Challenges of Using Machine Learning for Solar Flare Prediction and Prospective Solutions

Wendy Carande[1], James Craft[1], Chris Pankratz[1], Tom Berger[2], Justin Cai[1], Maxine Hartnett[1], William Newman[1]

Laboratory for Atmospheric and Space Physics
University of Colorado **Boulder**

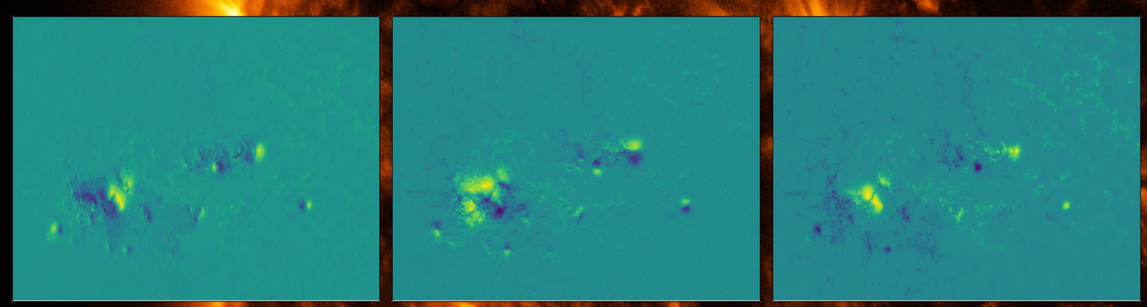
1 Laboratory for Atmospheric and Space Physics, Boulder, CO
2 SWx-TREC, University of Colorado, Boulder

Introduction

The question of how to use machine learning to predict solar flares has been explored extensively in recent years. Although the field has advanced dramatically, there are many challenges we still face in using machine learning for solar flare prediction. Here, we discuss some of these challenges and our proposed solutions to these challenges.

Methods

We use magnetogram data from the Solar Dynamics Observatory Helioseismic and Magnetic Imager (HMI) to train two types of deep learning models: convolutional neural networks (CNNs) and multilayer perceptron (MLP). This results in a binary classification indicating whether or not an M or X type solar flare will occur in the 24 hour time window.



Sample magnetogram: westward (left), southward (middle), and radial (right) components.

Data Imbalance

Solar flare data are inherently imbalanced; the vast majority of recorded events are "negative" (no flare), and the frequency of different solar flare types varies. Many machine learning classification algorithms work optimally on balanced data sets. Additionally, this imbalance needs to be considered when selecting a scoring metric. To address this imbalance, we use both under- and over-sampling techniques.

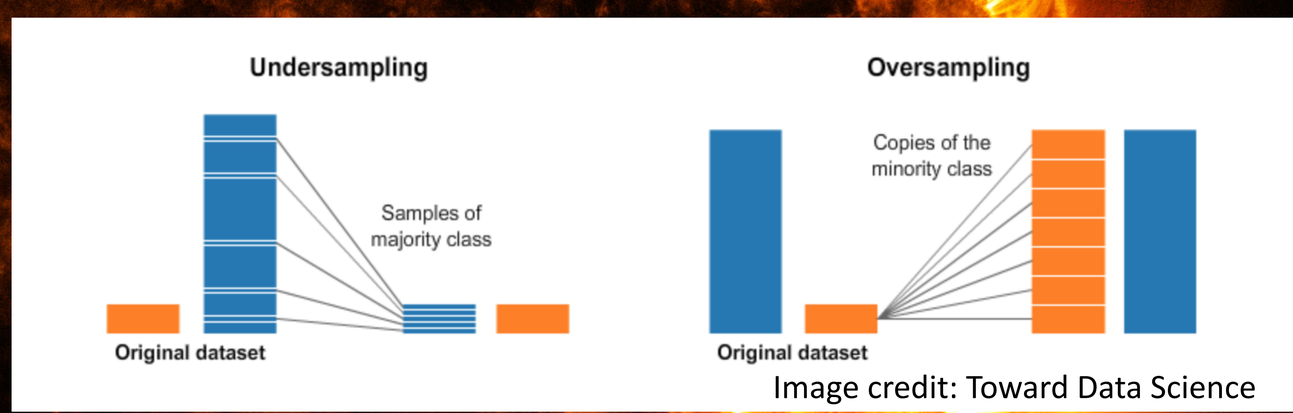


Image credit: Toward Data Science

Training Set Selection

To assess the performance of our deep learning methods, we split the data into training, testing, and validation sets. Although random sub-setting of data is desirable in some machine learning applications, for solar flare time series data this is not an ideal approach and can lead to an artificially high score. We recommend splitting data based on time rather than randomly.

Scoring Metrics

With the development of benchmark data sets to evaluate solar flare prediction methods, the question arises: What metric will be used to score and compare these methods? Although True Skill Score has become the metric of choice for solar flare prediction, we find that there are more descriptive metrics available, especially those derived from meteorological studies. Although our models score lower using these metrics, they are more indicative of what we truly care about when predicting solar flares, which is predicting positive events correctly.

CSI - Critical Score Index

$CSI = TP / (TP + FN + FP) \in [0, 1]$

- Measures the accuracy of positive predictions.
- Used for rare occurrences where we do not care about true negative predictions as much as true positive predictions.
- Does not account for true positives by random chance.

GSS - Gilbert Skill Score

$C = (TP + FP) * (TP + FN) / (TP + FP + TN + FN)$

$GSS = (TP - C) / ((TP - C) + FN + FP) \in [0, 1]$

- Similar to CSI, but corrects for random chance.
- Assume there is a random chance of guessing a positive event correctly (C).
- Subtract C from the True Positives to account for random chance.