Interpretable Flare Prediction with Integrated Data: SHARP parameters, Spatial Statistics Features and HMI Images[1]

Hu Sun [1]

Ward Manchester[2], Yang Chen[1]

[1]Department of Statistics, University of Michigan, Ann Arbor

[2]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor

Dec 16, 2021

[1]Our submitted manuscript can be viewed here, and the paper is to appear soon in *Space Weather*.

Overview

Flare Prediction with HMI Magnetograms

- Bobra, Sun, et al. (2014) introduced the Space-weather HMI Active Region Patch (SHARP) parameters, which are derived from the HMI/SDO magnetograms and have been used frequently for solar flare prediction models in recent years (e.g. Bobra and Couvidat, 2015; Florios et al., 2018; Chen et al., 2019; Camporeale, 2019; Jiao et al., 2020).

Flare Prediction with HMI Magnetograms

- Bobra, Sun, et al. (2014) introduced the Space-weather HMI Active Region Patch (SHARP) parameters, which are derived from the HMI/SDO magnetograms and have been used frequently for solar flare prediction models in recent years (e.g. Bobra and Couvidat, 2015; Florios et al., 2018; Chen et al., 2019; Camporeale, 2019; Jiao et al., 2020).
- There are efforts using deep neural network methods, which directly take the HMI/SDO magnetogram images to predict solar eruptions (e.g. the Long Short Term Memory network adopted by Chen et al. (2019) and Liu et al. (2019)).

Flare Prediction with HMI Magnetograms

- Bobra, Sun, et al. (2014) introduced the Space-weather HMI Active Region Patch (SHARP) parameters, which are derived from the HMI/SDO magnetograms and have been used frequently for solar flare prediction models in recent years (e.g. Bobra and Couvidat, 2015; Florios et al., 2018; Chen et al., 2019; Camporeale, 2019; Jiao et al., 2020).
- There are efforts using deep neural network methods, which directly take the HMI/SDO magnetogram images to predict solar eruptions (e.g. the Long Short Term Memory network adopted by Chen et al. (2019) and Liu et al. (2019)).
- Recent efforts (Deshmukh, Berger, Bradley, et al., 2020; Deshmukh, Berger, Meiss, et al., 2020) leverage the shape information contained in HMI magnetograms to construct interpretable and predictive new parameters for flare prediction.

Highlights of Our Work

1. Expand the feature set derived from the HMI magnetograms for flare prediction using tools from both *topological data analysis* and *spatial statistics*.

Highlights of Our Work

1. Expand the feature set derived from the HMI magnetograms for flare prediction using tools from both *topological data analysis* and *spatial statistics*.
2. Derive features not only from the PIL-masked HMI magnetograms but also from the spatial distribution of SHARP parameters.

Highlights of Our Work

1. Expand the feature set derived from the HMI magnetograms for flare prediction using tools from both *topological data analysis* and *spatial statistics*.
2. Derive features not only from the PIL-masked HMI magnetograms but also from the spatial distribution of SHARP parameters.
3. Marginally but steadily improved the skill score of the classification model of strong vs. weak solar flares.

Dataset

- We use the Geostationary Operational Environmental Satellites (GOES) flare list spanning 2010/12 - 2018/06 for collecting flare events, leading to 399 M/X class flares and 1,972 B class flares coming from 487 HARP regions.

Dataset

- We use the Geostationary Operational Environmental Satellites (GOES) flare list spanning 2010/12 - 2018/06 for collecting flare events, leading to 399 M/X class flares and 1,972 B class flares coming from 487 HARP regions.
- For each flare, we collect its corresponding high-resolution HMI magnetogram data from the JSOC at 4 time points: 1, 6, 12, 24 hours prior to the peak soft X-ray flux.

Dataset

- We use the Geostationary Operational Environmental Satellites (GOES) flare list spanning 2010/12 - 2018/06 for collecting flare events, leading to 399 M/X class flares and 1,972 B class flares coming from 487 HARP regions.
- For each flare, we collect its corresponding high-resolution HMI magnetogram data from the JSOC at 4 time points: 1, 6, 12, 24 hours prior to the peak soft X-ray flux.
- For each flare, at all four time points, raw data of the B_r, B_p, B_t components of the magnetic field are collected.

Derive SHARP Parameter Maps

- We derive features from the B_r component of the magnetic field but also from other secondary maps derived from the B_r, B_p, B_t components, which we call *SHARP parameter maps*.

Derive SHARP Parameter Maps

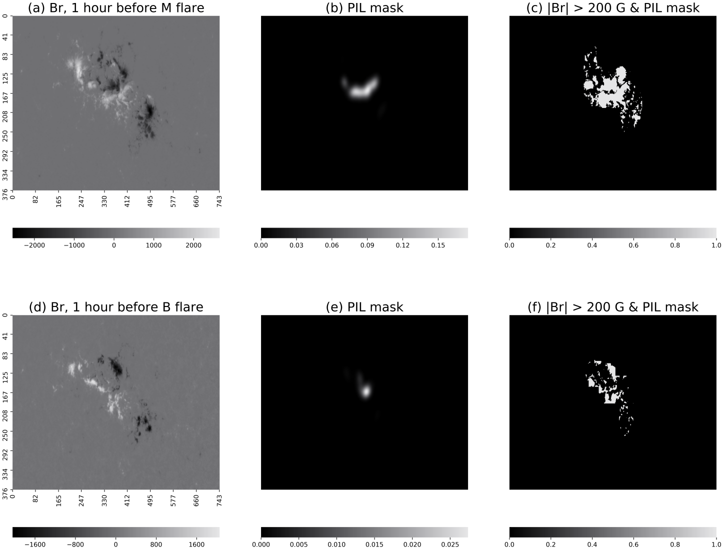
Channel	Formula	Unit
Br	$\mathbf{B}_z$	G
GAM	$\arctan\left(\frac{\mathbf{B}_h}{\mathbf{B}_z}\right)$	Degree
GBT	$\sqrt{\left(\frac{\partial \mathbf{B}}{\partial x}\right)^2 + \left(\frac{\partial \mathbf{B}}{\partial y}\right)^2}$	G$\times$ Mm^{-1}
GBH	$\sqrt{\left(\frac{\partial \mathbf{B}_h}{\partial x}\right)^2 + \left(\frac{\partial \mathbf{B}_h}{\partial y}\right)^2}$	G$\times$ Mm^{-1}
GBZ	$\sqrt{\left(\frac{\partial \mathbf{B}_z}{\partial x}\right)^2 + \left(\frac{\partial \mathbf{B}_z}{\partial y}\right)^2}$	G$\times$ Mm^{-1}
USJZ	$\lvert\left(\frac{\partial \mathbf{B}_y}{\partial x} - \frac{\partial \mathbf{B}_x}{\partial y}\right)\rvert$	A
USJH	$\lvert \boldsymbol{J}_z \times \mathbf{B}_z \rvert$	G^2 m $^{-1}$
POT	$\left((\mathbf{B}_x - \mathbf{B}_x^{POT})^2 + (\mathbf{B}_y - \mathbf{B}_y^{POT})^2\right)$	erg cm^{-3}
SHR	$\arccos\left(\frac{\mathbf{B}_x^{POT}\times\mathbf{B}_x+\mathbf{B}_y^{POT}\times\mathbf{B}_y+\mathbf{B}_z^2}{\sqrt{\mathbf{B}_x^{POT2}+\mathbf{B}_y^{POT2}+\mathbf{B}_z^2}\sqrt{\mathbf{B}_x^2+\mathbf{B}_y^2+\mathbf{B}_z^2}}\right)$	Degree

Table 1. SHARP parameter mask, formula applied to every pixel of the HMI magnetogram. Here, $\mathbf{B}_x, \mathbf{B}_y, \mathbf{B}_z$ are the x, y, z components of the magnetic field and $\mathbf{B}_x^{POT}, \mathbf{B}_y^{POT}$ the potential field components respectively. Detailed definition of the parameters can be found in Table 3 of Bobra et al. (2014).

Derive the Polarity Inversion Line (PIL)

- We focus specifically on the area adjacent to the polarity inversion line (PIL) by constructing the PIL mask.

Derive the Polarity Inversion Line (PIL)



Topological Feature: Betti Numbers

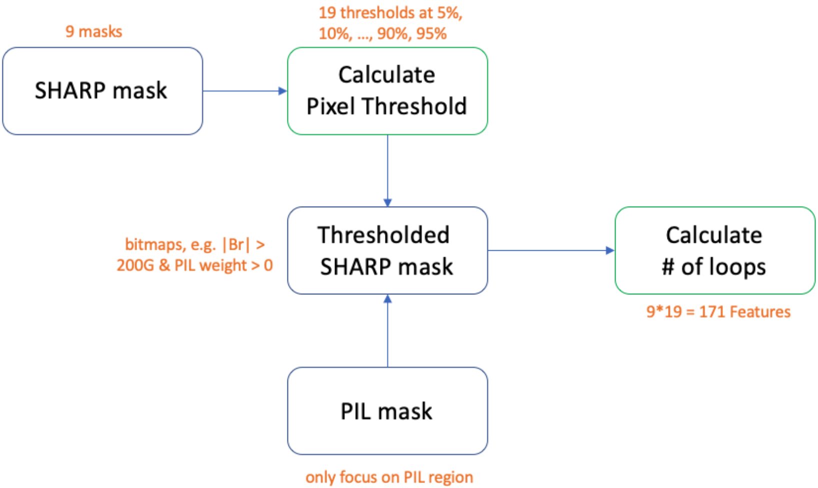


Figure: Feature Engineering Pipeline of Topological Features

Topological Feature: Betti Numbers

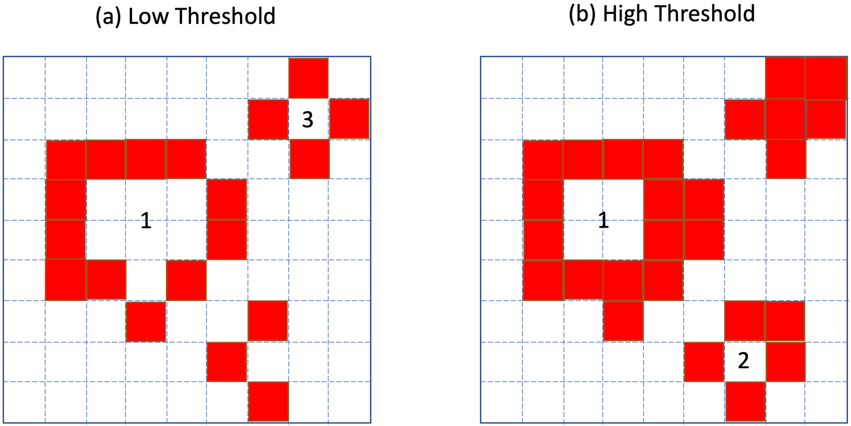


Figure: What is a loop?

Topological Feature: Betti Numbers

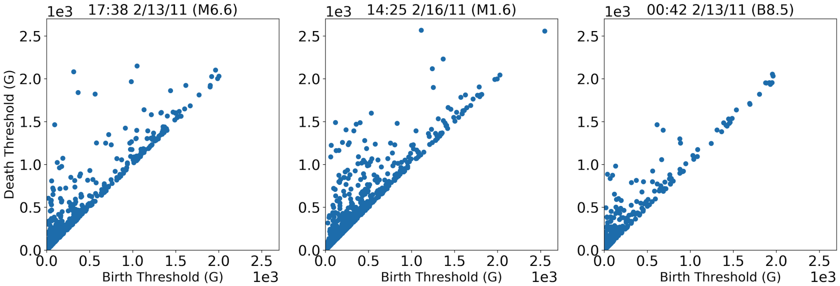
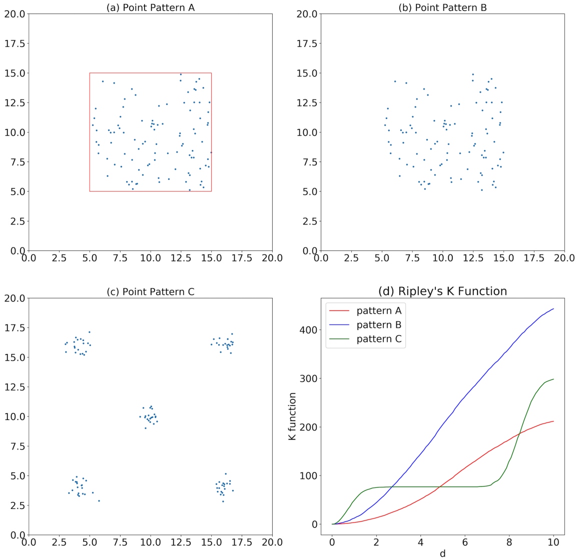


Figure: Final Topological Feature in Persistence Diagram

Spatial Feature I: Ripley's K Function



Spatial Feature I: Ripley's K Function

- For the thresholded B_r mask, we randomly pick 500 pixels, with sampling probability proportional to $|B_r|$, to construct a point cloud. Each picked pixel has a pair of (x, y) pixel coordinates in the 2D pixel grid.
- Ripley's K function:

$$L(d) = \sqrt{\frac{A \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} k_{i,j}}{\pi n(n-1)}},$$

where $k_{i,j} = 1$ if the i-th and j-th pixel are within distance d, and $n = 500$ in our case. A is the area size and is defined as the number of PIL pixels in our study.

Spatial Feature I: Ripley's K Function

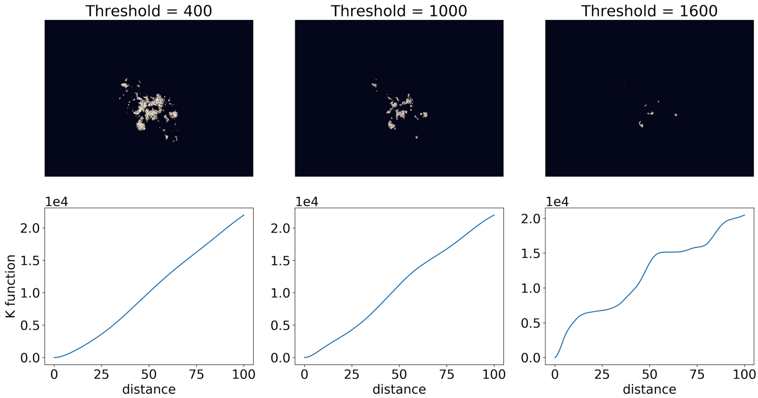
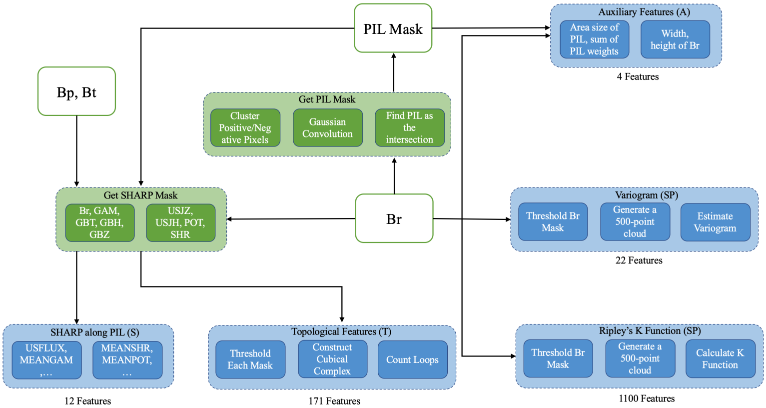


Figure: Point cloud and the corresponding Ripley's K function for the B_r mask collected from HARP 377, 1 hour before the M flare peaked at 2011.02.13 17:38. The top row includes 3 point clouds generated by 3 thresholds at 400G, 1000G, 1600G. The bottom row shows the 3 corresponding Ripley's K functions.

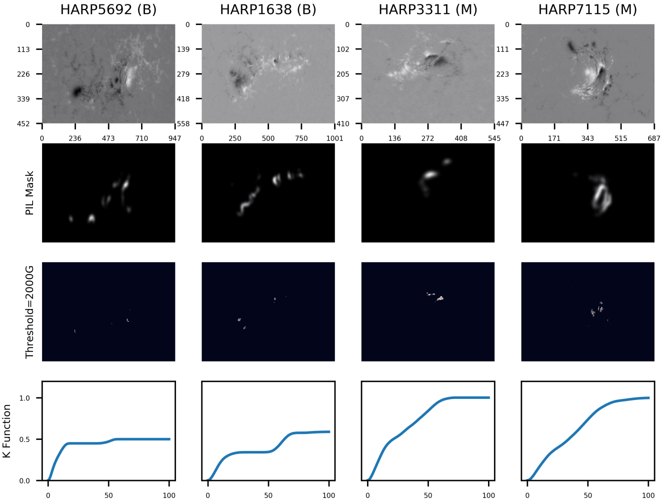
Feature Pipeline Summary



Result: True Skill Score for Xgboost Model

Feature Combination	Prediction Time (hour) 1	6	12	24
S	0.553 (0.075)	0.555 (0.071)	0.539 (0.068)	0.489 (0.077)
T	0.548 (0.069)	0.575 (0.071)	**0.561** (0.063)	**0.525** (0.069)
SP	0.558 (0.066)	**0.578** (0.076)	0.546 (0.071)	**0.528** (0.072)
S+T	**0.578** (0.071)	**0.581** (0.072)	0.554 (0.057)	**0.536** (0.052)
S+SP	0.56 (0.059)	**0.58** (0.073)	0.538 (0.078)	**0.533** (0.074)
S+T+SP	**0.586** (0.077)	**0.599** (0.068)	0.558 (0.08)	**0.57** (0.06)
S+T_PC+SP_PC	0.554 (0.075)	0.561 (0.077)	0.53 (0.082)	**0.533** (0.076)
S+T+SP+A	**0.587** (0.071)	**0.605** (0.063)	0.551 (0.077)	**0.55** (0.059)
S+T_PC+SP_PC+A	**0.578** (0.068)	0.561 (0.071)	0.533 (0.076)	**0.521** (0.089)

Result: Example of the Ripley's K function



Conclusion

In this project, we:

- Concentrate on SHARP parameter spatial distributions along the polarity inversion line regions.

Conclusion

In this project, we:

- Concentrate on SHARP parameter spatial distributions along the polarity inversion line regions.
- Engineered interpretable and predictive features summarizing the spatial variation, dispersion patterns of various SHARP quantities, especially the B_r component, using tools from TDA and spatial statistics.

Conclusion

In this project, we:

- Concentrate on SHARP parameter spatial distributions along the polarity inversion line regions.
- Engineered interpretable and predictive features summarizing the spatial variation, dispersion patterns of various SHARP quantities, especially the B_r component, using tools from TDA and spatial statistics.
- Obtained marginal but steady improvement on the solar flare classification task.

Conclusion

In this project, we:

- Concentrate on SHARP parameter spatial distributions along the polarity inversion line regions.
- Engineered interpretable and predictive features summarizing the spatial variation, dispersion patterns of various SHARP quantities, especially the B_r component, using tools from TDA and spatial statistics.
- Obtained marginal but steady improvement on the solar flare classification task.
- The spatial features derived solely from the B_r component are as good or better for flare prediction than full vector SHARP parameters. Theoretically interesting and important for future missions.

References I

Bobra, M. G., X. Sun, J. T. Hoeksema, M. Turmon, Y. Liu, K. Hayashi, G. Barnes, and K. D. Leka (Sept. 2014). “The Helioseismic and Magnetic Imager (HMI) Vector Magnetic Field Pipeline: SHARPs – Space-Weather HMI Active Region Patches”. In: *Solar Physics* 289.9, pp. 3549–3578.

Bobra, Monica G and Sebastien Couvidat (2015). “Solar flare prediction using SDO/HMI vector magnetic field data with a machine-learning algorithm”. In: *The Astrophysical Journal* 798.2, p. 135.

Camporeale, Enrico (July 2019). “The Challenge of Machine Learning in Space Weather Nowcasting and Forecasting”. In: *Space Weather* 17. DOI: 10.1029/2018sw002061.

References II

Chen, Yang, Ward B Manchester, Alfred O Hero, Gabor Toth, Benoit DuFumier, Tian Zhou, Xiantong Wang, Haonan Zhu, Zeyu Sun, and Tamas I Gombosi (2019). "Identifying solar flare precursors using time series of SDO/HMI Images and SHARP Parameters". In: *Space Weather* 17.10, pp. 1404–1426.

Deshmukh, Varad, Thomas Berger, James Meiss, and Elizabeth Bradley (2020). "Shape-based Feature Engineering for Solar Flare Prediction". In: *arXiv preprint arXiv:2012.14405*.

Deshmukh, Varad, Thomas E Berger, Elizabeth Bradley, and James D Meiss (2020). "Leveraging the mathematics of shape for solar magnetic eruption prediction". In: *Journal of Space Weather and Space Climate* 10, p. 13.

References III

Florios, Kostas, Ioannis Kontogiannis, Sung-Hong Park, Jordan A Guerra, Federico Benvenuto, D Shaun Bloomfield, and Manolis K Georgoulis (2018). "Forecasting Solar Flares Using Magnetogram-based Predictors and Machine Learning". In: *Solar Physics* 293.2, p. 28. DOI: doi:10.1007/s11207-018-1250-4.

Jiao, Zhenbang, Hu Sun, Xiantong Wang, Ward Manchester, Tamas Gombosi, Alfred Hero, and Yang Chen (2020). "Solar flare intensity prediction with machine learning models". In: *Space Weather* 18.7, e2020SW002440.

Liu, Hao, Chang Liu, Jason T. L. Wang, and Haimin Wang (June 2019). "Predicting Solar Flares Using a Long Short-term Memory Network". In: *The Astrophysical Journal* 877.2, p. 121. DOI: 10.3847/1538-4357/ab1b3c. URL: https://doi.org/10.3847%2F1538-4357%2Fab1b3c.