can be applied to pulse data to result in useful information.

This project will further characterize the behaviors of GRB pulses, and also offers the possibility of discovery of unique populations in the BATSE data, which may not be GRBs. If some pattern of pulse behaviors is discovered, the findings of this project may also place additional constraints on GRB physics.

III. METHOD

Using existing Interactive Data Language (IDL) programs, pulses in the BATSE archive will be fit to the standard empirical GRB pulse model [10]. This model is defined as

$$I(t) = A\lambda e^{-\tau_1/(t-t_s)-(t-t_s)/\tau_2} \tag{1}$$

where A is the pulse amplitude, t is the time elapsed since the trigger event, t_s is the pulse start time, τ_1 and τ_2 are respectively the pulse rise and decay parameters, and $\lambda = \exp[2(\tau_1/\tau_2)^{1/2}]$ is the normalization constant. The parameters of the pulse model determine pulse properties such as amplitude, duration, and asymmetry [13].

GRB pulse fits also have a characteristic residual structure that can be described with an empirical model [8]. This model is defined as

$$\mathrm{res}(t) = \begin{cases} AJ_0\sqrt{\Omega(t_0-t-\Delta/2)} & \text{if } t < t_0-\Delta/2 \\ A & \text{if } t_0-\Delta/2 \leq t \leq t_0+\Delta/2 \\ AJ_0\sqrt{s\Omega(t_0-t-\Delta/2)} & \text{if } t > t_0+\Delta/2 \end{cases}$$

where $J_0(x)$ is an integer Bessel function of the first kind, t_0 is the central time of the peak amplitude, A is the normalized amplitude of the peak, Δ is the duration of the peak, Ω is the Bessel function's angular frequency, and s is a scaling factor. The Bessel function describes the "wave-like" nature of the residuals. The parameters of the residual model correspond to pulse properties, and assist in determining pulse structure and behavior.

Though a catalogue of pulses has recently been created for Short bursts, more fits will need to be performed to achieve a complete, unbiased sample of Long and Intermediate pulses. As some data mining tools are sensitive to how the data is presented, and not just what data is available, a *complete* sample containing regular pulses of known behavior as well as irregular pulses with unknown or non-pulsed behavior is particularly important. A sequential sub-catalogue of GRBs with measured properties and classified pulses will be created for this purpose.

Once an appropriate sample of pulses has been selected, data mining tools will be applied to the dataset. The University of Waikato publishes Waikato Environment for Knowledge Analysis (Weka), a suite of machine learning algorithms. As each algorithm takes slightly different assumptions, and produces slightly different results, a variety of different algorithms will be used to discover statistical clusterings of pulses. There are two primary types of machine learning algorithms. Supervised algorithms are "trained" on known data to determine why a pulse belongs to a particular category, and

then applied to unknown data to categorize it; unsupervised algorithms are blindly applied to unknown data to discover any groupings that may appear. Both types will be used, both to check the results of the other (when possible) and to glean information the other may have missed as a result of a faulty heuristic.

IV. RESOURCES

Pulse fitting will be performed using programs written in IDL, the standard programming language for astronomical computing. All software licenses and computer hardware necessary for this purpose are owned by the department and easily acessible. The Weka machine learning suite is open source and publically available. Access to the BATSE archive has already been secured by the project advisor.

V. BUDGET

No department funding is necessary for this project.

VI. TIMELINE

- Produce sequential sub-catalogue of pulse fits for all burst classes, become familiar with data mining tools (Summer 2017)
- Identify and define pulse properties and extract them into database (Summer – September 2017)
- Perform cluster analysis on database (September – November 2017)
- Create plots of pulse properties against each other and look for correlations in pulse attributes; explain class behaviors (November 2017 – February 2018)
- Interpret results, begin to write paper and create poster (February 2018)
- Finish paper and poster, present at departmental poster session/South Carolina Academy of Sciences meeting (March – April 2018)

[1] Piran, T., 2004, arXiv:astro-ph/0405503
[2] Kouveliotou, C., et al. 1993, 1993ApJ...413L.101K
[3] Mukherjee, S., et al. 1998, arXiv:astro-ph/9802085
[4] Horvath, I. 1998, ApJ, 508, 757
[5] Hakkila, J., et al. 2003, arXiv:astro-ph/0209073v1
[6] Norris, J. P., Nemiroff, R. J., Bonnell, J. T., et al. 1996, ApJ, 459, 393
[7] Hakkila, J. and Preece, R.D., 2011, arXiv:1103.5434
[8] Hakkila, J. and Preece, R.D., 2014, arXiv:1401.4047
[9] Hofesmann, E. 2016, "Characterizing the Complexity of GRB Light Curves," `hofesmannen.stu.cofc.edu/files/GRB\%20Senior\%20Research.pdf`
[10] Norris, J.P., Bonnell, J. T., Kazanas, D. et al. 2005, arXiv:astro-ph/0503383
[11] Horvath, I., Balazs, L. G., Bagoly, Z., Ryde, F., Meszaros, A. 2006, arXiv:astro-ph/0509909v2
[12] Hakkila, J., Roiger, R. J., Haglin, D. J., et al. 2000, arXiv:astro-ph/0011583
[13] Hakkila, J., Lien, A., Sakamoto, T., et al. 2015, arXiv:1511.03695v1