R in the Midst of Exploding Stars:
Distributed, Time-Domain Transient Classification

Thomas J. Fuchs[1,2,*]**, Michael Turmon**[1]**, Matthew J. Graham**[2]**, Ciro Donalek**[2]**, Ashish Mahabal**[2]
Andrew J. Drake[2]**, S. George Djorgovski**[2]

1. Jet Propulsion Laboratory, California Institute of Technology
2. California Institute of Technology
[*]Contact author: thomas.fuchs@jpl.nasa.gov

Keywords: Astronomy, Machine Learning, Transients, Time Series, Feature Extraction

Modern synoptic sky surveys are enabling novel research frontiers in time domain astronomy and posing new machine learning challenges for early detection and classification Donalek et al. [2]. We present a novel framework for time domain astronomy using *R*. Our system incorporates machine learning algorithms for an iterative, dynamical classification of astronomical transient events, based on the initial detection measurements, archival information, and newly obtained follow-up measurements from robotic telescopes. Specifically we use *R*for (i) preprocessing, (ii) to post observed light curves to the Caltech Time Series Characterization Service for feature extraction and (iii) for classification.

Astronomical time series tend to be heterogeneous. They can vary widely in their temporal coverage, sampling rates and regularity, number of points and error bars, even from the same survey, and hence need to be homogenized prior to analysis. For a large number of learning algorithms each time series needs to be recast in terms of a set of features – individual measurable heuristic properties of an object that can be used to characterize it. The Caltech Time Series Characterization Service (CTSCS: nirgun.caltech.edu:8000) aims to provide easy access to different time series characterization statistics used in the literature. These descriptors are then used to train and test a hierarchical model based on random forest classifiers. We compare various modeling choices and analyze the contribution of CTSCS features to the classification accuracy.
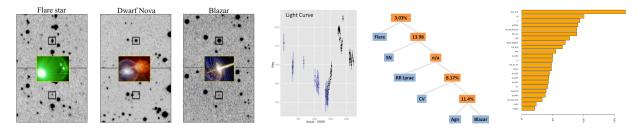
Figure 1: **Left:** Examples of transient events from the Catalina Real-time Transient Survey (CRTS: crts.caltech.edu) Djorgovski et al. [1]. Images in the top row show objects which appear much brighter that night, relative to the baseline images obtained earlier (bottom row). On this basis alone, the three transients are physically indistinguishable, yet the subsequent follow-up shows them to be three vastly different types of phenomena: a flare star (left), a cataclysmic variable powered by an accretion to a compact stellar remnant (middle), and a blazar, flaring due to instabilities in a relativistic jet (right). Accurate transient event classification is the key to their follow-up and physical understanding. **Center left:** A single light curve plotted with **ggplot**. **Center right:** Classification error of a hierarchy of random forest models. **Right:** Random forest based importance plot of CTSCS features.

References

[1] Djorgovski, S. G., A. Mahabal, C. Donalek, M. J. Graham, A. J. Drake, B. Moghaddam, and M. Turmon (2012). Flashes in a star stream: Automated classification of astronomical transient events. In *E-Science (e-Science), 2012 IEEE 8th International Conference on*, pp. 1–8. IEEE.

[2] Donalek, C., A. Kumar, S. G. Djorgovski, A. Mahabal, M. J. Graham, T. J. Fuchs, M. J. Turmon, N. S. Philip, M. T.-C. Yang, and G. Longo (2013). Feature selection strategies for classifying high dimensional astronomical data sets. In *BigData Conference*, pp. 35–41.