Physics and Astronomy

Acta Spartae Vol. - No. 1 20–
Pages 1–4

Burst Chaser: Classifying Gamma-Ray Bursts with Citizen Science and Machine Learning

Carter Murawski and Amy Lien[1]

Department of Physics & Astronomy, University of Tampa, Tampa, FL 33606, [1]*FacultyAdvisor*

ABSTRACT

The Burst Chaser project aims to evaluate the accuracy of using citizen science to classify Gamma Ray Bursts (GRBs) and to identify methods for enhancing that accuracy. Through the Zooniverse platform, 1,480 gamma ray bursts were distributed to scientists of varying skill levels worldwide, providing an opportunity for classification, discussion, and reporting of interesting findings. The accuracy of the citizen scientist classifications was evaluated using a golden sample, with a resulting accuracy of 58%. The classifications of the remaining bursts were then verified using proportional analysis and statistical methods. Using the verified labels, a machine learning algorithm was developed that increased the accuracy on the golden sample to 80%.

1 INTRODUCTION

Gamma-Ray Bursts are explosions in space that can be detected out to large distances because of their extraordinary brightness. These events are typically associated with high energy phenomenon such as the death of massive starts or the merger of compact stellar remnants and can release more energy in a few seconds than the Sun will over its entire lifetime. Understanding GRBs is important for gaining insight into the early universe, stellar evolution, and the physics of extreme environments.

Because of their variability in both shape and duration, GRBs are difficult to classify. The light curves produced by GRBs can differ widely, ranging from sharp, simple pulses to long, complex sequences with multiple peaks. As the number of observed GRBs increases, classifying them becomes increasingly difficult.

To address this, the Burst Chaser project uses citizen science to assist in classification. Volunteers on the Zooniverse platform were presented with light curve graphs and asked to assign each burst to a category based on its shape. This approach builds on the idea that people, regardless of formal training, can be effective at spotting visual patterns in data.

However, while the human brain is excellent at visual recognition, it is still prone to bias and inconsistency. To improve accuracy and scalability, the project combined human-labeled data with machine learning techniques. Using statistical methods to verify labels and reduce uncertainty, a machine learning model was trained to learn from these patterns and apply them at a much larger scale. By combining citizen science with machine learning, this study demonstrates a collaborative method to improve GRB classification and support future astrophysical research.

2 METHODS

Gathering Citizen Science Classifications

This study aims to classify gamma ray bursts with assistance from citizen scientists. The Zooniverse platform was used, providing access to a large number of participants. These scientists were presented with a sample of 1,480 Gamma Ray Bursts in the form of light curve graphs. The light curves were distributed through workflows that instructed participants to both classify and locate the pulses. However, to explore the effectiveness of citizen science, only the classification data was used. This approach allows for the evaluation of citizen scientist accuracy.

To support the training and evaluation of the classification data, 48 bursts were labeled by professional scientists, accompanied by explanations of the reasoning behind each classification. These 48 bursts were included in a tutorial workflow designed to provide feedback to citizen scientists, with the aim of improving classification accuracy. This golden sample also serves as a benchmark for comparing citizen science and machine learning results.

For the sake of simplicity, this study focuses on the categorization of bursts into four groups: simple, extended, other, and too noisy. Simplifying the classification task facilitates pattern recognition by both citizen scientists and machine learning algorithms.

Proportions Verification

With over 64,000 classifications from citizen scientists, a method for verifying these classifications was necessary. Initially, the results of the citizen science categorization data were compared on an individual basis. Each burst had counts in each category, which could be normalized to account for the varying number of classifications per burst. To verify a burst, those with over 70% of classifications in a single category were considered verified as belonging to that category. In Table 1, GRB160630A exceeded the 70% threshold in the extended category and was therefore classified as an extended burst.

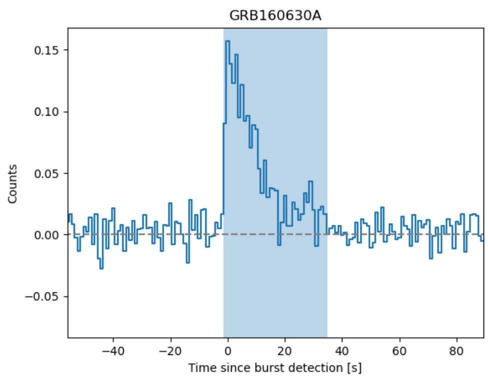

Figure 1. Light Curve Data of GRB160630A

Category	Simple	Extended	Other	Too Noisy
Counts	3	13	2	0
Proportions	0.167	0.722	0.111	0

Table 1. Vote Counts for GRB160630A

When analyzing the results using the 70% threshold, some bursts showed a strong leaning toward two out of the four categories. In these cases, there was clear evidence supporting classification in either category. To account for these discrepancies, the option to classify a burst as a combination of two categories was introduced. If a burst received over 40% of votes in two categories, it was considered to belong to both. This approach acknowledges that some bursts may not fit neatly into a single category. An example of this is GRB150301B, which exhibits a smaller tail than most extended bursts, making it appear similar to a simple burst as well.

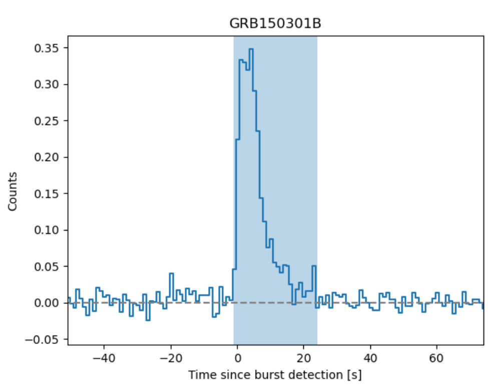

Figure 2. Light Curve Data of GRB150301B

Category	Simple	Extended	Other	Too Noisy
Counts	27	23	2	3
Proportions	0.491	0.418	0.036	0.055

Table 2. Vote Counts for GRB150301B

Statistical Classification

Classifying gamma-ray bursts using proportion data proved to be simple and effective for identifying bursts that clearly fit into a single category. However, this approach led to the classification of 242 Simple, 13 Extended, 151 Other, and 28 Too Noisy bursts. From this, it might appear that there are significantly more simple and other gamma-ray bursts than extended and too noisy ones. However, certain bursts possessed specific features that were often missed by most citizen scientists. For example, GRB110503A was categorized as a simple burst using the proportions method, but the long, extended blue region suggests the telescope detected activity beyond the initial pulse. Although it may visually resemble a simple burst, the light curve data indicates extended emission.

While the proportions-based classification is straightforward and useful, the potential for human categorization error remained significant. To address this limitation, an alternative approach was explored. Instead of evaluating how individual bursts were classified by citizen scientists, the analysis focused on how all bursts were voted on across each category. As shown in Figure 5, the simple and other categories exhibit a wider distribution of votes, whereas the too noisy and extended categories display a narrower distribution. Although this may be expected given the lower number of bursts in these categories, even the outliers in the extended and too noisy groups fall below the original threshold for classification.

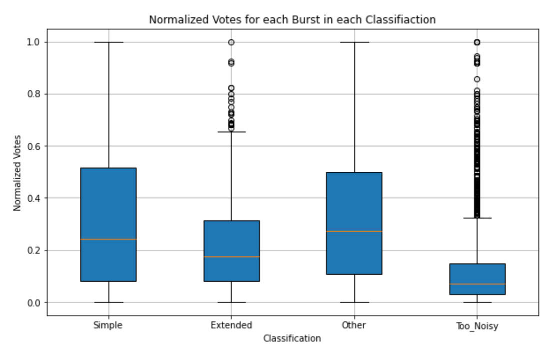

Figure 3. A box and whisker plot showing the distribution of votes for Gamma Ray Bursts.

To address the differences between categories, a statistical test was developed to classify the bursts more rigorously. When analyzing how a burst was voted on individually, it was observed that a category receiving 0% of the votes could reasonably be excluded. Therefore, if it can be shown that the proportion of votes for a specific category is significantly different from zero, the burst can be classified under that category. This method enables the identification of bursts with distinguishing features that consistently lead participants to classify them in a particular way.

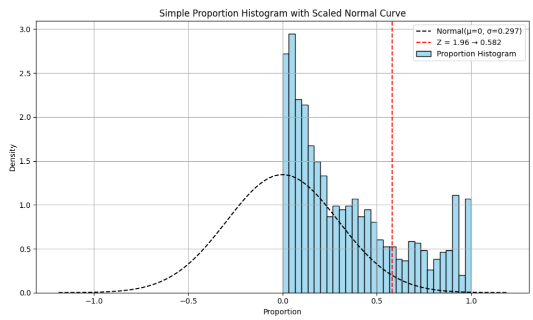

Figure 4. Histogram for Simple Category

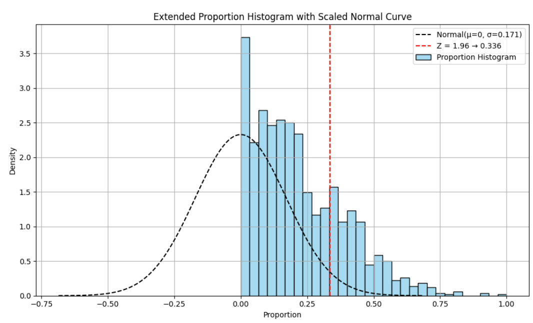

Figure 5. Histogram for Extended Category

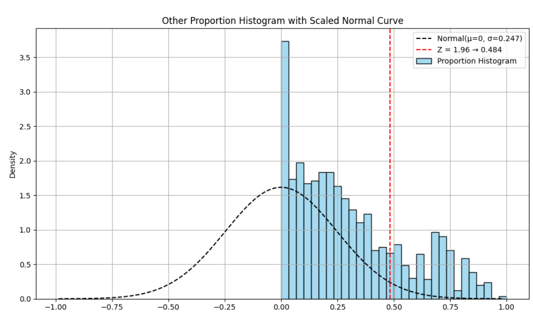

Figure 6. Histogram for Other Category

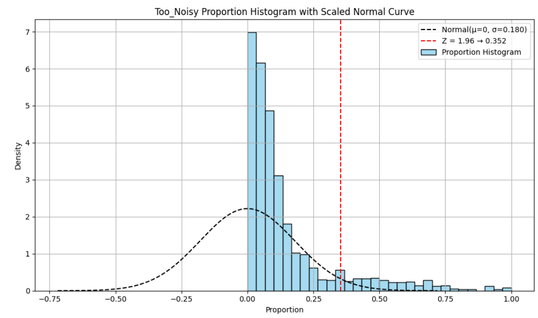

Figure 7. Histogram for Too Noisy Category

Based on this, if a burst sample had a proportion greater than the population p-value threshold of 0.05, it could be classified into that category. This test was also performed using a p-value of 0.01 to generate multiple labeled datasets for training the machine learning model. By analyzing the data using a distribution-based method rather than raw proportions, the number of bursts verified in each class became more balanced, allowing bursts with distinct features to remain.

Machine Learning Classification

To enhance the accuracy of the classification process, the citizen scientist data was utilized to create a machine learning model. Given that the citizen scientists interacted with light curves in graphical form, the decision was made to process the light curve data as images, mirroring how the data was presented to the participants. This allowed the model to learn from the visual patterns that citizen scientists used in their classifications. Prior to training the model, a statistical method was applied to verify the burst classifications and ensure they were correctly assigned. The statistical test, as explained previously, was used to determine if a burst should be classified into one or more categories based on the proportions of votes it received.

For the machine learning model, a Random Forest classifier was selected from the Scikit-learn library. Random Forest is a powerful ensemble learning method that creates multiple decision trees and merges them together to improve classification accuracy. This model is particularly useful for handling complex, high-dimensional data and provides robust results even with noise in the input data Scikit-learn (n.d.). The Random Forest algorithm's ability to assess feature importance and its resistance to overfitting make it ideal for this task.

After training the model with the citizen scientist data, the testing accuracy was found to be 93%, with a training accuracy of 100%. These results indicate that the model effectively captured the patterns in the data and outperformed the initial citizen scientist classifications.

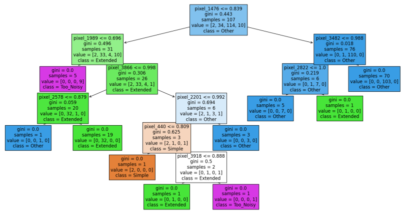

Figure 8. Decision Tree from the Random Forrest Classifier

3 RESULTS

The analysis of the citizen scientist classifications revealed an accuracy of 58% on the golden sample. These classifications were then used to train a machine learning model, which was designed to improve the classification process. By leveraging the model, the accuracy of the golden sample was increased to 80%. This improvement demonstrates the potential of combining citizen science with machine learning techniques to enhance classification accuracy.

4 DISCUSSION/CONCLUSION

The primary objective of this citizen science project was to enhance the classification of astronomical phenomena by leveraging the pattern-recognition capabilities of non-expert participants. Analysis of the collected data demonstrated that accurate identification of shapes and patterns within astronomical graphs does not always require expert knowledge. However, interpretation can be challenging without familiarity with the subject matter.

To address this limitation, machine learning techniques can be employed. By using the categorization data generated by citizen scientists, consistent and commonly agreed-upon patterns can be extracted. These patterns serve as a training dataset for a machine learning model, allowing it to emulate collective human insight. The resulting model has the potential to significantly improve the efficiency and accuracy of astronomical classification tasks, representing a fusion of human intuition and computational scalability.

REFERENCES

Fryer, C. L., Ellinger, C., Smidt, J., & Wollaeger, R. 2021 arXiv:2112.00643

Moss, M., Lien, A., Kocevski, D., et al. 2022 arXiv:2111.13392

Scikit-learn developers. n.d. Scikit-learn Documentation