Predicting Coronal Mass Ejections Using Topological Data Analysis-based Classifier

E. S. Ivshina

November 15, 2020

1 Introduction

A Coronal Mass Ejection (CME) is a significant release of plasma and magnetic field from the Sun into interplanetary space. Despite the progress in numerical modeling, it is still unclear which conditions will produce a CME; however, it is known that CMEs and solar flares are associated as "a single magnetically driven event" (Webb and Howard, 2012). In this study, we use topological data analysis-based classifier to predict whether an M- or X-class flaring active region will produce a CME in the next 24 hours.

2 Data

We build a catalog of active regions (ARs) that either produced both a flare and a CME (the positive class) or only a flare (the negative class). To determine if an AR produced a CME, we retrieve data from the *SOHO* Large Angle and Spectrometric Coronagraph Experiment (LASCO) instrument and both coronagraphs on the STEREO Sun Earth Connection Coronal and Heliospheric Investigation (SECCHI) instruments by querying the DONKI database[1] for events between 2010 May 1 and 2019 July 1. We reject flares unassociated with an AR (as it is not possible to compute features for such events) and flares below the M1.0-class (as they release a limited amount of energy). To determine whether an AR produced an X- or M-class flare, we retrieve a GOES flare list by querying the Heliophysics Events Knowledgebase[2] (Hurlburt et al., 2012) using a SunPy python library (SunPy Community et al., 2015). We reject flares in the flare list unassociated with an AR and that are not within ±70 degrees of central meridian during the GOES X-ray Flux peak time because of the significant decrease in the signal-to-noise ratio in the HMI vector magnetic field data.

[1] `http://kauai.ccmc.gsfc.nasa.gov/DONKI/`
[2] `http://www.lmsal.com/hek`

We use Spaceweather HMI Active Region Patches (SHARPs) data[3] to characterize every event in our catalog throughout a 24 hour period before the GOES X-ray flare peak. We do this by choosing 4 points throughout this period, at t=24, 18, 12, and 6 hours before the GOES X-ray flare peak. SHARPs data contains 18 features that parameterize the magnetic field within ARs observed by the SDO, such as the magnetic flux contained in an AR and the current helicity (for details, see Bobra et al. (2014)).
We also restrict events to those where (1) the absolute value of the radial velocity of SDO is less than 3500 m s^{-1} and (2) HMI data is of high quality at this time. In the end, we are left with 62 events in the positive class and 338 events in the negative class in our catalog. In this paper, we extend the work of Bobra and Ilonidis (2016) to include four more years of data and implement a new classifier.

3 Topological Data Analysis-based Classifier

Topological Data Analysis (TDA) is a set of tools for studying the "shape" of data using geometry and methods from algebraic topology (Carlsson, 2014).
The TDA-based classifier presented here begins by randomly sampling each sample using a number of data points[4]. Then, the algorithm applies two one-dimensional filter functions (the first and the second principal components in our study) to the data space. Next, the range of values created by each filter function is divided into non-overlapping intervals of some arbitrary length. Within these intervals, local clustering is conducted. The linkage method can be chosen; in our study, the metric is Euclidean distance, and the linkage is complete linkage. Since the created clusters contain a number of data points from each of the samples, an $n \times m$ matrix is constructed as input to a Feedforward Neural Network with one hidden layer where n is the number of samples, m is the number of clusters created after applying two filter functions to the data space, and entries are the number of data points in a given cluster.

4 Results

The algorithm's sampling rate can be chosen; since each event is represented with only 4 data points in our study, we use all 4 data points to characterize every event. 50 runs of the TDA-based classifier were conducted to obtain the average classification accuracy (the ratio of the number of true positives to the total number of predictions made). The results are presented in Figure 1. The TDA-based classifier has 80.5% accuracy when using 15 training examples, increasing to 86.2% with 285 training examples. In general, the classifier seems to be more stable (the classification accuracy deviates less among 50 runs) when increasing the number of training examples.

[3] `http://jsoc.stanford.edu`
[4] A "data point" is a single set of 18 SHARP features. Each "sample" is an event in the

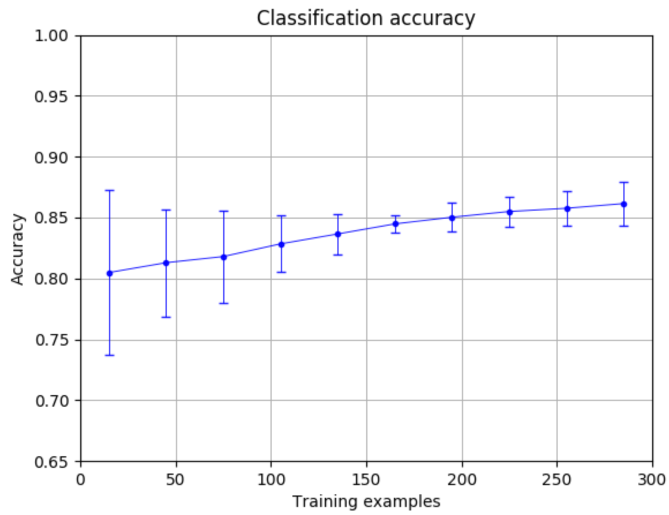

Figure 1: The average prediction accuracy of the TDA-based classifier as a function of the number of training examples. The error bars show the standard deviation from 50 independent runs.

5 Code

The code is available here: `https://zenodo.org/badge/latestdoi/208948304`

6 Acknowledgements

The author would like to thank M. G. Bobra for reviewing this paper.

References

Bobra, M., Holdgraf, C., Mason, J., et al. (2019). Helioml/helioml: Helioml 0.2.0 (2019-02-22).

Bobra, M. G. and Ilonidis, S. (2016). PREDICTING CORONAL MASS EJECTIONS USING MACHINE LEARNING METHODS. *The Astrophysical Journal*, 821(2):127.

Bobra, M. G., Sun, X., Hoeksema, J. T., et al. (2014). The helioseismic and magnetic imager (hmi) vector magnetic field pipeline: Sharps – space-weather hmi active region patches. *Solar Physics*, 289(9):3549–3578.

catalog associated with four data points.

Carlsson, G. (2014). Topological pattern recognition for point cloud data. *Acta Numerica*, 23:289–368.

Hurlburt, N., Cheung, M., Schrijver, C., et al. (2012). Heliophysics event knowledgebase for the Solar Dynamics Observatory (SDO) and beyond. *Solar Physics*, 275(1):67–78.

Riihimäki, H., Chacholski, W., et al. (2019). A topological data analysis based classification method for multiple measurements. *CoRR*, abs/1904.02971.

SunPy Community, T., Mumford, S. J., Christe, S., et al. (2015). SunPy Python for solar physics. *Computational Science and Discovery*, 8(1):014009.

Webb, D. F. and Howard, T. A. (2012). Coronal mass ejections: Observations. *Living Reviews in Solar Physics*, 9(1):3.