Solar flare forecasting using sunspot-groups classification and photospheric magnetic parameters

Yuan Yuan[1,2], Frank Y. Shih[2], Ju Jing[1] and Haimin Wang[1]

[1]Space Weather Research Lab, New Jersey Institute of Technology,
323 MLK Blvd, Newark, New Jersey, United States
email: yy46@njit.edu

[2]Computer Vision Lab, New Jersey Institute of Technology,
323 MLK Blvd, Newark, New Jersey, United States

Abstract. In this paper, we investigate whether incorporating sunspot-groups classification information would further improve the performance of our previous logistic regression based solar flare forecasting method, which uses only line-of-sight photospheric magnetic parameters. A dataset containing 4913 samples from the year 2000 to 2005 is constructed, in which 2721 samples from the year 2000, 2002 and 2004 are used as a training set, and the remaining 2192 samples from the year 2001, 2003 and 2005 are used as a testing set. Experimental results show that sunspot-groups classification combined with total gradient on the strong gradient polarity neutral line achieve the highest forecasting accuracy and thus it testifies sunspot-groups classification does help in solar flare forecasting.

Keywords. Sun: flares, sun: magnetic fields, sunspots

1. Introduction

Sunspot-groups characteristics have long been used in solar flare forecasting and still being used extensively. Contarino *et al.* (2009) studied sunspot-groups parameters (i.e., Zrich class, magnetic configuration, area, morphology of the penumbra), and then performed a flare forecasting campaign based on the results. They claimed that the results obtained by comparing the flare forecasting probability with the number of flares that have actually occurred are quite encouraging. Kasper & Balasubramaniam (2010) found out that the penumbral area, umbral area and irradiance showed promise as possible parameters for predicting solar flares, particularly M-class flares. Qahwaji & Colak (2007) compare the performances of several machine learning algorithm on flare forecasting using classification of sunspot groups and solar cycle data. They found out that Support Vector Machines provide the best performance for predicting whether a classified sunspot group is going to flare.

On the other hand, photospheric magnetic parameters derived form line-of-sight magnetograms are becoming more and more popular in solar flare forecasting. Jing *et al.* (2006) studied the mean value of spatial magnetic gradients at strong-gradient magnetic neutral lines, the length of strong-gradient magnetic neutral lines and the total magnetic energy. They found out there exist statistical correlations between the three parameters of magnetic fields and the flare productivity of solar active regions. Yuan *et al.* (2010) proposed a cascading forecasting approach using total unsigned magnetic flux, length of the strong-gradient magnetic polarity inversion line, and total magnetic energy dissipation. Experimental results show that photospheric parameters is indeed can be used a precursor for solar flares forecasting.

(a) A cropped MDI magnetogram containing active regions

(b) A map indicating the regions used to calculate total unsigned magnetic flux with white color

(c) Magnetic polarity inversion line overplotted on a smoothed MDI magnetogram

(d) Magnetic polarity inversion line with strong gradient overplotted on a smoothed MDI magnetogram
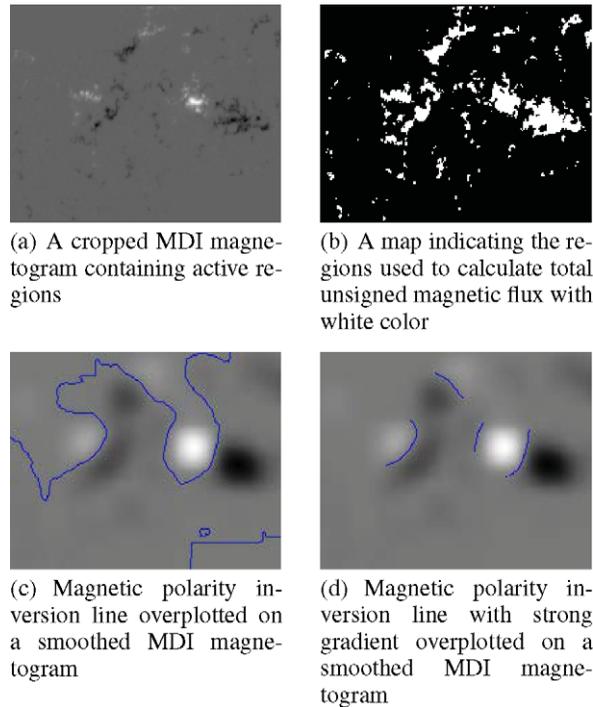
Figure 1. Illustration of the calculation of total unsigned magnetic flux and total unsigned gradient on strong gradient magnetic polarity inversion line

In this study, aiming to improve the solar flare forecasting performance in our previous study (Song *et al.* 2009) ,we use both sunspot-groups classification and photoshperic magnetic parameters. We view the solar flare forecasting as a classification problem in machine learning field. To forecasting a flare event is usually converted to classify one sample as a flaring sample or a non-flaring sample. Previously, researchers usually adopt support vector machines, such as Qahwaji & Colak (2007), or neural networks, such as Wang *et al.* (2008). The outputs of support vector machines and neural networks are binary labels indicating flaring or nonflaring. However, people sometimes prefer to get a probability instead of a binary label, just like what people get from daily weather reports. In our previous studies by Song *et al.* (2009) and Yuan *et al.* (2010), we have shown that logistic regression, which is a statistical learning method for probability estimation, can be used for flare forecasting. In this paper, the solar flare forecasting is regarded as a classification problem in machine learning field, i.e., flaring population vs. non-flaring population.

2. Dataset

The dataset used in our experiments includes 4913 samples from the year 2000 to 2005, in which 2721 samples from the year 2000, 2002 and 2004 are used as the training set, and the remaining 2192 samples from the year 2001, 2003 and 2005 are used as the testing set. Each sample is a pair of values describing the properties of an active region. A sample composed of a label indicating whether the active region produces a flare or not, a label indicating the classification of the sunspot-groups within the active region, a number indicating the total unsigned magnetic flux within the active region,
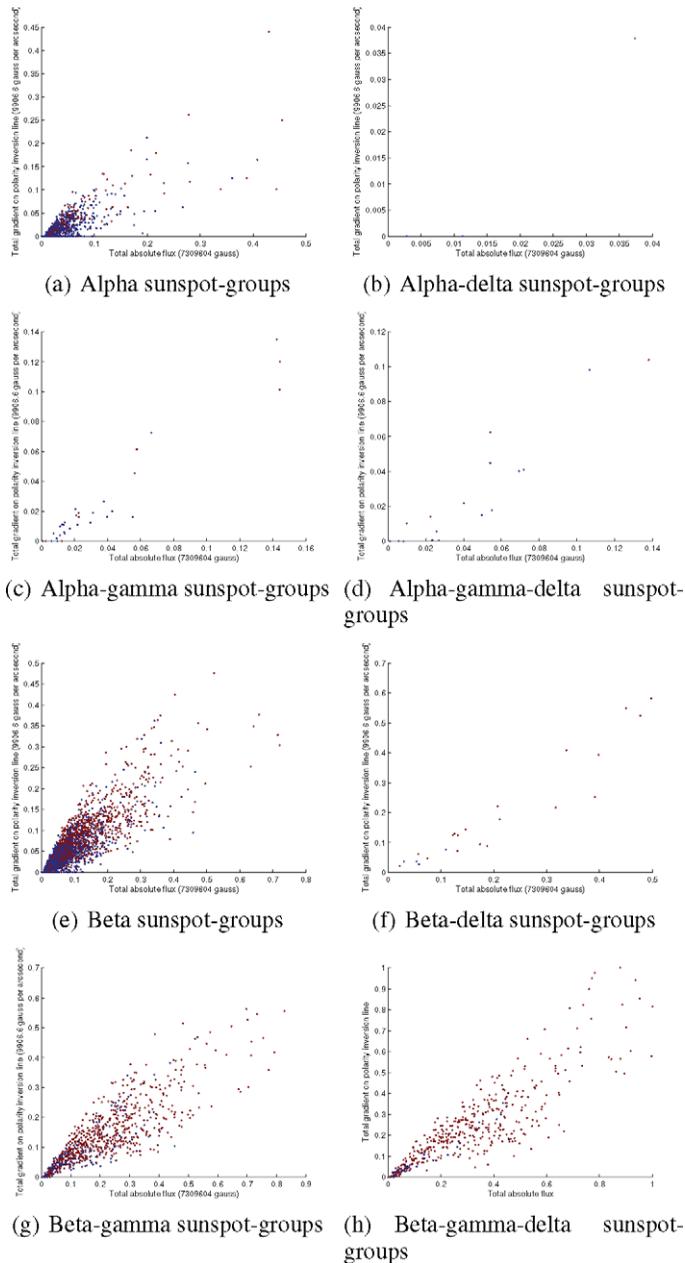
Figure 2. Distribution of dataset with respect to different sunspot-groups classifications.

and a number indicating the total gradient of the strong gradient polarity neutral line. Figure 2 illustrates the distribution of our dataset with respect to different sunspot-groups classifications.

Total unsigned magnetic flux is the integration of pixel intensity over the strong magnetic flux region of an active region. In this study, we define strong flux region as the region composed of pixels with intensity greater than median value plus 80 gauss, and pixels with intensity less than median value minus 80 gauss. Total unsigned magnetic

flux T_{flux} can be calculated as following:

$$T_{flux} = \int_{B_Z(x,y) \leqslant median-80 \ or \ B_Z(x,y) \geqslant median+80} |B_Z(x,y)| \, dx dy \qquad (2.1)$$

where $B_Z(x,y)$ is the intensity of a pixel at location (x,y) of a MDI magnetogram.

The total gradient of the strong gradient polarity neutral line is the integration of the gradient over the pixels whose intensities are zeros and their gradient is greater than a threshold (here we choose 5). The total gradient of the strong gradient polarity neutral line T_{grad} can be calculated as following;

$$T_{grad} = \int_{B_Z(x,y) \equiv 0 \ ,m>5} \sqrt{\left(\frac{\partial B_Z(x,y)}{\partial x}\right)^2 + \left(\frac{\partial B_Z(x,y)}{\partial y}\right)^2} \, dx dy \qquad (2.2)$$

where

$$m = \sqrt{\left(\frac{\partial B_Z(x,y)}{\partial x}\right)^2 + \left(\frac{\partial B_Z(x,y)}{\partial y}\right)^2} > 5 \qquad (2.3)$$

Figure 1 contains one sample illustrating the calculation of total unsigned magnetic flux and total unsigned gradient on strong gradient magnetic polarity inversion line. To calculate total unsigned magnetic flux, a binary mask (illustrated as fig. 1(b)) is generated which indicating the regions where magnetic flux is greater than median plus 80 or less than median minus 80. And then the summation of pixel values inside those regions are figured out as total unsigned magnetic flux T_{flux}. To figure out magnetic polarity inversion line, a MDI magnetogram is firstly smoothed with a Gaussian filter with the standard deviation 10 and of size 30 by 30. And then contour lines at height zeros are find out(illustrated as fig. 1(c)). At last, the contour lines with strong gradient is kept (illustrated as fig. 1(d)). The summation of the gradient on strong gradient magnetic polarity inversion line are figured out as the total gradient of the strong gradient polarity neutral line T_{grad}.

	1	2	3	4	5	6	7
■ Accuracy	0.6939	0.7163	0.6586	0.7211	0.7060	0.7299	0.7266
■ Recall	0.5630	0.5882	0.3659	0.5943	0.5653	0.5966	0.5959
■ Precision	0.7385	0.7677	0.8287	0.7734	0.7621	0.7905	0.7839
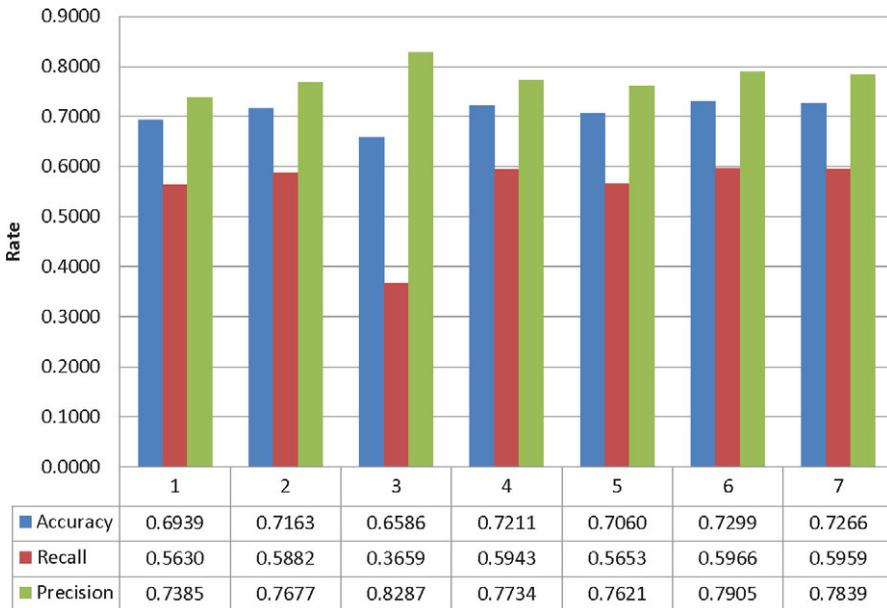
Figure 3. Performance evaluation

3. Experimental Results

Figure 3 illustrates the rate of accuracy, recall and precision of the solar flares forecasting method with seven different combinations of input parameters: 1. use T_{flux} alone, 2. use T_{grad} alone, 3. use sunspot-groups classification alone, 4. use T_{flux} and T_{grad} , 5. use T_{flux} and sunspot-groups classification, 6. use T_{grad} and sunspot-groups classification, 7. use T_{flux}, T_{grad} and sunspot-groups classification.

From figure 3, we can see that solar flares forecasting using T_{grad} and sunspot-groups classification achieves best accuracy and recall. Solar flare forecasting using sunspot-groups classification alone achieve best precision. The average performance (measured by accuracy, recall and precision) of solar flare forecasting using T_{grad} and sunspot-groups classification is the best.

References

Contarino, L., Zuccarello, F., Romano, P., Spadaro, D., Guglielmino, S. L., & Battiato, V. 2009, *Acta Geophysica*, 57, 52

Jing, J., Song, H., Abramenko, V., Tan, C., & Wang, H. 2006, *Astrophys. J.*, 644, 1273

Kasper, D. & Balasubramaniam, K. S. 2010, *Bulletin of the American Astronomical Society*, 41, 291

Qahwaji, R. & Colak, T. 2007, *Solar Phys.*, 241, 195

Song, H., Tan, C., Jing, J., Wang, H., Yurchyshyn, V., & Abramenko, V. 2009, *Solar Phys.*, 254, 101

Wang, H. N., Cui, Y. M., Li, R., Zhang, L. Y., & Han, H. 2008, *Advances in Space Research*, 42, 1464

Yuan, Y., Shih, F. Y., Jing, J., & Wang, H.-M. 2010, *Research in Astronomy and Astrophysics*, 10, 785