Avoiding More Accurate and Less Robust Models: Mistakes and Pitfalls in Training Flare Forecasting Models

Authors

Azim Ahmadzadeh, Berkay Aydin, Dustin Kempton, Sushant Mahajan, Manolis Georgoulis, and Rafal Angryk

Abstract

In spite of more than 20 years of substantial advances, solar flare prediction remains a largely outstanding problem. This is partly because of the scarcity of major flares. Effective flare prediction, if ever achieved, would help mitigate a substantial projected economic damage, with a long-range magnitude of 1 to 2 trillion dollars for the US alone. Prediction could also help mitigate, or even prevent, serious health risks to astronauts exposed to flares' electromagnetic radiation and particulate. While many recent flare prediction studies have opted to employ Machine Learning techniques to better tackle the problem, a lack of sufficient understanding of how to properly treat the data often leads to overly optimistic results. We use the recently generated GSU solar flare benchmark dataset, called Space Weather ANalytics for Solar Flares (SWAN-SF), to show how a 'mediocre' forecast model can turn into an 'impressive' one, by simply overlooking some basic practices in data mining and machine learning. The benchmark is a multivariate time series collection, extracted from magnetographic measurements in the solar photosphere and spans over eight years of the Solar Dynamics Observatory Helioseismic and Magnetic Imager (SDO/HMI) era. We briefly explain the data collection process, the sampling and the slicing of time series, and then outline a series of experiments using machine learning models to illustrate the common mistakes, fallacies and pitfalls in forecasting rare events. We particularly elaborate on how and why imbalanced datasets, in general, impact the models' performance, and how different under- or over-sampling methodologies and weighting practices could introduce accurate but often weak models. Concluding, we aim to draw attention to the impact of these practices on the flare forecasting models and how to train models by accentuating the statistical robustness over a relative accuracy in prediction.