SOLAR FLARE PREDICTION USING PYTHON

Muhammad Usman

USED LIBRARIES

NUMPY, PANDAS

Code can be found at https://github.com/musman2012/ml-data_science

Introduction

The solar system we live in is one of the most complex and tightly bounded system known to the human knowledge. It follows certain laws with perfection and consists of some interesting events which help the entire system lively. Sun, being the center of solar system, is source of various events happening in the system and the space. These events and activities commonly referred as Space Weather.

Two of the main phenomena of space weather are Sunspots and Solar flares. Both these events are connected with the extreme heat being produced from the core of sun and with the magnetic field flux. Sunspots are the darker regions at the surface of sun having more magnetic activity and less energy consumption as compared to the other surrounding areas (Solanki, 2003). The core of sun being a continuous source of heat energy consistently produce heat towards the surface and the strong magnetic fields around the sunspot region stops this heat to reach at the surface making that region darker than the neighboring areas (Erickson, 2019). Once the sunspot are made, some of them, which are referred as 'complex' sunspot groups by Zirin and Wang, can originated one of multiple solar flares as the magnetic field at the surface reorganizes after some time (Zirin and Wang, 1993).

These two events of the space weather have been the center of enormous research and attention because of their potential affect at the earth as well. Mitra discussed the impact of solar flares and geomagnetic storm connected with them to argue the criticality of these flares for human beings (Mitra, 1974). These solar flares sometimes contains sun surface particles radiations which are collectively called as coronal mass ejection (CMEs). The practical impact of these flares and accompanied CMEs on surface of earth varies with the intensity of flares and in some severe cases, the CME particles can disturb the neutral nature of earth making it a slightly charged body which can then interfere power grids. Such an event happened in 2003 when a multiple solar flares and associated CMEs reached the earth surface and high magnetic fields disturbed the Polar Regions causing power outage in Sweden for about an hour (Dunbar, 2019).

These sort of events are cause damages to the human race and have been started to be counted as a natural disaster as well (Zielinski, 2019). Also, predicting their occurrence can be very helpful in preparing the required measures and finding methods to predict solar flares and CME have been a popular research area in the past as well (Qahwaji and Colak, 2007) (Bobra and Couvidat, 2015). As Zirin and Wang pointed in their work that complex or unstable sunspots usually originated solar flares, this report will correlate the sunspots and solar flares data to see if growth rate of sunspot size can be a measure of sunspot instability and a potential source of solar flares or not.

This work will analyze the growth of sunspot size with time and the overall rate of change in the sunspot sizes and will map that rate to the see **whether a high growth rate can predict an intense type of flares (e.g. M or X type flare) or not?**

Background

Relevant Work

The solar flares' prediction had been a popular research problem among the academia and various methodologies have been proposed for predicting these events. A very recent work is done by Zheng et. al as they used Hybrid Deep Convolutional Neural Network for predicting the solar flares (Zheng, Li and Wang, 2019). A similar effort focusing on predicting an expected solar flare is proposed by Florios et al. in which they have used machine learning and magnetogram-based predictors for achieving the goals (Florios et al., 2018).

Overview of dataset

As mentioned above, the space weather has been the focus of research interest over the past few decades and NASA is maintaining a very rich data repository of space events. Sunspots and the solar flares' data, because of their criticality, have also been recorded with details. As the focus of this report is to correlate sunspots' data with solar flares' data, both of these datasets were taken from the official website of NASA.

The sunspots and solar flares data contain record of sunspots and solar flares, respectively, between years 1981 to 2017. Following are the useful features of sunspot and flares with their explanation:

Sunspots		Solar Flares	
Feature Name	Detail	Feature Name	Detail
Date	Date of sunspot	Date	Date of sunspot
Time	Time of sunspot	Time	Time of sunspot
NOAA Number	Identification number	NOAA Number	Identification number of associated sunspot
Area	Area of sunspot in millionths of solar hemisphere	Type	Type of solar flares (A,B,C,M,X) with X having extreme intensity

Experimental Details Main Part

Experimental aim

The purpose was to analyze the data for mapping sunspot growth rate with the corresponding flares. That is, if a sun-spot's is being changing rapidly, is it more prone to producing sun flares or not. Also, if it does, what kind of flares, it can emit.

Data Preprocessing

After reading the data description, it was decided to separate out NOAA number and area of sunspots from the file using excel and made a separate excel sheet for that.

Next task for this is to import this data into Python and visualize the rate of growth of sunspots. To see how the size of every sunspot is changing.

Experimental Phases

Before moving on towards finding rate, it was decided to find the magnitude of change for each of the sunspot. That is, two phases of experiment were decided to achieve the experimental aim:

1. Finding correlation of magnitude of change in sunspot area with solar flare types

2. Finding correlation of rate of change of sunspot area with relevant solar flare types

Correlating magnitude of change in Sunspot area with Flare types

The collective change of a particular sunspot was found by iterating the following algorithm over sunspot data

This collective change was found using the following code:

```
for i in range(0,length):
    sunspot = c.NOAA[i]
    if math.isnan(sunspot) or math.isnan(c.Size[i]):
        continue
    if sunspot in sunspot_number:
        sunspot_number[sunspot] = sunspot_number[sunspot] + 1
        sunspot_change[sunspot] = sunspot_change[sunspot] + abs(prev_sunspot[sunspot] - c.Size[i])
        prev_sunspot[sunspot] = c.Size[i]
    else:
        prev_sunspot[sunspot] = c.Size[i]
        sunspot_number[sunspot] = 1
        sunspot_change[sunspot] = c.Size[i]
```

After finding rate of growth, these lists were sorted w.r.t the rates and top 8 entries were plotted using `pyplot`.

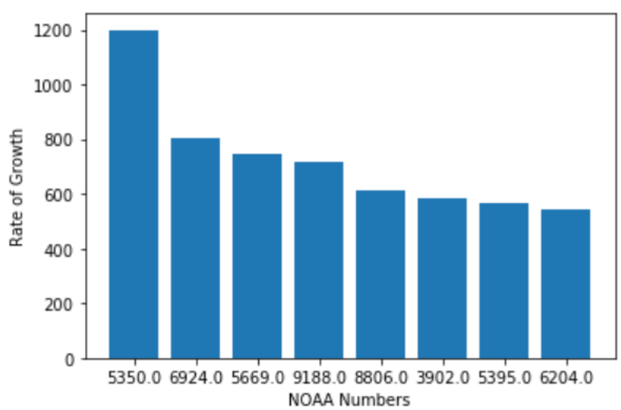


Figure 1: Sunspots having the highest growth change

Just to analyze importance of this finding, flare types against some of top sunspots were found.

X Flares among the top 100 spots were plotted in this figure.

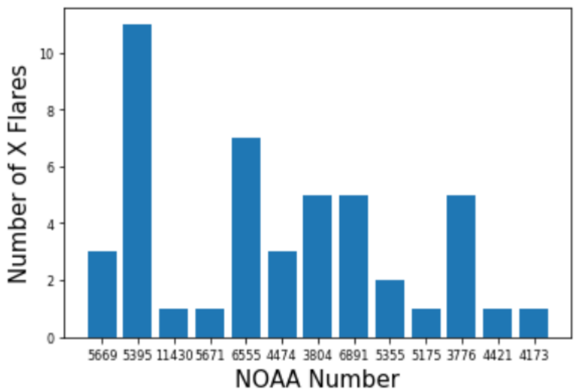


Figure 2: Sun-flares originated from sunspots having the highest change

After this, those sunspots were analyzed for which this change was the least. For this purpose, the lowest 100 spots were picked and mapped with the flares data. Interestingly, there was not a single sunspot in those which had produced an X type solar flare.

As the sunspots were already sorted with respect to the change of size, it was decided to count total number of X flares originated from the first half of the sunspots and from the second half. It was achieved by the code shown in Figure 1 of appendix. Likewise, observation for M flares was made using the code shown in Figure 2 in appendix.

Results of Experimental Phase-I

Following observation was made:

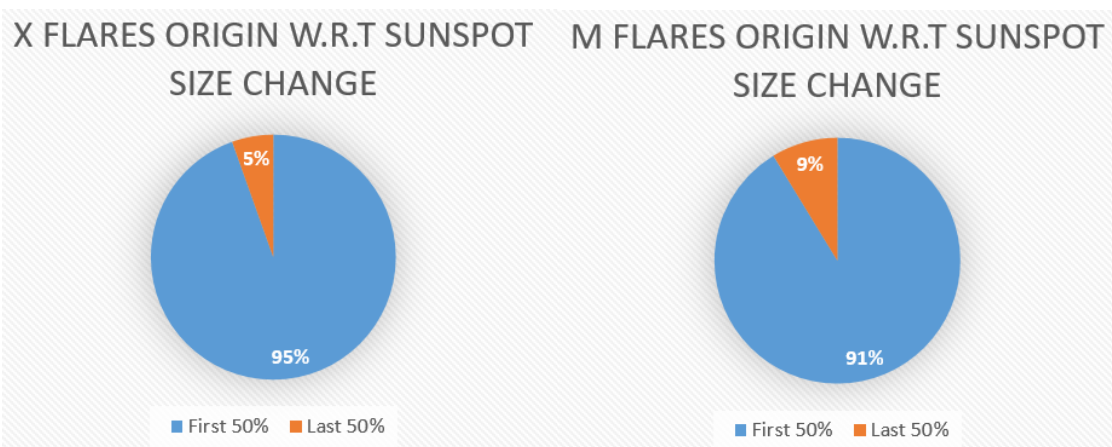


Figure 3: Results of Phase-I sun-flares from the top 50% sunspots having the high area change and those having the last 50% of the change

Analysis and Discussion on Phase-I Experiment

Figure 3 shows a clear difference in number of flares in both of the halves for each of the solar flare type. That is, from the top 50% of sunspots having the high size change, 329 X type flares were originated. On the other hand, for the last 50% of sunspots, this number was reduced to 19, which is a drastic change. Similarly, for the M flare types, there was a significant change in the top half and the lowest half of the sunspots. On the basis of this observation, it can be deducted that the magnitude of change in area of sunspot is related with the emission probability of an intense type flare. This can be concluded from the above observation that for the spots having high size changes, there is a higher probability to originate an X and M type flares than those having a low or no size changes.

Correlating rate of change of sunspot area with relevant solar flare types

Aim for this phase

It was decided to compare the first reading of the day for size comparison. That is, we wanted to calculated rate of change of size per day for a spot and map that to the corresponding flare types. Also, once the rate of change for those two days were found, we had to check the corresponding flare type within those two days for analyzing what kind of flare type was originated in those days.

Data Preprocessing for Phase-II

For persistence of format across flares and sunspot data, date format in flares was changed such that initial three bytes were removed and next two bytes of were replaced with 11. That is, the date changed from 31777810102 to 11810102, which was the same format as that in sunspot file.

Implementation Details

As the aim was to compare first readings of the day for each unique sunspot, we had to maintain details of sunspot number and the corresponding date so that we can search for the same date and number in solar flares data. This was achieved by iterating through all the solar flares data first and maintaining a separate Python dictionary for X and M flare types with NOAA as their keys pointing to the list of dates at which that flare was occurred.

This was done using the piece of code shown in figure 4:

```
x_flares_day = {}

flares.NOAA = flares.NOAA[:].astype(numpy.float)
flares.date = flares.date[:].astype(numpy.float)

for noaa in flares.NOAA:
    if math.isnan(flares.date[counter]) or math.isnan(noaa):
        counter += 1
        continue
    if flares.type[counter] == 'X':
        if noaa in x_flares_day:
            x_flares_day[noaa].append(flares.date[counter])
        else:
            x_flares_day[noaa] = [flares.date[counter]]
    counter += 1
```

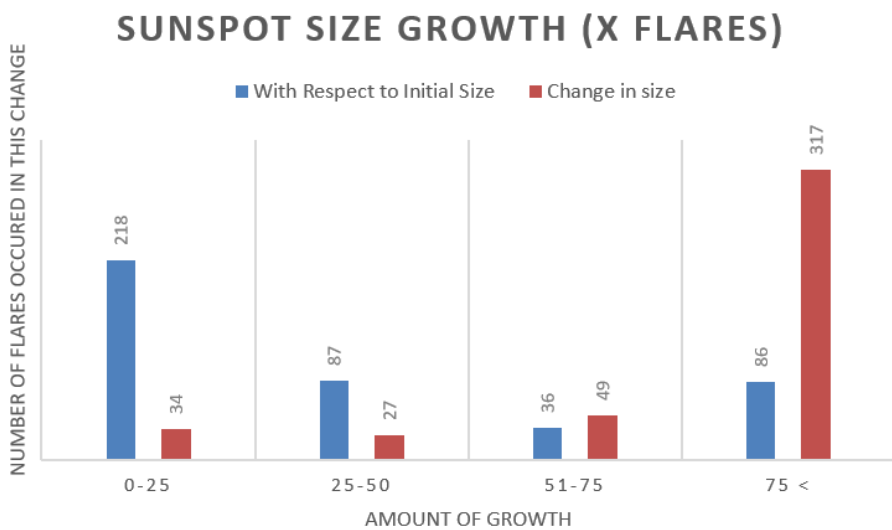
Figure 4: Code for Solar Flares

The next task was to iterate through the sunspots data and correlate it to the solar flares dictionary we had. For this, we iterate through the sunspots data and for every sunspot, we checked if it is already in our solar flares dictionary. If we found a record for which there was solar flare of type X or M against a specific sunspot on the same day as that of our record, we computed the difference of areas and appended that to our list of area changes. Here we were maintaining a list of dates corresponding to x type flares for a specific sunspot as shown in figure 5. Once the area difference was found, we also gathered a new value by dividing the difference with initial size (commented line of code in figure). So, for all the sunspots record, we computed the magnitude of area differences and their values with respect to the initial value of area. Same was done for m type flares.

```
if sunspots.NOAA[counter] in x_flares_day: # the same sunpot is present in x flares
    temp = x_flares_day[sunspots.NOAA[counter]]
    for day in temp:
        if day == prev_day or day == sunspots.date[counter]:
            diff = abs(sunspots.Size[counter] - prev_size)
            #diff = (diff / prev_size)*100
            rates.append(diff)
```

Figure 5: Code for rates

Experiment Phase-II Results



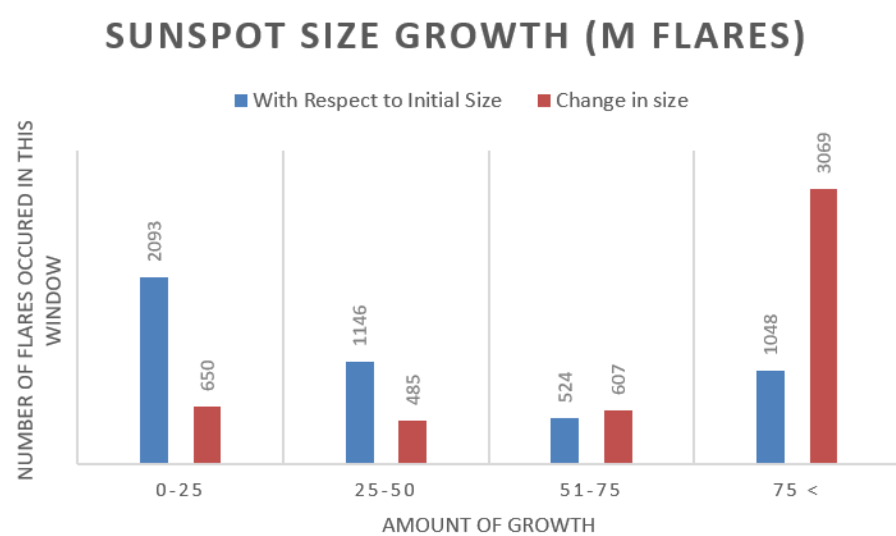


Figure 6: Change in area with number of flares

Experiment Phase-II Analysis and Discussion

The results of this experiment were interesting. Figure 6 shows a relation for amount of area change and the number of X and M type solar flares against the sunspots having that change in area.

These results indicated that, as compared to the rate of growth with respect to the initial area of sunspot, a rapid change, despite the initial area of the sunspot is a better indication of unstable sunspot which is prone to high intensity solar flares. Moreover, if we use change in area of sunspots, without relating it to their initial size, we can see that when the change in area of a sunspot is more than 75 millionths of solar hemisphere, that sunspot is more prone to emitting M and X flares. On a more general level, it can also be concluded from the results that when there is an abrupt change in a sunspot, it is more likely that it will emit an intense type of flare.

Further Investigation using AI Techniques

As it is discussed in the relevant work that machine learning and artificial intelligence have been used for prediction of solar flares. Specifically, the findings and results of these reports can be used for better predictions of solar flares. To illustrate, the report found a clear relation between the abrupt change of a sunspot area with the intensity type of corresponding solar flare. A machine learning model can be trained on the basis of area change with the flares type so that it can identify the expected type of solar flare once it is given area change as input to the system.

References

Solanki, S. (2003). Sunspots: An overview. *Astronomy and Astrophysics Review*, 11(2-3), pp.153-286.

Erickson, K. (2019). *Sunspots and Solar Flares*. [online] NASA Space Place. Available at: https://spaceplace.nasa.gov/solar-activity/en/ [Accessed 16 Dec. 2019].

Zirin, H. and Wang, H. (1993). Narrow lanes of transverse magnetic field in sunspots. *Nature*, 363(6428), pp.426-428.

Mitra, A. (1974). *Ionospheric Effects of Solar Flares*. Dordrecht: Springer, Dordrecht.

Dunbar, B. (2019). *Halloween Storms of 2003 Still the Scariest*. [online] NASA. Available at: https://www.nasa.gov/topics/solarsystem/features/halloween_storms.html [Accessed 16 Dec. 2019].

Zielinski, S. (2019). *Could The Sun Set Off The Next Big Natural Disaster?*. [online] Smithsonian. Available at: https://www.smithsonianmag.com/science-nature/could-the-sun-set-off-the-next-big-natural-disaster-78352782/ [Accessed 16 Dec. 2019].

Qahwaji, R. and Colak, T. (2007). Automatic Short-Term Solar Flare Prediction Using Machine Learning and Sunspot Associations. *Solar Physics*, 241(1), pp.195-211.

Bobra, M. and Couvidat, S. (2015). SOLAR FLARE PREDICTION USINGSDO/HMI VECTOR MAGNETIC FIELD DATA WITH A MACHINE-LEARNING ALGORITHM. *The Astrophysical Journal*, 798(2), p.135.

Andrews, D. (1972). Plots of High-Dimensional Data. *Biometrics*, 28(1), p.125.

Tableau Software. (2019). *Tableau: Business Intelligence and Analytics Software*. [online] Available at: https://www.tableau.com/ [Accessed 16 Dec. 2019].

Zheng, Y., Li, X. and Wang, X. (2019). Solar Flare Prediction with the Hybrid Deep Convolutional Neural Network. *The Astrophysical Journal*, 885(1), p.73.

Nishizuka, N., Sugiura, K., Kubo, Y., Den, M., Watari, S. and Ishii, M. (2017). Solar Flare Prediction Model with Three Machine-learning Algorithms using Ultraviolet Brightening and Vector Magnetograms. *The Astrophysical Journal*, 835(2), p.156.

Appendix

```
counter = 1
sum_first = 0
sum_second = 0

for key in sorted_dict:
    if key in flares_type_x:
        if counter < len(sorted_dict) // 2:
            sum_first = sum_first + flares_type_x[key]
        else:
            sum_second = sum_second + flares_type_x[key]

    counter = counter + 1

print("X Flares Result ", sum_first, " ", sum_second)
```

Figure 3

```
for key in sorted_dict:
    if key in flares_type_m:
        if counter < len(sorted_dict) // 2:
            sum_first = sum_first + flares_type_m[key]
        else:
            sum_second = sum_second + flares_type_m[key]

    counter = counter + 1

print("M Flares Result ", sum_first, " ", sum_second)
```

Figure 4