PROJ 201 Project Proposal Report

Project Title: Morphological Classification of Magnetar Bursts using High-Energy Astrophysical Data

Name, Surname & ID of group members:
Şefik Efe Altınoluk 35801
Muhammed Abtaha Farooq 36715

Supervised by:Ersin Göğüş

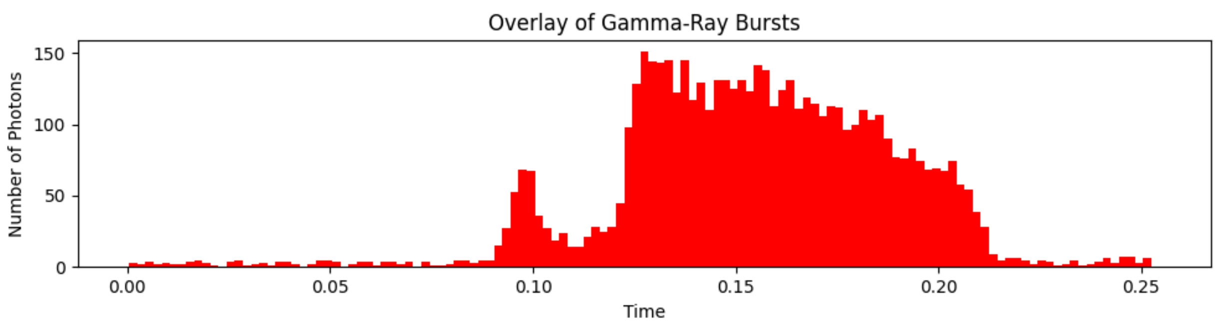
1/11/2024

Abstract

This project focuses on studying magnetars, a small class of highly magnetized neutron stars, which exhibit short but powerful bursts of energy in the form of hard X-rays and gamma rays. These bursts are caused by crustal fractures in the star due to extreme magnetic stresses. Using data from NASA's Fermi Gamma-ray Burst Monitor (GBM), we shall analyze magnetar bursts by generating light curves across various energy bands. We shall filter the GBM data, plot the light curves of each burst, and classify them based on their morphological characteristics using specialized machine learning techniques. The analysis aims to enhance understanding of the unique features of magnetar bursts.

Introduction

Magnetars, a rare and highly magnetized type of neutron star, are known for their intense bursts of hard X-rays and gamma rays, often triggered by fractures in their magnetically stressed crusts. These bursts are typically detected in the form of light curves, which graph the intensity of the emissions over time, providing a visual representation of their energy release. The light curve is an essential tool for studying these events and reveals the rapid and intense nature of magnetar bursts.



Our awareness of magnetar bursts has been significantly enhanced by the Fermi Gamma-ray Space Telescope, equipped with the Gamma-ray Burst Monitor (GBM), which detects and monitors these high-energy events. Fermi, a space-based observatory launched in 2008, has expanded our capacity to observe these phenomena, offering critical data on the temporal and spectral characteristics of bursts. The GBM is specifically designed to detect and monitor high-energy events from 8 keV to 40 MeV, making it well-suited for studying transient cosmic phenomena such as gamma-ray bursts (GRBs), solar flares, and magnetar outbursts.

Leveraging data from Fermi's GBM, we shall use machine learning algorithms, specifically neural networks, support vector machines (SVMs), and random forest classifiers, to systematically and morphologically categorize these bursts, offering new insights into the extreme environments that drive such powerful emissions.

Purpose

The overarching aim of this study is to enhance our understanding of magnetar bursts through a systematic analysis and classification of high-energy events utilizing hard X-ray data obtained from NASA's Fermi Gamma-ray Burst Monitor (GBM). By implementing a structured approach to filter and extract relevant magnetar burst data, we shall facilitate the generation of light curves across various energy bands, thereby allowing for a comprehensive examination of the temporal and energy distribution characteristics of these phenomena.

Furthermore, the application of machine learning algorithms to classify magnetar bursts based on their morphological features, such as intensity and duration, represents a significant advancement in the automation of data analysis in astrophysics. This methodological innovation not only promises to uncover previously obscured patterns within the data but also enhances the potential for identifying correlations that may elucidate the underlying mechanisms of magnetar behavior.

Ultimately, this study aims to contribute meaningfully to the field of high-energy astrophysics by integrating modern computational techniques with astrophysical observations. The insights gained from this research may provide a deeper understanding of the dynamics governing magnetar emissions and facilitate future investigations into these rare yet potent cosmic phenomena. Through this endeavor, we hope to advance the knowledge of high-energy astrophysics, paving the way for further exploration of the extreme conditions present in the universe.

Description of Specific Steps

Following an exploration of magnetar bursts and the tools necessary for their analysis, the project will focus on selecting classification parameters from a limited set of magnetar burst data. The approach will involve several distinct phases, each employing specific machine learning techniques.

1. **Data Acquisition and Preprocessing**:
    - **Data Collection**: Gather data from NASA's Fermi Gamma-ray Burst Monitor (GBM), including various energy bands and associated burst characteristics. For now, a subset of data has been collected. Over the course of this project, we shall have more than a 100 samples of burst data collected from the GBM.
    - **Data Cleaning**: Remove any irrelevant or corrupted entries in the dataset, ensuring high-quality data for analysis.
    - **Feature Extraction**: Extract relevant morphological features, including:
        - **Amplitude-based Features**:
            - **Peak Intensity**: The maximum number of photons detected during the burst.

        - **Mean and Standard Deviation**: Calculate the mean and standard deviation of photon arrival times or energy levels to capture burst variability.
        - **Signal-to-Noise Ratio**: Assess the prominence of bursts compared to background noise.
    - **Time-based Features**:
        - **Rise Time**: The time taken for the burst to reach its peak intensity.
        - **Decay Time**: The time taken for the burst intensity to return to baseline.
        - **Duration**: The total time span of the burst from onset to end.
        - **Centroid**: The time at which the burst is centered.
        - **Skewness**: Measure the asymmetry of the burst profile.
    - **Morphological Features**:
        - **Area Under the Curve**: Integrate the histogram or time-series data to calculate the total energy of the burst.
        - **Morphological Transforms**: Apply binary dilation and erosion, and calculate the number of distinct regions post-dilation to indicate burst complexity.

2. **Feature Engineering**:
    - Transform and normalize the extracted features to prepare them for machine learning algorithms. This step may include scaling the features using the softmax function to ensure that all variables contribute equally to the model training.
3. **Labeling the Data**:
    - Define labels based on the extracted features to classify the bursts. Tentatively we have decided on the following:
        - **Duration-based Labels**: Short, Intermediate, Long.
        - **Energy-based Labels**: Low Energy, Medium Energy, High Energy.
        - **Shape-based Labels**: Single-Peak, Multi-Peak, Complex.
4. **Training and Testing Split**:
    - Divide the dataset into training and testing subsets using an 80-20, to evaluate the model's performance effectively. This ensures that the model can generalize well to unseen data.
    - Ensure that the models do not overfit by using techniques such as regularization, dropout in neural networks, and cross-validation during model evaluation.
5. **Model Development**:
    - **Neural Networks**:
        - Design and implement a neural network architecture suitable for classifying magnetar bursts based on the extracted features. This may involve using a feedforward neural network with one or more hidden layers.
        - **Activation Functions**: Employ activation functions like ReLU (Rectified Linear Unit) for hidden layers to capture non-linear relationships and a softmax activation function in the output layer for multi-class classification.

        - **Training**: Train the model using backpropagation and an optimizer (such as Adam) to minimize the loss function, typically categorical cross-entropy for multi-class tasks.
        - **Evaluation**: Assess the model's performance on the testing dataset using metrics like accuracy, precision, recall, and F1-score.
    - **Support Vector Machines (SVMs)**:
        - Implement an SVM classifier to separate different classes of magnetar bursts based on the feature set. SVMs are particularly effective for high-dimensional data.
        - **Kernel Selection**: Experiment with different kernel functions (linear, polynomial, radial basis function) to find the best fit for the data.
        - **Training and Testing**: Train the SVM model on the training dataset and evaluate its performance on the testing dataset using similar metrics as for the neural network.
    - **Random Forest Classifiers**:
        - Utilize random forests as an ensemble learning method that builds multiple decision trees and merges them to improve classification accuracy.
        - **Feature Importance**: Analyze feature importance scores to determine which morphological characteristics contribute most significantly to the classification, aiding in further feature selection.
        - **Training and Evaluation**: Train the random forest classifier on the training dataset and evaluate its performance on the testing dataset, focusing on metrics like accuracy and confusion matrix to identify any misclassifications.
6. **Model Comparison and Selection**:
    - Compare the performance of the three models (neural networks, SVMs, and random forests) using cross-validation techniques and evaluation metrics. This comparison will help identify the most effective approach for classifying magnetar bursts.
    - Select the best-performing model for further analysis and refinement.
7. **Final Model Training and Validation**:
    - Conduct hyperparameter tuning for the selected model(s) using techniques like grid search or random search to optimize model parameters (e.g., learning rate, number of trees in the random forest, regularization parameters in SVM).
    - Train the final model on the entire dataset (combining training and testing sets) to maximize its performance.
    - Validate the model against additional datasets, if available, to assess its robustness and generalizability.
8. **Result Interpretation and Visualization**:
    - Visualize the classification results through confusion matrices, ROC curves, and feature importance plots to interpret the model's performance and understand the underlying patterns in magnetar bursts.

- Generate reports detailing the findings, including key morphological features that distinguish different types of bursts.

We shall utilize a sprint approach for each of these steps, distributing tasks across team members. As this is a research project, we remain flexible to adapt and modify features, labels, and models as necessary throughout the process.

By following these steps, our project aims to develop a comprehensive framework for the classification of magnetar bursts using advanced machine learning techniques. The insights gained from this analysis will contribute to a deeper understanding of the unique characteristics of these high-energy events and the physical processes driving them.

Description of Responsibilities of Individual Members

This section outlines the responsibilities of each team member, including their specific tasks, time allocation, and meeting schedules.

Muhammad Abtaha Farooq

- **Primary Responsibilities:**
  - Lead the data acquisition and preprocessing efforts, including data collection and cleaning.
  - Oversee feature extraction and engineering processes to ensure high-quality input for machine learning models.
  - Design and implement neural network architectures, including training and evaluation.
  - Conduct hyperparameter tuning and validation for the selected model(s).
  - Edit the documentation and visualization processes
  - Oversee the overall coherence and clarity of project documents, ensuring high-quality writing and adherence to academic standards.
  - Finalize and format reports and presentations, ensuring consistency in formatting and citation style.
- **Time Allocation:**
  - Devote approximately 10 hours per week to project-related tasks, focusing on model development and analysis.

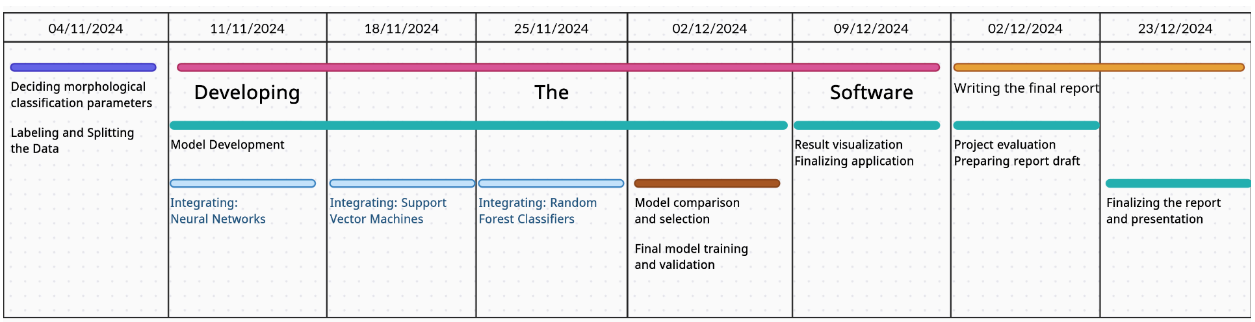
Şefik Efe Altınoluk

- **Primary Responsibilities:**
  - Assist in data collection and preprocessing tasks, including data cleaning and basic feature extraction.
  - Support the implementation and evaluation of Support Vector Machine (SVM) and Random Forest classifiers.
  - Participate in model training and testing processes
  - Draft sections of the project report, focusing on data collection, methodology, and results.
  - Prepare visualizations and summaries of findings to support the written content.
- **Time Allocation:**
  - Allocate around 8 hours per week to project-related tasks, focusing on research and assisting with model implementation.

Meeting Schedule:

- The team will meet with their advisor every Monday at 1:40 PM to discuss project progress, address challenges, and plan next steps. In addition, informal check-ins will be held weekly to ensure ongoing collaboration and alignment on project goals.

Gantt Chart



References

1. NASA. (n.d.). *Fermi Gamma-ray Space Telescope overview*. Retrieved from https://www.nasa.gov/mission_pages/GLAST/overview/index.html
   Fermi GBM Team. (n.d.). *Fermi Gamma-ray Burst Monitor (GBM) homepage*. Retrieved from https://heasarc.gsfc.nasa.gov/W3Browse/fermi/fermigbrst.html
2. Wikipedia contributors. (2023). *Magnetar*. In *Wikipedia, The Free Encyclopedia*. Retrieved from https://en.wikipedia.org/wiki/Magnetar
3. Wikipedia contributors. (2023). *Gamma-ray burst*. In *Wikipedia, The Free Encyclopedia*. Retrieved from https://en.wikipedia.org/wiki/Gamma-ray_burst
4. Rea, N., & Esposito, P. (2011). Magnetar outbursts. In *The Astronomy and Astrophysics Review*, 19(1), 9-25. https://doi.org/10.1007/s00159-010-0021-5
5. Cavanagh, B., et al. (2021). An introduction to machine learning in astronomy. *Publications of the Astronomical Society of the Pacific*, 133(1021), 075001. https://doi.org/10.1088/1538-3873/abf5a9
6. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. Retrieved from https://www.deeplearningbook.org
7. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. https://doi.org/10.1007/BF00994018
   Schölkopf, B., & Smola, A. J. (2002). *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. MIT Press.
8. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324
   Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81-106. https://doi.org/10.1007/BF00116251
9. Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning* (ICML-10) (pp. 807-814). Retrieved from http://www.icml-2010.org/papers/432.pdf
   Dhingra, B., et al. (2017). A framework for improving the convergence of deep learning models. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 1065-1075). https://doi.org/10.18653/v1/D17-1124
   Zhang, Y., et al. (2019). Softmax activation function in neural networks: A practical guide. *IEEE Access*, 7, 98827-98839. https://doi.org/10.1109/ACCESS.2019.2921382
10. NASA Goddard Space Flight Center. Retrieved from https://www.nasa.gov/goddard
    Fermi Science Support Center. Retrieved from https://fermi.gsfc.nasa.gov/ssc/
11. SciPy. (n.d.). SciPy: Scientific Library for Python. Retrieved from https://www.scipy.org/
12. PyTorch. (n.d.). PyTorch: An open source machine learning framework. Retrieved from https://pytorch.org/
13. Matplotlib. (n.d.). Matplotlib: Visualization with Python. Retrieved from https://matplotlib.org/
14. Pandas. (n.d.). Pandas: Powerful data structures for data analysis in Python. Retrieved from https://pandas.pydata.org/