solar flare and CME prediction: 3 ways

Monica Bobra
Stanford University

Jonas & Bobra, 2016, in development
[1] predicting flares using features

sdo observes the vector magnetic field

Schou et al., 2012
step 1: detect active region

Turmon et al., 2014
Bobra et al., 2014
step 2: identify mask
step 3: compute features

e.g.: $j_z \sim \frac{\partial}{\partial x} B_y - \frac{\partial}{\partial y} B_x$
step 3: compute features

\[ j_z \sim \frac{\partial}{\partial x} B_y - \frac{\partial}{\partial y} B_x \]
flare forecasting

features: 25 parameterizations of the photospheric magnetic field, e.g. current, flux, magnetic energy

- e.g. Tian et al., 2005; Török and Kliem, 2005; LaBonte, Georgoulis, and Rust, 2007; Moore, Falconer, and Sterling, 2012; Hagineal et al., 1984; Leka and Barnes, 2003a, 2003b, 2007; Cui et al., 2006; Barnes and Leka, 2006; Georgoulis and Rust, 2007; Mason and Hoeksema, 2010; Falconer, Moore, and Gary, 2008; Schrijver, 2007; Fisher et al., 2012; Bobra et al. 2014

model: machine-learning algorithm — binary classifier


first study to use a large sample of vector magnetic field data: can the HMI vector field tell us something new?

Bobra & Couvidat, 2015
step 1: build flare catalog

positive class: 303 flares; negative class: 5,000 no-flares
step 2: feature selection

compute the univariate f-score

inter-class distance

intra-class distance
step 2: feature selection

- number of features from lowest to highest univariate score
- univariate score
- included
- rejected
step 2: feature selection

#1: current helicity

\[ H_{c_{\text{total}}} \propto \sum |B_z \cdot J_z| \]

J \sim \text{small}  

J \sim \text{large}
step 2: feature selection

#2: magnitude of Lorentz force acting on upper solar atmosphere

\[ F \propto \sum B^2 \]
step 2: feature selection

#3: free magnetic energy

\[ \rho_{\text{tot}} \propto \sum \left( B^{\text{Obs}} - B^{\text{Pot}} \right)^2 dA \]
step 3: support vector machine

a non-linear binary classifier
step 3: support vector machine

benefit: non-linearity

linearly separable  non-linearly separable  inseparable
step 3: support vector machine

training set:
- 70% of data
- labels provided

testing set:
- 30% of data
- no labels provided
step 4: determine algorithm performance (by accounting for class imbalance problem)

Only TSS independent of imbalance ratio. (Woodcock, 1976; Bloomfield et al. 2012)
results

The graph shows the skill scores over the number of features from lowest to highest univariate score. The skill scores are represented by different lines:

- HSS\(_1\)
- HSS\(_2\)
- GS
- TSS

The x-axis represents the number of features, ranging from 25 to 5, and the y-axis represents the skill scores, ranging from 0.2 to 0.7.
[2] predicting coronal mass ejections using features

results: cme prediction

TSS per number of features

number of features

TSS

Bobra & Ilonidis, 2016
results: cme feature selection

#1: gradient of horizontal field

Bobra & Ilondis, 2016
Sun et al., 2015
Török & Kleim, 2005
reproducibility matters
[3] predicting flares using images

Jonas & Bobra, 2016, in development
improvement 1: add coronal information
improvements 2 and 3: add temporal information + flare history
improvement 4: use the entire image
With multiple non-linear layers, a system can implement extremely intricate functions of its inputs that are simultaneously sensitive to minute details — distinguishing Samoyeds from white wolves — and insensitive to large irrelevant variations such as the background, pose, lighting and surrounding objects.
summary

• we can predict whether or not an active region flares or a cme erupts better than flipping a coin
• codes are public
• since sdo takes so much data (~1.5 tb/day) it is important to analyze it automatically and quickly.
  one way to do this is by using machine-learning algorithms.