Twenty Feature Selection Algorithms, One Dataset, One Problem: Flare Forecasting

A. Yeolekar, S. Patel, S. Talla, K. R. Puthucode, K. R. Puthucode, A. Ahmadzadeh, V. M. Sadykov,
and *R. A. Angryk*
Department of Computer Science
Georgia State University

Solar Energy Particles (SEPs) can be associated with solar flares and coronal mass ejections (CMEs) and offer energy spectra ranging from few KeVs to many GeVs. These events can occur without any notable indication and alter the radiation environment of the inner solar systems, which can potentially lead to precarious conditions for humans in space, affect the interior of spacecraft's sensitive electronics, and trigger radio blackouts. Identifying the most critical physical parameters of the Solar Dynamic Observatory (SDO) to detect SEPs can allow for a swift response against its adverse effects.

With the profusion of high-quality time series data from the SDO, which accounts for the modulating background of magnetic activity and the inherently dynamic phenomenon of pre-flares and post-flare phases; antithetical to non-representative data with the point-in-time measurements employed earlier, selection of vital parameters for solar flare classification using machine learning algorithms appears to be a well-fitted problem in this realm. The primary issue of dealing with multivariate time series data (mvts) is the large number of physical parameters operating at a rapid frequency, making the data dimensionality very high and thus causing the learning process to curb. Moreover, manually selecting vital parameters is a tedious and costly task on which experts may not always agree on the results. In response, we examined feature subset selection using multiple algorithms on both mvts data and the statistical features derived from mvts segments (vectorized data). We used the SWAN-SF (Space Weather Analytics for Solar Flares) benchmark dataset collected from May 2010 - September 2018 to conduct our experiments. The comprehensive study gives a stable scheme to recognize the critical physical parameters, which boosts the learning process and can be used as a blueprint to foretell future solar flare episodes.