PROCEEDINGS OF SCIENCE

Search and Machine Learning Classification of Short Untriggered Transients

Martin Topinka[∗†]
Czech Technical University, Technická 2, Prague 6, 162 00, Czech Republic
E-mail: martin.topinka@gmail.com

René Hudec
Czech Technical University, Technická 2, Prague 6, 162 00, Czech Republic

An intense search through the INTEGRAL IBIS/ISGRI data archive revealed a large number of untriggered flares on a millisecond time scale. As the IBIS detector uses a coded mask, localisation of those flares is hard due to low number statistics in the deconvolution process. The identified flares are used as an input for further data processing. A machine learning clustering algorithm reveals several groups of objects. This fast classification method can separate e.g. short gamma-ray bursts from soft gamma-ray repeater flares. Discriminating between different classes of objects allows to adjust strategies for real-time follow-up observations and to constrain fraction of SGRs among short GRBs.

Swift: 10 Years of Discovery,
2-5 December 2014
La Sapienza University, Rome, Italy

∗Speaker.

†This publication was supported by the European social fund within the framework of realising the project „Support of inter-sectoral mobility and quality enhancement of research teams at Czech Technical University in Prague", CZ.1.07/2.3.00/30.0034. Period of the project's realisation 1.12.2012–30.6.2015.

1. Introduction

ESA INTEGRAL satellite was launched in 2002 [1]. The most sensitive instrument on-board is the IBIS imager. IBIS [2] uses a coded mask for the source localisation. The instrument is composed of 16384 CdTe detectors, located 3.4 m from a tungsten mask which projects a shadowgram onto the detector plane. Maps of the sky are reconstructed by decoding the shadowgram with the mask pattern. Since the upper tail of the operational energy region works in Compton scattering mode, the best sensitivity is achieved between 20 and 400 keV in photoelectric mode. The IBIS field of view is $19^\circ \times 19^\circ$ (at 50% coding).

2. Search

Short intense flares, mostly short gamma-ray bursts (sGRBs [3]) and soft gamma-ray repeaters (SGRs [4]), have been triggered through the on-board burst alert system, IBAS [5]. However, the triggering efficiency drops down on a very short time scale < 0.4 s. Therefore, a deep offline search for short transient events has been performed [6].

The INTEGRAL observations spanned > 64 Ms scattered in the period from Oct 2002 to Nov 2013 ($>$ Terabytes of data). For each observation block (Science Window, ScW) a "light curve" of counts ($20-200$ keV) from all pixels of the detector was composed and scanned for flares with 10σ significance above the actual background. Identified flares were confirmed using Bayesian block algorithm [7, 6]. Coincident solar flares and hot spot triggers caused most likely by cosmic rays were filtered out using data from the SPI ACS instrument [8] and the list of reported solar flares [9]. An example of such "light curve" is shown in Fig. 1.

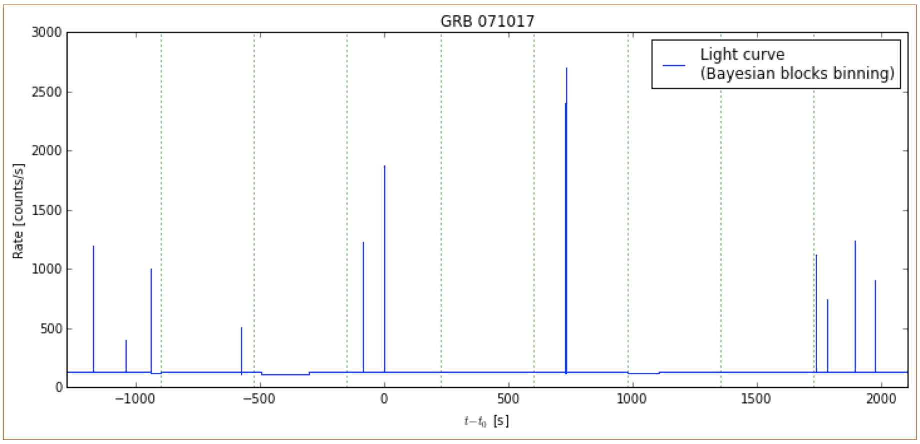

Figure 1: An example of a "light curve" constructed from all counts that hit the detector pixels (solar flares and cosmic rays filtered out). The result reveals additional flares around the short GRB 071017 (here located at $(t-t_0) = 0$ s) triggered and localised by INTEGRAL. The other flares in the ScW are of unidentified origin.

The search revealed $> 40'000$ events. Only 6% of those events were long enough (> 30 ms) to show structure and were suitable for further analysis. Each flare was characterised by its duration, peak flux, fluence, spectral lag, hard/soft ratio, symmetry (defined from the rise and decay time of a flare) and temporal variability. Galactic latitude of actual pointing and a total number of flares within the same ScW were also recorded. These properties were used as features in the further clustering analysis.

3. Short GRBs vs SGRs

The identity of the majority of the detected flares is unknown. After rejecting solar flares and cosmic rays, sGRBs and SGRs remain plausible explanations for the origin of at least some of those flares. The similarity between the time profiles of sGRBs and SGR giant flares lead to a several early misclassifications of SGRs as sGRBs. If the giant flare from the magnetar SGR 1806-20 from Dec 2004 happened in a distant galaxy, the periodic tail following the outburst would remain hidden below the detection threshold and the event would resemble a sGRB [10, 11]. An estimated $\gtrsim 4\%$ of sGRBs in the BATSE catalog could be distant SGRs [12].

4. Clustering

There is many SGR flares identified within the INTEGRAL observations, e.g. SGR 1550-5418 flaring in Jan 2009 [6]. On the contrary, the number of known short GRBs detected by INTEGRAL is low. To avoid a bias in classification, unsupervised machine learning was applied to the entire set of flares identified in the search.

The sample of the flares and their features were normalised to zero mean and unit variance for calculation purposes. Principal component analysis (PCA) of the feature space was applied to reduce the dimensionality of the problem [13], to evaluate which flare properties are important for the classification and to simplify visualisation. The technique of PCA transforms the feature vector space to a space of the same or lower dimension by finding such a linear combination defined by eigenvectors that provides the highest variance along the new axis. Consequently, high variance reflects high importance of a feature in machine learning classification (Fig. 2).

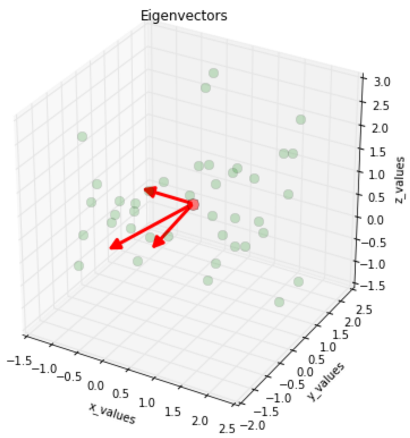

Figure 2: An illustration of the PCA transformation. The length of the eigenvectors (red) represents the variance along a new transformed axis of the feature space.

Then k-mean clustering [13] was used to discover classes in the reduced sample by finding similarities between the data points. k-mean clustering is simple yet powerful algorithm that groups nearby data points based on the metric in the feature space. 1) Initially, k centres of mass are chosen 2) Points become members of the group defined by the nearest centre 3) Centres of mass in each group are recalculated (Fig. 3). An iteration over steps 2) and 3) always converges to a solution. Since the found solution is not always the globally best solution, cluster centres are initiated 100 times randomly and the best result is accepted.

The k-mean clustering works the best for isotropic data, cluster of similar sizes and centroid shapes, while noise and irrelevant dimensions can disrupt performance.

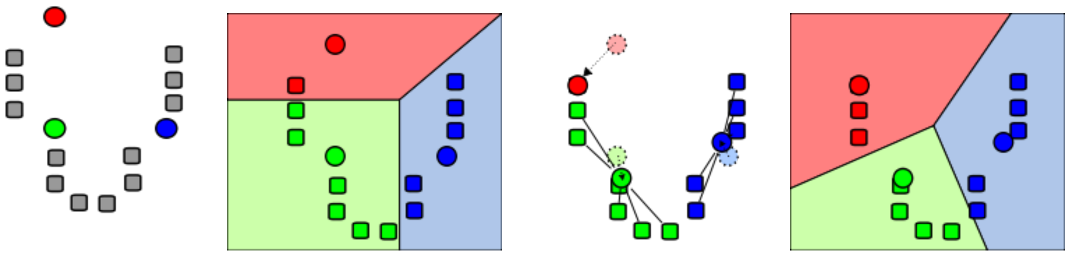

Figure 3: Illustration of the k-mean clustering algorithm for $k = 3$.

Sample splitting technique validated that the membership to the found cluster centres holds up to 86%. The identified clusters were a posterior interpreted by projecting data points representing sGRBs and SGR flares detected by INTEGRAL and known from literature.

5. Results

The distributions of individual flare attributes (features) do not show any significant grouping. The number of clusters $k = 5$ yielded the largest stability. Projecting of the flares of known origin revealed that the three INTEGRAL sGRBs seem to be clear outliers, while GRB 071017 is closer to the SGR group (Fig. 4)

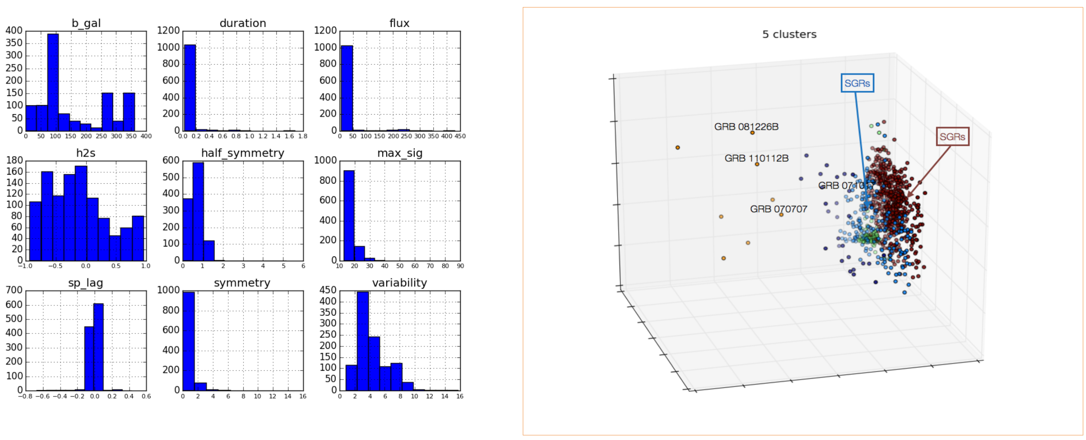

Figure 4: *Left:* The distributions of flare attributes used features for the machine learning clustering analysis. *Right:* 3D projection of the PCA transformed of flares represented as vectors ins the feature space. Data points belonging to the known sGRBs and SGR flares identified by INTEGRAL are labelled.

6. Conclusions

More than 40'000 untriggered flares with significance highly exceeding random fluctuation were found in the IBIS data. A posteriori visualisation of known sources identified separation between GRBs and SGRs. GRB 071017 is very likely a magnetar misclassified as a short gamma-ray burst [14]. This division could help in determining further observational strategy of an unknown flare and to put constraint on a fraction of magnetars among short GRBs. This simple but powerful technique could be applied without the need for localisation of the flaring event to any time series. Selected flares could then be inspected individually enhancing the offline search for untriggered short GRBs.

References

[1] C. Winkler et al., *The INTEGRAL mission*, *A&A*, **411**, L1–L6 (2003)

[2] F. Lebrun, J.P. Leray, P. Lavocat et al., *ISGRI: The INTEGRAL Soft Gamma-Ray Imager*, *A&A*, **411**, L141 (2003)

[3] E. Nakar, *Short-hard gamma-ray bursts*, *Physics Reports*, **442**, Issue 1-6:166-236 (2007)

[4] S. Mereghetti, *The strongest cosmic magnets: soft gamma-ray repeaters and anomalous X-ray pulsars*, *A&A Reviews*, **15**, 255-287 (2008)

[5] S. Mereghetti, D. Gotz, J. Borkowski, R. Walter and H. Pedersen, *The INTEGRAL Burst Alert System*, *A&A*, **411**, L291 (2003)

[6] M. Topinka, *Studies of Cosmic Gamma-Ray Bursts and a Search for Other Transient Events Detected by the INTEGRAL*, PhD thesis (2011)

[7] J. D. Scargle et al., *Studies in Astronomical Time Series Analysis. VI. Bayesian Block Representations*, *The ApJ*, **764**, Issue 2:167 (2013)

[8] A. von Kienlin, V. Beckmann, A. Rau et al., *INTEGRAL Spectrometer SPI's GRB detection capabilities. GRBs detected inside SPI's FoV and with the anticoincidence system ACS*, *A&A*, **411**, L299 (2003)

[9] R. P. Lin, B. R. Dennis, G. J. Hurford et al., *The Reuven Ramaty High-Energy Solar Spectroscopic Imager (RHESSI)*, *Sol. Phys.*, **210**, 3 (2002)

[10] Hurley E.P, et al., *An exceptionally bright flare from SGR 1806-20 and the origins of short-duration γ-ray bursts*, *Nature*, **434**, 1098–2005 (2005)

[11] K. Hurley, *Soft gamma repeaters*, *Advances in Space Research*, **47**, 1337 (2011)

[12] D. Lazzati, G. Ghirlanda, G. Ghisellini, *Soft gamma-ray repeater giant flares in the BATSE short gamma-ray burst catalogue: constraints from spectroscopy*, *MNRAS*, **362**, L8-L12 (2005)

[13] Ž. Ivezic, A. J. Connolly, J. T. VanderPlas and A. Gray, *Statistics, Data Mining and Machine Learning in Astronomy*, Princeton University Press (2014)

[14] S. Mereghetti, P. Esposito, A. Tiengo et al., *The magnetar candidate AX J1818.8-1559*, *A&A*, **546**, A30 (2012)