Space Weather of the Heliosphere: Processes and Forecasts
Proceedings IAU Symposium No. 335, 2017
C. Foullon & O.E. Malandraki, eds.

doi:10.1017/S1743921317008134

Comparison of Predictive Efficiency of Topological Descriptors and SHARP in Solar Flares Forecasting

Irina Knyazeva[1,2], **Fedor Urtiev**[1] **and Nikolay Makarenko**[1]

[1]Pulkovo Observatory, Saint-Petersburg
email: `iknyazeva@gmail.com`

[2]Saint-Petersburg State University
Saint-Petersburg, Russia
email: `iknyazeva@gmail.com`

Abstract. In the current paper, we investigate topological invariants, calculated by HMI LOS magnetograms as complexity descriptors of solar magnetic fields. We compared them with the physical parameters provided by the Space-weather HMI Active Region Patches (SHARP). We have repeated forecasting experiment of Stanford Solar Observatories Group with the same positive and negative active region patches database, but replace SHARP parameters with topological invariants of corresponding LOS magnetograms. The classification results turned out practically identical to those obtained by the Stanford Solar Observatory group. This means that using LOS magnetograms retains enough complexity for magnetic field description.

Keywords. methods: numerical, methods: statistical,Sun: magnetic fields, Sun: flares

1. Introduction

The growth of the quantity and quality of the solar magnetic field observational data and development of machine learning methods stimulate improved forecasting techniques for the solar flare forecasting. Most of techniques are based on the complexity of the photospheric magnetic field of the Sun's active regions. There are a large number of different characteristics that can be used for magnetic field complexity description. Due to many empirical assumptions during their calculation, they are hardly reproducible. For HMI/SDO vector magnetograms an automated active region tracking system exist called Spaceweather HMI Active Region Patch (SHARP). For each active region, key features called SHARP parameters were calculated and are available online. Computation of these features is based on SDO vector magnetograms. The line-of-sight component of the magnetic field (HMI/SDO LOS) includes only one projection of three-dimensional vector but is measured with higher quality. It is impossible to compute physically based parameters with only one component of a magnetic field but a large number of different proxy are known. We propose to use topological invariants based on a topological data analysis approach. They could be computed directly from the data without any assumptions and parameters. Using topological invariants as complexity description (in terms of critical points) is used by researchers in random field analysis Adler (2010). In this paper, we compared the efficiency of these features computed only by the LOS magnetograms with the SHARP parameters as solar flare forecasting model.

2. Methods

We perform a classification of positive and negative events that correspond with flaring and non-flaring active regions on the sun with two different sets of features. The first set of parameters are the physical parameters provided by the Space-weather HMI Active Region Patches (SHARP) calculated by vector magnetic data (Bobra *et al.* 2014). The second set of parameters is topological features calculated by corresponding LOS data. The number of possible combinations of flaring and especially non-flaring regions are very large. We reproduce exactly the same experiment made by Stanford Solar Observatories Group published openly with the data and active regions timing. With these two sets of features, we trained and tested a support vector machine classifier and evaluated its performance.

Topological features were calculated with the topological data analysis approach, see Edelsbrunner & Harer (2010). The main idea of TDA is to describe the shape of the data in term of some topological invariant and after that use these invariants as features. For random field, the basic research topic is understanding the distribution of geometric and topological features generated by a stochastic process. We could consider each magnetogram as a realization of some stochastic process drawn from some distribution, see Adler (2010). The structure of the random field is determined by the content of the local neighborhoods for the maxima and minima points, namely how many and at which level peaks appear that are close to the given maxima or minima. Also, we would like to know up to which level field maxima and minima are isolated in some local neighborhood. We can measure the lifetime of each isolated peak as the length of the interval during which it is separated from others. It is useful to draw it on the plane using the beginning and the end of the barcode as point coordinates. As the result, we obtain a set of points that lie above the diagonal that corresponds to barcodes of zero length. This graph is called a persistence diagram (PD). For each magnetogram we received two persistence diagram, roughly speaking one of them corresponds to minima neighborhood, the other to maxima neighborhood. In this work, we have used our own program with algorithm fully described in the paper Makarenko *et al.* (2014) and CAPD RedHom (Juda & Mrozek 2014) for PD estimation.

3. Results

The database was formed as follows. A positive event is an active region that flares with a peak magnitude above the M1.0 level, as defined by the GOES database. A negative event would be an active region that does not have such an event within a 24-hour time span. Two sets of features were formed. The first set was based on SHARP parameters that were taken from the page of the Stanford Solar Observatories Group. For the second feature set, PD for each LOS magnetogram were computed (see examples of AR and corresponding PD in Figures 1-4. For feature vector construction descriptive statistics of these PD were used. After that, these two data sets were run through a SVM. The performance of the classifier was esimated by using 6-fold cross validation and represented as ROC curve for each fold. The ROC curve is created by plotting the true positive rate or truly predicted flares events (TPR) against the false positive rate or predict flare when it was not (FPR) at various threshold settings. In the case of perfect classification ROC values equal to 1 in all points except 0. The closer the curve is to the diagonal the worse the forecasting ability. The cumulative area under the curve (AUC) measure summarize the quality of the ROC.

The result of classification for first (see Figure 5) and second set (see Figure 6) turned out to be practically the same.

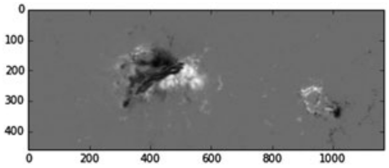

Figure 1. Positive AR

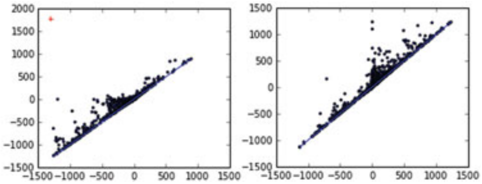

Figure 2. PD for AR of Figure 1

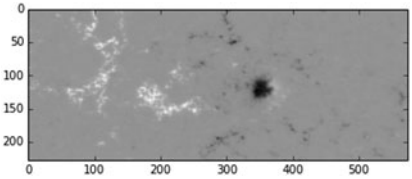

Figure 3. Negative AR

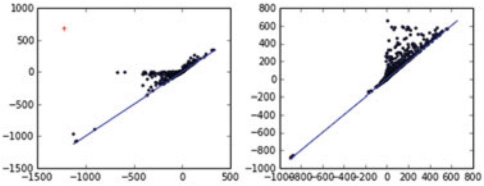

Figure 4. PD for AR of Figure 3

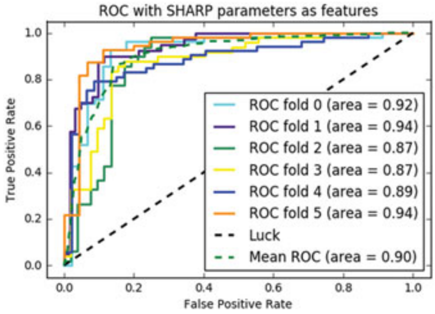

Figure 5. ROC curve based on SHARP parameters

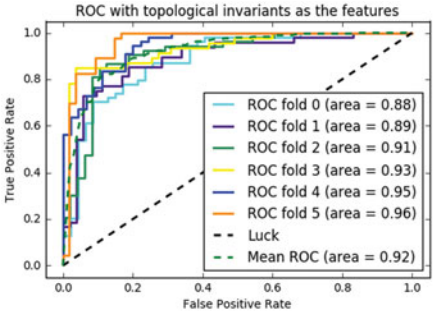

Figure 6. ROC curve based on topology

4. Conclusions and Forthcoming Research

We find that the forecasting efficiency of topological features computed only by the LOS magnetograms is practically the same as for vector magnetic data parameters, therefore LOS data could be useful in forecasting systems. Using distances between PD evolution of active region could be included. Also, PD depends only on the inner structure of the magnetogram and does not depend on absolute values, so potentially SOHO/MDI or other magnetic field data could be combined in one system.

Acknowledgments

This work was supported by RFBR grant number 16-32-00251

References

Bobra, M. G., Sun, X., Hoeksema, J. T., Turmon, M., Liu, Y., Hayashi, K., Barnes, G. & Leka K. D. 2014, *SoPh*, 289, 3549

Makarenko N. G., Malkova D. B., Machin M. L., Knyazeva I. S. & Makarenko I. N., 2014, *Fundam. Prikl. Mat. (Russian), translation in J. Math. Sci.* 203, 806

Edelsbrunner, H. & Harer, J. 2010, Computational topology: an introduction. *AMS*

Adler, Robert J. 2010, The geometry of random fields, volume 62 of Classics in Applied Mathematics. *SIAM*, Philadelphia, PA

Juda M. & Mrozek M. 2014, *ICMS* 160–166