Using Machine Learning Techniques to Forecast Solar Energetic Particles

Pedro Brea

Abstract
Solar energetic particles (SEPs) endanger satellites, disrupt air traffic over the polar regions and high frequency radio communication. The capacity to predict SEP events on Earth ahead of time can lead to the preservation of important space and aeronautical assets as well as protection for humans in space. Considering that the methods of acceleration and transport of these particles is still an area of active research and that physics-based models are, at the moment, slower than empirical models, forecasters at the National Oceanic and Atmospheric Administration (NOAA) Space Weather Prediction Center (SWPC) use the latter, combined with forecaster heuristics, to make real time decisions. This project attempts to improve upon the results of the current statistical model in use at SWPC (Proton Prediction Model) by using machine learning classification techniques. Machine learning models learn and make decisions using an observational training set and are currently much quicker than numerical models for issuing a forecast. SEP forecasts are made based on physical parameters associated with solar flares and coronal mass ejections. Preliminary results show that the logistic regression, Adaboost decision tree, and support vector machine algorithms show an improvement in forecasting skill over the current SWPC Proton Prediction Model.
Introduction

The NOAA Space Weather Prediction Center relies on real-time observations of the space environment to forecast solar energetic particle events. SWPC defines an SEP event as when the 10 MeV proton flux at the GOES (Geostationary Operational Environmental Satellite) satellite in geostationary orbit exceeds 10 particles per \( \text{cm}^2 \cdot \text{s} \cdot \text{sr} \cdot \text{m} \). The purpose of the GOES satellite is to continually monitor the meteorological and space environment to protect life and property in the United States (GOES-R Series Concept of Operations (CONOPS)). SEPs can damage electronics on satellites causing them to malfunction, disrupt communication technologies on Earth, and they pose a radiation hazard to astronauts and flight crew and passengers flying over the poles. These high-energy events are associated with solar flares and coronal mass ejections (Schrijver & Siscoe, 2011) and accurately predicting these events allows SWPC’s customers (including NASA, SpaceX, electrical power grid companies, the International Civil Aviation Organization (ICAO), and even pigeon racing associations, among others) to prepare and take the necessary actions to mitigate the impact of these solar radiation storms.

Since the method of acceleration and transport of these particles is not entirely understood and considering that SEPs can reach Earth as soon as 40 minutes, numerical physics-based models cannot yet provide quick, accurate forecasts of SEP events. Empirical proton forecasting models run much quicker and are more adequate for making SEP predictions at the moment. SEPs are strongly correlated with coronal mass ejections (CMEs) which eject significant amounts of magnetized plasma into space. However, forecasters do not have images of CMEs from the LASCO coronagraph in real time, so proxy data of CMEs must be used to predict SEP events. Currently, forecasters rely on the Proton Prediction Model that is outlined in Balch (1999) and Balch (2008), which relies on observational data that is readily available in the forecasting office in real time. These include flare duration, X-ray flux, radio bursts, and other data that can be obtained in real time. SWPC’s reliance on real-time observations, and swift statistical modeling presents a prime opportunity for the implementation of machine learning models in SEP forecasting. These models can be trained to recognize relationships between inputs (real-time observational data in this case) and the occurrence of events, which in this project is identified by the presence or lack of SEPs hitting Earth in response to a solar flare occurring on the Sun.

Recently, NASA announced its Moon to Mars mission, in which the agency plans to establish a long-term human residence on the Moon beginning in 2022 to make new scientific discoveries and start a lunar economy with the intent of eventually pushing the boundaries of human exploration to Mars in the 2030s (NASA Lunar Gateway). However, this ambitious venture will be extremely dangerous because astronauts will no longer have the protection of Earth's magnetosphere once they venture out into interplanetary space. One risk is exposure to radiation and high-energy solar events. In order to protect the lives of the astronauts and the integrity of
the hardware necessary for the trips, it is essential for NASA to receive forecasts of SEP events, which they will be receiving from SWPC. Methods such as those outlined in this paper could make a positive contribution to this pioneering endeavor.

Figure 1: Coronagraph images of a solar flare from the SOHO LASCO instrument overlaid by the SDO AIA 304 nanometer coronagraph on April 21, 2001.

Figure 2: Coronagraph images of a coronal mass ejection from the SOHO LASCO instrument overlaid by the SDO AIA 304 nanometer coronagraph on September 10, 2017.
Methods

TRAINING

Supervised machine learning models are trained to map a given input to an output and they are ‘supervised’ in the sense that the user trains the algorithm with a training set where the inputs and outputs are known by the user. The models of interest within this project were initially trained on a catalog of flares from 1986 to 2004 (Balch, 2008) containing data associated with flares recorded by the GOES satellite. The machine learning models were given this information to draw relationships between these various physical parameters and how they interact to produce SEPs. Theoretically, a model that recognizes these relationships should accurately be able to predict SEP events, the accuracy being limited to the extent that the model parameters are indicative of SEPs.

The models were trained on 80% of the original dataset (the training set) and tested on the remaining 20% of the original dataset. In this way the model can be trained and tested on distinct sets of data which ensures that the model is evaluated against an unseen dataset. One challenge in regards to training the model was the class imbalance between positive and negative classes in the dataset. Only about 3% of all the events in the dataset were associated with an SEP event. This imbalance could lead to models that are not trained well enough to properly recognize the intricacies associated with SEP events. To increase the number of positive events, 40 flares associated with SEPs were added to the database. Every C-class flare or higher from 2005-2017 (over 9000 flares) is being added to the database of events at the time of this writing but they were not used for the results outlined in this paper.

Model performance was evaluated using the Heidke Skill Score (HSS)

\[
HSS = \frac{2([TP \times FN] - [FP \times FN])}{(TP + FN)(FN + TN) - (TP + FP) (TP + TN)}
\]

where TP stands for true positives (number of events correctly classified as having SEPs), TN stands for true negative (number of events correctly classified as not being associated with SEPs), FP stands for false positive (number of events incorrectly classified as having SEPs, i.e., a false alarm), and FN stands for false negative (number of events incorrectly classified as not being associated with SEPs, i.e., a missed event)(www.eumetrain.org). This was the scoring metric used in Balch 2008 which allows a direct comparison between the operational SWPC model and the machine learning algorithms presented herein.
FEATURES

The original features used to develop the Balch model were the X-ray peak flux, integrated X-ray flux, and the presence of a type II or type IV radio burst. The machine learning models were initially given these same parameters for an “apples to apples” comparison of the two models. Solar flares emit highly energetic electromagnetic radiation in the form of X-rays, increasing the X-ray flux in the heliosphere. For our purposes, the X-ray peak flux is identified by the maximum flux observed by the GOES satellite 1.0 - 8.0 Angstroms channel during the event. SEPs tend to be associated with large flares (flares with large peak flux values). This makes the X-ray peak flux a strong indicator of an event capable of accelerating particles to SEP velocities. The flare integrated flux is defined by the integration of the X-ray flux between the flare start time the point of half maximum power after the flare peak. The integrated flux is correlated with the occurrence of CMEs. Long duration flares with a greater integrated flux are indicative of a CME and are a good proxy in lieu of CME images, and since CMEs are correlated with SEPs it is expected that the integrated flux is a predictive feature. Type II and type IV radio bursts are radio signatures of electrons being accelerated in CME shocks. Radio signatures provide information regarding ambient density, magnetic field, and turbulence in the heliosphere (Gopalswamy, 2016).

Figure 3: Example of the GOES X-ray lightcurve for September 2017 indicating the flare peak flux.
Furthermore, since the Earth is magnetically connected to the western limb of the Sun, after the initial comparison the models were given east-west flare location on the solar disk to test whether the location of a CME or solar flare on the solar disk improves forecasting accuracy. Any large event occurring on the western limb of the Sun is likely to hit Earth but this is not the case for the eastern limb. Therefore, this information should also be predictive of SEPs hitting Earth. The flare X-ray temperature and emission measure were calculated using the GOES soft X-ray(1-8 Angstroms) data (Garcia, 1994). The newest features introduced into the model were CME parameters such as CME speed and width. CME speed is the propagation speed (km/s) of a CME through interplanetary space and it is measured by the LASCO coronagraph along with the angular width of the CME measured by the coronagraph. Although forecasters are not presently able to obtain CME speed and width in real time, the NOAA SWFO-L1 mission will launch in 2024, which will allow forecasters to look at this data in real time. Therefore, since it is believed that CME speed and width are related to the acceleration of energetic particles, this information was used in the most recent model runs in anticipation of the forecasting capabilities forecasters will have in 2024. Another feature that was added to the catalog is flare persistence. Flare persistence is the peak GOES X-ray flux six hours before the beginning of a flare. This parameter is being used to test whether different levels of X-ray flux in the space environment before a flare affects the capability of flares and CMEs to accelerate charged particles to the threshold of solar energetic particles. A flare occurring six hours before another could leave a seed population of highly energetic particles in the interplanetary environment making an SEP event more likely if another flare were to occur soon after.

**LOGISTIC REGRESSION**

A logistic regression model is a linear model that is used for classification rather than regression (Pedregosa et. al, 2011). The model assigns weights to the given features using methods such as gradient descent. The algorithm then calculates the probability of each event having SEPs based on the weights the model assigns to each feature and the corresponding feature values. This is done by utilizing the sigmoid function

\[
sig(z) = \left(1 + e^{-z}\right)^{-1}
\]

where,

\[
z = \overline{w} \cdot \overline{x}
\]

is equal to the inner product of the feature weights (\(\overline{w}\)) and their corresponding features (\(\overline{x}\)). The sigmoid function returns values between 0 and 1, where its output is the probability of an SEP event occurring. If the sigmoid function's output was greater than or equal to the threshold value (0.5), then the model would predict SEPs to be associated with that event. If the output was less
than 0.5 than the model predicts that SEPs will not be associated with that event. The model
decision threshold can be adjusted, as well as the relative weights of the classes (SEP and
non-SEP events) and the regularization tuning factor, among others. Regularization is an
important hyperparameter that controls the model's flexibility by preventing feature weights from
becoming unrealistically large which makes it useful to avoid overfitting (towardsdatascience.com, 2017). Overfitting occurs when a model is really well suited to a
training set but is not well generalized to unseen data.

![Figure 4: Example of the sigmoid that the logistic regression function outputs where the x-axis represent the z values and the y-axis is how the output of the sigmoid varies as a function of z. The red line represents the decision threshold of 0.5.](image)

**ADABOOST**

Decision trees attempt to make classifications by making splits within the feature data that
optimize information gain, e.g. splitting to give the purest classifications to give the fewest
misclassifications. Specifically, boosted decision trees were utilized in an attempt to reduce
overfitting and obtain the best results. The boosting algorithm used in the model, AdaBoost, uses
an ensemble of decision trees with a single split, known as "weak learners", to make
classifications. Data points are classified "democratically" in the sense that a label is attached
according to the cumulative opinion of all voting members. The AdaBoost algorithm attempts to
minimize the loss function

\[
\frac{1}{m} \sum_{i=1}^{m} \exp(y_i F(x_i))
\]
where

\[ F(x) = \sum_{i=1}^{T} \alpha_i h_i(x), \]

being that \( y \) is the class label (-1,+1), \( \alpha \) equals the assigned feature weight, \( h \) is the predicted class label the feature corresponds to (-1,+1) based on the split made by a particular weak classifier, \( t \) and \( i \) are summing indices, \( m \) and \( T \) are the total number of events, and \( x \) is a feature value (Schapire, 2013). A classification is then made based on the sign of \( F(x) \).

Hyperparameters such as the number of weak estimators utilized and the learning rate, which determines how quickly a model converges to a solution, were tuned using the GridSearchCV algorithm.

---

**SUPPORT VECTOR MACHINES (SVMs)**

Support vector machines are supervised machine learning models that determine an optimum line or hyperplane in a multidimensional feature space to classify data where events falling on one side of the hyperplane are classified positive on those on the other are classified negative. SVMs are flexible in the sense that they can be tuned to draw a complex hyperplane to categorize data instead of simply taking a linear cut through the data to separate one class from another. In the training phase of the model, SVMs were used to find the optimal hyperplane to correctly distinguish between flares that were associated with solar energetic particles and those that weren't, in hopes that the model would be able to predict the occurrence of an SEP event given a new set of data from a new flare. To get a reasonable idea of the probability associated

---

Figure 5: Visualization of how Adaboost decision trees create an ensemble of weak learners to create a classification model.
with an event classified by an SVM, the orthogonal vector distance between an event and the hyperplane in feature space can be used as a proxy for the probability of occurrence.

Figures 6&7: How SVMs use kernels to transform datasets to a higher dimension to draw a linear cut which corresponds to a hyperplane in the original data feature space. Images are credited to Chris Kettelsen, University of Colorado, Boulder.

**Results**

Results were acquired in three runs, each of which with a different set of features. In the first run only the original Balch features were utilized to train the model. In the second all of the features excluding CME speed and width were used. Finally, in the third and final run all of the features were utilized. The scores for the runs are listed in the tables below. It is important to note that the SWPC Proton Prediction Model has an HSS of 0.47.

**TEST SET 1 - BALCH FEATURES**

<table>
<thead>
<tr>
<th></th>
<th>Logistic Regression (HSS)</th>
<th>Adaboost Decision Trees (HSS)</th>
<th>Support Vector Machine (HSS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Score</td>
<td>0.3622</td>
<td>Training Score: 0.7754</td>
<td>Training Score: 0.4824</td>
</tr>
<tr>
<td>Test Score</td>
<td>0.4192</td>
<td>Test Score: 0.3623</td>
<td>Test Score: 0.4566</td>
</tr>
</tbody>
</table>
**TEST SET 2 - ALL FEATURES MINUS CME SPEED AND CME WIDTH**

<table>
<thead>
<tr>
<th>Model</th>
<th>Training Score</th>
<th>Test Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression</td>
<td>0.4162</td>
<td>0.4464</td>
</tr>
<tr>
<td>Adaboost Decision Trees</td>
<td>1.0000</td>
<td>0.2366</td>
</tr>
<tr>
<td>Support Vector Machine</td>
<td>0.4605</td>
<td>0.4840</td>
</tr>
</tbody>
</table>

**TEST SET 3 - ALL FEATURES**

<table>
<thead>
<tr>
<th>Model</th>
<th>Training Score</th>
<th>Test Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression</td>
<td>0.4914</td>
<td>0.5149</td>
</tr>
<tr>
<td>Adaboost Decision Trees</td>
<td>0.7295</td>
<td>0.1840</td>
</tr>
<tr>
<td>Support Vector Machine</td>
<td>0.5159</td>
<td>0.5313</td>
</tr>
</tbody>
</table>

Although in the first run with the Balch features all three models score below 0.47, in the subsequent tests there were three runs that score over a 0.47. In the second one with all the features minus CME speed and width the SVM model slightly edges the Balch model. In the third run with all the features the logistic regression and SVM all score over 0.5 showing visible improvement upon addition of the CME features. Although Adaboost consistently scores the highest in training, it significantly underperformed on the test sets. This large discrepancy between training and test score is a sign that the Adaboost model is overfitting to the training data and that the hyperparameters of the model need to be adjusted.

**Discussion**

The results are promising considering that three of the runs were able to score higher than the Balch model. The second run, where the SVM model scored 0.4840, contained features that can all be obtained in real-time by forecasters. This shows that the models, even in their early stages of development, are able to outperform SWPC’s operational model.

There was a significant rise in scores for both the logistic regression and SVM models upon addition of the CME features. The logistic regression model improved from an HSS score of 0.4464 in the second run to a score of 0.5149 in the third run while the SVM model improved from a score of 0.4840 to 0.5313. The Adaboost model performed worse upon addition of the CME features but this is likely because, similarly to how it was overfitting in the first two runs, it is overfitting to an even more complex dataset than the first two runs leading to an even less generalized model. However it has been seen in the past that the Adaboost decision trees have
the capability to edge the Balch model’s HSS score. The performance of the models depend on the tuning of the hyperparameters of each model and while the logistic regression and SVM models seem to have been well tuned, other methods of tuning Adaboost should be considered to prevent overfitting.

**Conclusion**

It seems that the implementation of these machine learning models along with the addition of the new aforementioned features could increase forecasters’ skill in predicting SEPs. The hyperparameters of the models should continue to be investigated to optimize model performance and avoid overfitting (especially in the case of the Adaboost decision trees). The models have already shown potential outperform the Proton Prediction Model with data that is available in real-time to forecasters, and with the launching of the SWFO-L1 in 2024 forecasters will be able to obtain features such as CME speed and width in real time and the results show that adding these features to the dataset significantly improve SEP forecasting skill. Overall, according to the results outlined by the test runs of this project, it seems like the logistic regression and SVM models have the potential to be used operationally by SWPC.

**References**


https://towardsdatascience.com/regularization-in-machine-learning-76441ddcf99a

Heidke Skill Score (HSS). *Physical Background.*

Loff, S. (2018), Lunar Gateway, NASA. Available from:
https://www.nasa.gov/topics/moon-to-mars/lunar-gateway


US Department of Commerce (DOC), National Oceanic and Atmospheric Administration (NOAA), NOAA Satellite and Observation Services (NESDIS), and National Aeronautics and Space Administration (NASA) (2016): GOES-R Series Concept of Operations (CONOPS),