PROCEEDINGS OF SCIENCE

New results in applying the machine learning to GRB redshift estimation

Istvan I. Racz*
Eötvös University, RCAES HAS Konkoly Observatory
E-mail: racz@complex.elte.hu

Dezso Ribli
Eötvös University

Zsolt Bagoly
Eötvös University

Istvan Csabai
Eötvös University

Istvan Horvath
National University of Public Service

Lajos G. Balázs
Eötvös University, RCAES HAS Konkoly Observatory

Gamma-ray bursts (GRBs) are the most energetic transients in the far Universe. Several thousands of GRBs have been observed so far but we could measure the distance of only a few hundreds. We studied the parameters of GRBs with available spectroscopic redshift in order to be able to estimate the redshift of those GRBs without a measured one. To calculate their distances we applied two machine-learning estimator methods: random forest regressor and XGBoost. For the process we used selected gamma, x-ray and ultraviolet parameters from the Swift GRB catalog, which contains the measured spectroscopic redshift of 328 GRBs. We found a significantly higher correlation between the measured and estimated redshift, we have improved the correlation in multiple steps from 0.57 (published by Ukwatta et al., 2016) to 0.67. It seems that both the random forest and the XGBoost methods give similarly high correlation. For further improvements additional redshift measurements are required.

7th Fermi Symposium 2017
15-20 October 2017
Garmisch-Partenkirchen, Germany

*Speaker.

PoS(IFS2017)079

1. Introduction

Gamma-ray bursts (GRBs) are the most energetic explosions in distant galaxies. The duration of these bursts can range from tens of milliseconds to thousands of seconds [1]. However, the phenomenological classification scheme remains an open question since the 80s. It is generally accepted that there are short and long GRBs [2], the former is thought to be originating from binary neutron star mergers and the latter from hypernovas. Lately, with the help of multi-and uni-variate statistical analysis techniques an intermediate group was found [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13]. Furthermore, recent publications have discovered large structures based on GRBs: the Hercules-Borealis great wall at $z \approx 2$ [14, 15] and the Giant GRB ring at $z \approx 0.8$ [16, 17]. More GRBs with known redshift are necessary to understand the genesis of these structures. There are more attempts to estimate/determine the redshifts of GRBs without exact redshift measurement, such as [18], which describes a "machine-z" redshift prediction algorithm. Their method could show a 0.57 correlation coefficient between machine-z predictions and the real redshift measurements.

2. Data

The Swift robotic telescope is a multi-wavelength space observatory specialized to study GRBs and their afterglows with its three instruments on board working in gamma-ray (BAT), X-ray (XRT), ultraviolet, and optical (UVOT) wavebands. In this work we used data from all three instruments, but in contrast with [18] we used not only the Swift GRB Table from the NASA website but we merged this table with the Swift-XRT GRB Catalogue from the UK Swift Science Data Centre [19, 20, 21], so our new table contains all X-ray spectral fitting data, which are more precise compared to the data of the Swift GRB Table. Finally, we used similar parameters like [18].

3. Mathematical summary

XGBoost ("Extreme Gradient Boosting") is an advanced machine learning algorithm based on the decision tree method and uses "boosting" to improve a single weak model by combining it with a number of other weak models in order to generate a collectively strong model [22].

It is applied for supervised learning problems. To predict a target variable y_i we use an x_i training sample with multiple features, where usually the parameters follow the variance of the near normal, gamma or Poisson distribution.

Because there are relatively few GRBs with measured redshift, the training sample should contain the most objects possible, that's why we decided that the size of the training sample is 90% of the size of the original sample and we cross-validate with 10% of the objects. The method, with the help of a bootstrap algorithm, based on numerous – partially different – training samples, made the final decision tree.

4. Results

Fig. 1 shows how we improved the correlation on a perspective block diagram. We checked the input data, we decided that we skip the records where the parameters are uncertain. We defined some limits for the XRT first time observation and errors of data. We used the values if the errors were not too bad (errors $\leqq 100\%$). These limits were given by the examination of the parameter distribution. On the other side, we only used the X-ray and UVOT data if the XRT/UVOT first time observation was not too late ($T_{first} \leqq 150/200s$). Because these late measurements were conducted in a different setup, sometimes hours or days later, therefore the measured values represent different quantities than in all other data points. Additionally, we checked the redshift values too, and also skipped the photometric redshift data.

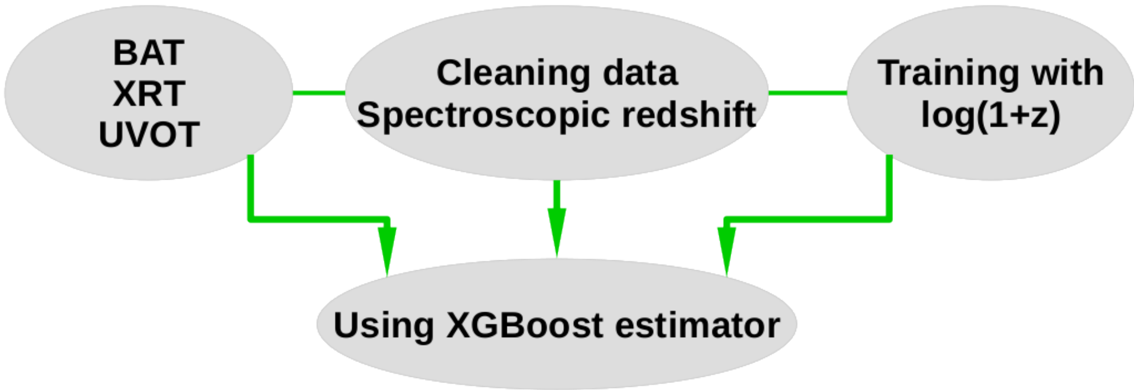

Figure 1: This block diagram shows the main steps for improving the goodness of redshift estimation. We checked the data and skipped the record which were not precise enough. Then we trained the method with $log_{10}(1+z)$ data because this was following normal distribution.

Examining the redshift distribution we found that on a logarithmic scale it can be estimated with a normal distribution. Taking the fact into consideration that the used algorithm gives the best results with parameters following normal distribution we decided that we train XGBoost with $log_{10}(1+z)$ samples.

After sufficiently cleaning the data –skipping the uncertain records (see above) – from all three Swift instruments, we used the $log_{10}(1+z)$ data as a training sample with which a 0.67 linear correlation (Fig. 2) was achieved between the estimated and measured redshifts. We calculated both (Spearman and Pearson) correlation coefficients and their significances and found similar correlation and high significances is both cases (sign. level: $< 10^{-16}$).

5. Summary

We examined the Swift BAT-XRT-UVOT data. Using the XGBoost estimator we could successfully improve the redshift estimations. The $log_{10}(1+z)$ correlations improved between the measured and calculated data from 0.57 to 0.67, the significance of this latter correlation – similarly to the previous case – strong (sign. level: $< 10^{-16}$). Our results show that with the increase in the number of GRBs with measured redshift the redshift estimation can improve further. We plan to take the errors of the parameters into consideration in our analysis, also to test other machine learning algorithms for this task.

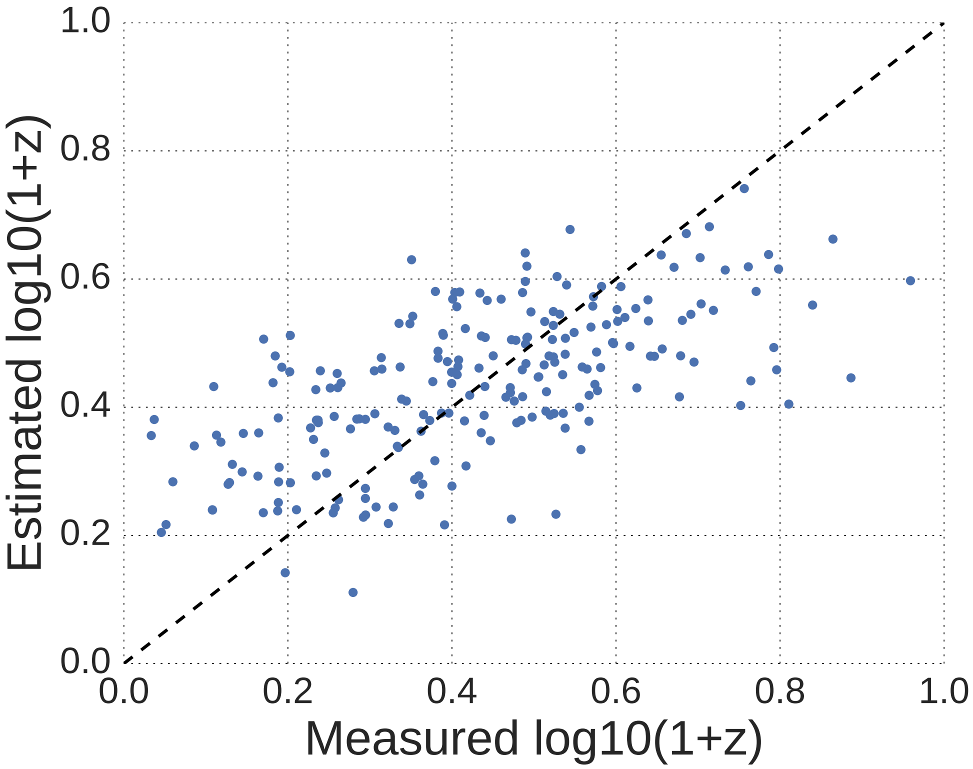

Figure 2: We used the data of all three Swift instruments and sufficiently cleaned them. Using the $log_{10}(1+z)$ data we could establish a 0.67 linear correlation (both Pearson and Spearman correlation) between the estimated and measured redshifts without the errors of measured redshift (for now, the errors were not taken into account). The figure shows this correlation between the logarithms of the estimated and the measured redshifts.

Acknowledgements This work was supported by the Hungarian OTKA NN-111016 grant and the ÚNKP-17-3 New National Excellence Program of the Ministry of Human Capacities. This work made use of data supplied by the UK Swift Science Data Centre at the University of Leicester.

References

[1] P. Kumar and B. Zhang, *The physics of gamma-ray bursts and relativistic jets*, *Phys. Rep.* **561** (Feb., 2015) 1–109, [1410.0679].

[2] J. P. Norris, T. L. Cline, U. D. Desai and B. J. Teegarden, *Frequency of fast, narrow gamma-ray bursts*, *Nature* **308** (Mar., 1984) 434.

[3] I. Horváth, *A Third Class of Gamma-Ray Bursts?*, *ApJ* **508** (Dec., 1998) 757–759, [astro-ph/9803077].

[4] S. Mukherjee, E. D. Feigelson, G. Jogesh Babu, F. Murtagh, C. Fraley and A. Raftery, *Three Types of Gamma-Ray Bursts*, *ApJ* **508** (Nov., 1998) 314–327, [astro-ph/9802085].

[5] I. Horváth, *A further study of the BATSE Gamma-Ray Burst duration distribution*, *A&A* **392** (Sept., 2002) 791–793, [astro-ph/0205004].

[6] I. Horváth, L. G. Balázs, Z. Bagoly, F. Ryde and A. Mészáros, *A new definition of the intermediate group of gamma-ray bursts*, *A&A* **447** (Feb., 2006) 23–30, [astro-ph/0509909].

[7] I. Horváth, L. G. Balázs, Z. Bagoly and P. Veres, *Classification of Swift's gamma-ray bursts*, *A&A* **489** (Oct., 2008) L1–L4, [0808.1067].

[8] I. Horváth, *Classification of BeppoSAX's gamma-ray bursts*, *Ap&SS* **323** (Sept., 2009) 83–86, [0905.0860].

[9] P. Veres, Z. Bagoly, I. Horváth, A. Mészáros and L. G. Balázs, *A Distinct Peak-flux Distribution of the Third Class of Gamma-ray Bursts: A Possible Signature of X-ray Flashes?*, *ApJ* **725** (Dec., 2010) 1955–1964, [1010.2087].

[10] A. de Ugarte Postigo, I. Horváth, P. Veres, Z. Bagoly, D. A. Kann, C. C. Thöne et al., *Searching for differences in Swift's intermediate GRBs*, *A&A* **525** (Jan., 2011) A109, [1006.4469].

[11] M. Tarnopolski, *Analysis of Fermi gamma-ray burst duration distribution*, *A&A* **581** (Sept., 2015) A29, [1506.07324].

[12] I. Horváth and B. G. Tóth, *The duration distribution of Swift Gamma-Ray Bursts*, *Ap&SS* **361** (May, 2016) 155, [1604.00887].

[13] M. Tarnopolski, *Analysis of gamma-ray burst duration distribution using mixtures of skewed distributions*, *MNRAS* **458** (May, 2016) 2024–2031, [1506.07801].

[14] I. Horváth, J. Hakkila and Z. Bagoly, *Possible structure in the GRB sky distribution at redshift two*, *A&A* **561** (Jan., 2014) L12, [1401.0533].

[15] I. Horváth, Z. Bagoly, J. Hakkila and L. V. Tóth, *New data support the existence of the Hercules-Corona Borealis Great Wall*, *A&A* **584** (Dec., 2015) A48, [1510.01933].

[16] L. G. Balázs, L. Rejtő and G. Tusnády, *Some statistical remarks on the giant GRB ring*, *MNRAS* **473** (Jan., 2018) 3169–3179, [1710.01621].

[17] L. G. Balázs, Z. Bagoly, J. E. Hakkila, I. Horváth, J. Kóbori, I. I. Rácz et al., *A giant ring-like structure at* $0.78 < z < 0.86$ *displayed by GRBs*, *MNRAS* **452** (Sept., 2015) 2236–2246, [1507.00675].

[18] T. N. Ukwatta, P. R. Woźniak and N. Gehrels, *Machine-z: rapid machine-learned redshift indicator for Swift gamma-ray bursts*, *MNRAS* **458** (June, 2016) 3821–3829, [1512.07671].

[19] P. A. Evans, A. P. Beardmore, K. L. Page, L. G. Tyler, J. P. Osborne, M. R. Goad et al., *An online repository of Swift/XRT light curves of* γ*-ray bursts*, *A&A* **469** (July, 2007) 379–385, [0704.0128].

[20] M. R. Goad, L. G. Tyler, A. P. Beardmore, P. A. Evans, S. R. Rosen, J. P. Osborne et al., *Accurate early positions for Swift GRBs: enhancing X-ray positions with UVOT astrometry*, *A&A* **476** (Dec., 2007) 1401–1409, [0708.0986].

[21] P. A. Evans, A. P. Beardmore, K. L. Page, J. P. Osborne, P. T. O'Brien, R. Willingale et al., *Methods and results of an automatic analysis of a complete sample of Swift-XRT observations of GRBs*, *MNRAS* **397** (Aug., 2009) 1177–1201, [0812.3662].

[22] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, *ArXiv e-prints* (Mar., 2016) , [1603.02754].