Short-term solar flare forecast

Victor Chernyshov, Dmitry Laptev

Moscow State University, Department of Computational Mathematics and Cybernetics

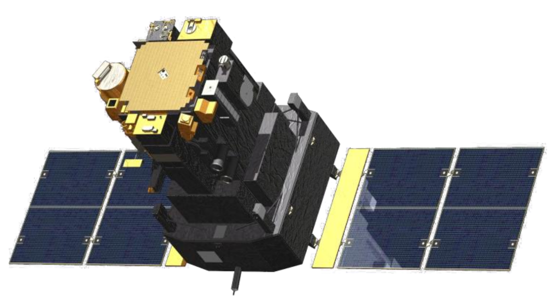
Introduction

In this poster we present our automated system for short-term solar flare forecast, using machine learning algorithms and high-quality satellite data. Any comments and suggestions are highly appreciated!

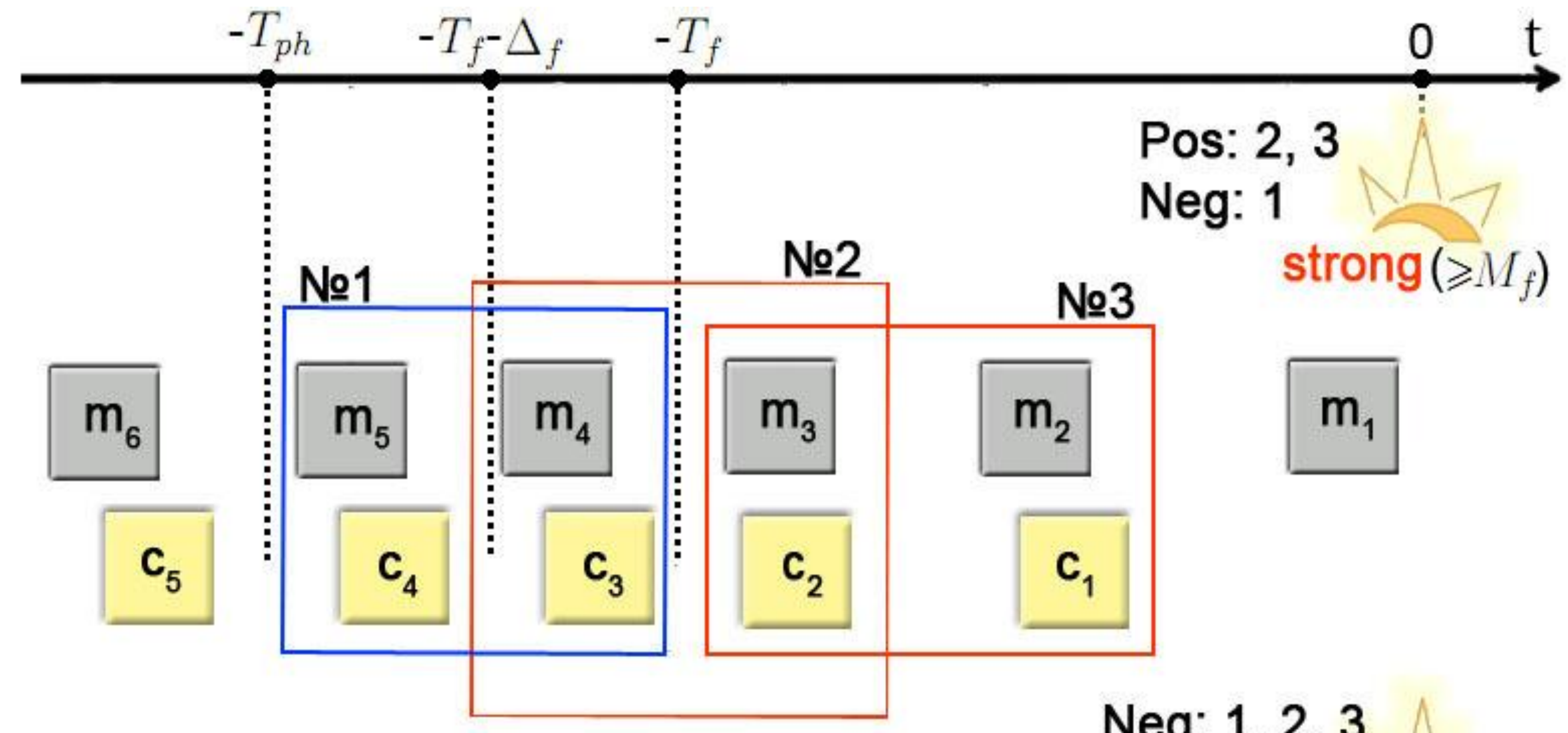
Physical basics. A *solar flare* is a sudden brightening observed over the Sun surface or the solar limb, which is interpreted as a large energy release. *Sunspots* are temporary phenomena on the photosphere of the Sun that appear in visible spectrum as dark spots compared to surrounding regions. An *active region (AR)* on the Sun is an area with an especially strong magnetic field; it is a place of sunspots and flares occurrence.

Data

SOHO/MDI provides both continuum ("visible", 2–5 per day) and magnetogram observations (in the vicinity of the Ni I 6767.8 Å photospheric absorption line, 6–14 per day) of the Sun. We created a flare base (NOAA flare catalogue), consisting of 24658 events, and an image base, including 40573 magnetograms and 14927 continuum images from 1996–2009 years interval (SOHO data).



Precedent model



$$M_f \in M_{set} = \{C5.0, M1.0, M5.0, X1.0\}$$
$$T_f \in T_{set} = \{1.0, 1.5, 2.0\}$$
$$\Delta_f = 0.25 \quad T_{ph} = 4.0$$

Stage 0. For simplicity, let's fix flare F, which will be corresponding to a precedent. It uniquely localizes an active region on every image within flare prehistory; of course, we suppose tracking AR stay on a visible hemisphere of the Sun throughout the prehistory). We consider only images within flare T_{ph} days prehistory.

Stage 1: base precedent. We choose a magnetogram image ("head magnetogram") and the nearest preceding magnetogram located not closer than Δ_f days. For the head magnetogram we determine the most closely located continuum image ("head continuum"), for which we also find the nearest preceding image located not closer than Δ_f days. Two pairs of images and active region on them we denote as *base precedent*. Time from base precedent to flare F := time from the head magnetogram to flare F.

Stage 2: positive and negative precedents. *Positive precedent (class 1)* is a base precedent, which meets following requirements:
1. Time from the head magnetogram to the flare is not more than T_f.
2. Flare strength is not less than M_f.
3.The nearest flare, which meets 1. and 2., corresponds to the head magnetogram.
Negative precedent (class 0) is a base precedent, which is not a positive one. To a negative precedent we correspond the strongest flare in a T_{ph} time interval from the head magnetogram.
M_f, T_f, Δ_f, T_{ph} — model structural parameters.

Features extraction scheme

1. We fix a head magnetogram and build a precedent according the model (after it we have two pairs of images and a flare).
2. On the nearest to the flare cont-image sunspot groups are localized.
3. If there is a no correspondence between the flare and one of the sunspot groups (using tracking procedure), we start processing another precedent. Else we have the sunspot group and all it's characteristics (including bounding box parameters).
4. Steps 2–4 are perfomed for the second cont-image. After that we are able to extract continuum features from the pair of continuum images.
5. We apply tracking procedure to the found sunspot group (to be precise, to the center of it's bounding box) to locate active regions on both magnetograms and extract from them magnetic features).

The initial localization of sunspot group is made on a continuum image, then we specify active region more exactly on a magnetogram image because direct active region extraction from a magnetogram is rather sophisticated problem.

Continuum features extraction

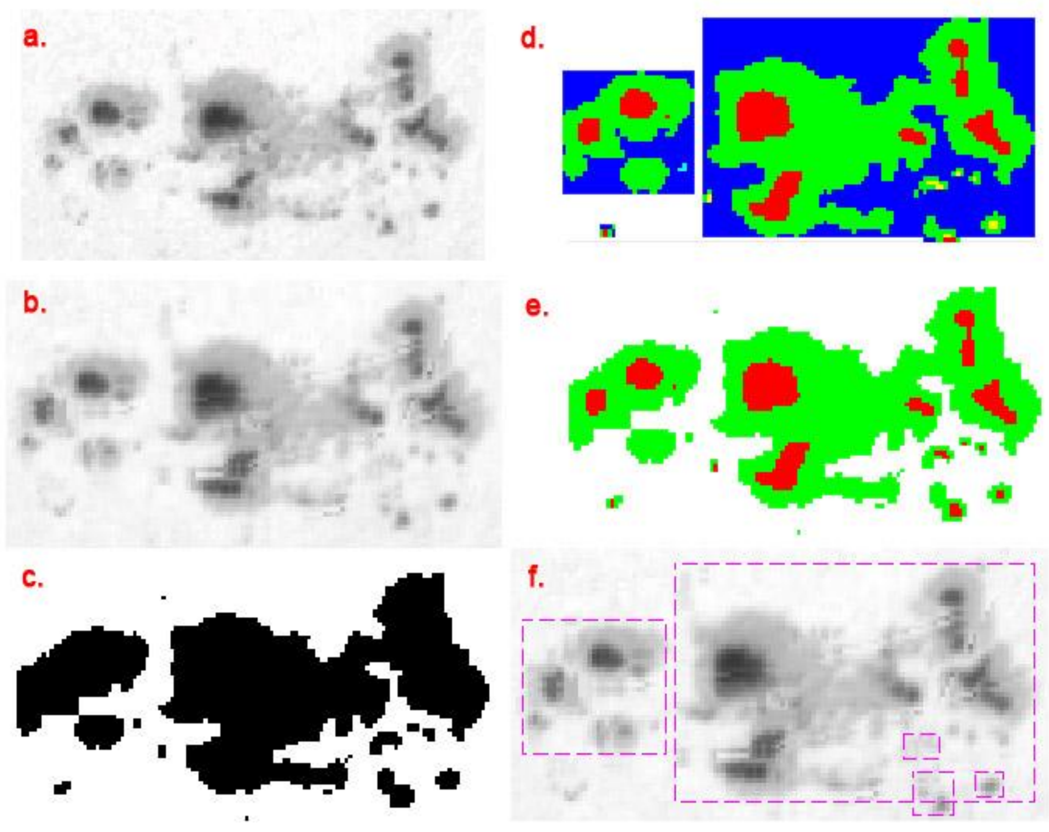


Fig. 1: Vicinity of the sunspot group, localized on cont-image 4240.0003, at different stages of features extraction: (a) — initial, (b) — after spherical correction, (c) — the result of adaptive binarization, (d) — segmented umbra (red) and penumbra (green) using Zharkov et. al method, (e) — segmented umbra and penumbra using our method, (f) — after clustering all neighboring sunpots form a sunspot group.

Main stages: preprocessing (applying spherical correction, filters etc.), adaptive binarization for sunspots localizing, umbra and penumbra segmentation, sunspots clustering.

Extracted continum features: umbra/penumbra square of the sunspot group, it's speed of change, and some others.

Magnetic features extraction

Main stages:
1. Segmentation of a magnetogram (based on variational approximation with the use of global constraints, which give us a possibility to include some physically-driven conditions in our model (i. e. the equality of positive and negative fluxes within AR). See Fig. 2 and 3 (red color corresponds to strong positive fluxes, green — strong negative, blue — neutral).
2. Finding active regions (more exactly, their refining; based on branch and bounds approach to maximizing the functional defined on rectangle) — see Fig. 2.
3. *Neutral line* is a line separating the regions of strong magnetic fields obtained from segmentation with different polarities. Neutral line extraction is applied to a refined AR (Fig. 3).

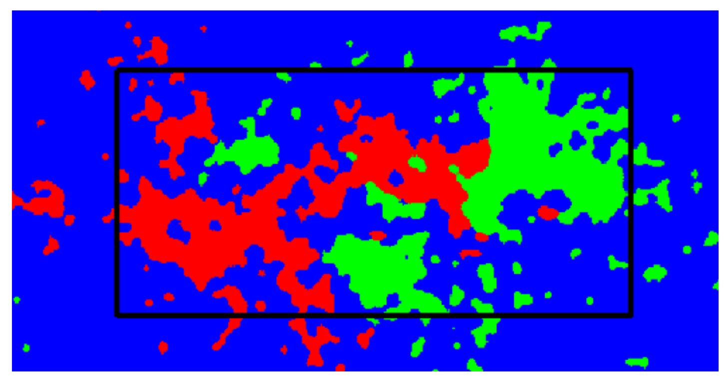


Fig. 2: The localized active region.

Fig. 3: Standard (white) and robust (black) versions of the neutral line.

Extracted magnetic features: positive/negative magnetic flux and it's speed of change, standard/robust neutral line parameters and their speed of change, et al.

Testing and results

M_f \ T_f	1.00	1.50	2.00
C5.0	36.0 **26.9**	36.6 **29.3**	37.0 **30.1**
M1.0	31.7 **22.3**	36.2 **22.2**	36.5 **22.2**
M5.0	26.7 **12.3**	29.3 **13.7**	25.7 **13.7**
X1.0	19.9 **9.4**	17.3 **10.2**	19.0 **11.6**

Table: Average error rates (%) on TEST dataset: for balanced (first number) and unbalanced it's versions.

Testing features:
1. We picked up the most informative features and perform full search over this set; features subset is a structural parameter.
2. Linear SVM with structural parameter C.
3. Dataset partitioning: every calculated dataset is divided into three ~equal parts: train (to learn classificator), test (to find the most optimal joint configuration of the features subset and SVM parameter), and TEST (to get testing results).
4. After multiple partitioning and averaging obtained results we get optimal configuration of the structural parameters and final results (see Table).

Conclusion

In this research a new automated hybrid method for short-term flare forecasting is introduced.
At the initial stage we created a respectable flare and images bases covering1996–2009 years.
Further, we worked out simple and efficient parameterized precedent model, which turned ourprediction problem into two-class classification problem, and developed machine learning-based procedures for features extraction from SOHO/MDI both magnetograms and continuum images. For minimization of area distortions due plane projection of a semisphere spherical correction procedure was proposed.
Finally, a testing system for obtaining unbiased test error rates was implemented. The accuracy of built system depends on chosen precedent model configuration: 63% to 82% on class-balanced data (maximum worst-case), **73% to 90% on real data**.
In the nearest future we intend to incorporate in our method some additional physically-driven features, include SHO/HMI images support, and build fully-automated web-compliant prediction system. You can vizit http://workshop18.org to get full research information.