ВАК 2024 MODERN ASTRONOMY: FROM THE EARLY UNIVERSE TO EXOPLANETS AND BLACK HOLES
300 years Russian Academy of Sciences

Recognition of solar flares in the microwave range of observations using machine learning

J. Shamsutdinova and L. Kashapova

Institute of Solar-Terrestrial Physics SB RAS, Irkutsk, 69167 Russia

Abstract. We present the results of testing machine learning methods for the recognition of solar flares observed in the microwave range. The catalogue of observations by the Nobeyama spectropolarimeters (NoRP) was used for both training and analysis of method effectiveness. The input data are one-dimensional temporal profiles with a frequency of 9.4 GHz, which are mostly associated with the emission of an optically thin source and the non-thermal emission from accelerated electrons. The input dataset consisted of 100 events and included temporal profiles with simple and complex structures. The aim of recognition is to distinguish between the temporal profiles of “classical” or simple-looking shapes, and “complex” ones. We compared the classification results provided by the correlation coefficient application with the results obtained using the Support Vector Machine (SVM) method with different parameters. True Positive (TP) and True Negative (TN) are values describing the number of events which the model correctly recognized, in our case, “classical” and “complex” profiles. False Postive (FP) and False Negative (FN) indicate incorrect model predictions of “classical” or “complex” profiles. The number of correctly recognized events using the correlation coefficient method was 75, compared to 90 events obtained using the SVM method. The number of false predicted events resulting from the application of the correlation coefficient was 25, versus 10 by SVM application. These results indicate the advantage of the SVM technique over the correlation coefficient approach. The SVM method allows us to process datasets using different functions as kernels - linear, polynomial and Radial Basis Function (RBF). In our study, we found that the most accurate result was provided by the RBF kernel with a gamma hyperparameter of 1.

Keywords: Sun: solar flare, radio radiation; methods: data analysis

DOI: 10.26119/VAK2024.113

1 Introduction

Nowadays, it becomes more and more important to identify events in information streams obtained from new astronomical instruments. Machine learning is a useful tool for this purpose. However, in the field of solar physics, these methods have been used to predict solar flares (Asensio Ramos et al. 2023) or identify different types of radio bursts in dynamic spectra (Xu et al. 2019). Machine learning methods have not been used for the detection and identification of solar flares in the microwave range. This paper presents the first results of testing the Support Vector Machine (SVM) for detecting solar flares using a classical temporal profile shape.

2 Dataset and the Support Vector Machine method

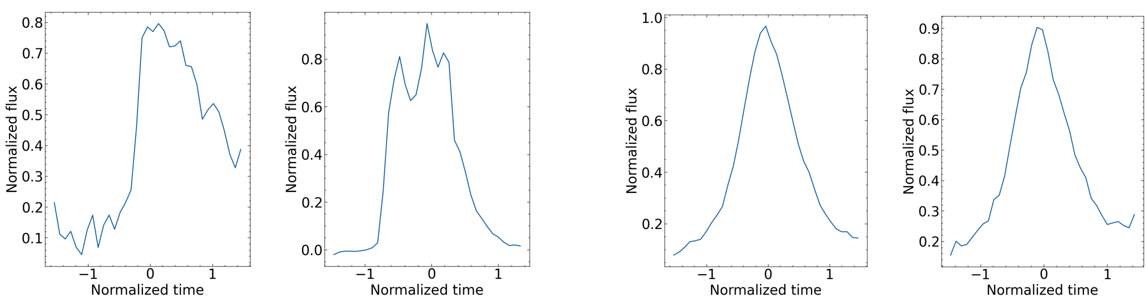

Fig. 1. Example of dataset after primary processing. Left panel: Temporal profiles with complex structure (class=0). Right panel: Temporal profiles with simple structure (class=1).

The dataset for training and testing was formed from the temporal profiles of microwave emission from the database of events observed by the Nobeyama spectropolarimeters. (NoRP) (Torii et al. 1979). The NoRP receives the intensity (I) and circular polarization (V) at frequencies of 1, 2, 4, 9.4, 17, and 35 GHz with a time resolution up to 0.1 second. We selected 100 events from the NoRP catalogue during periods of high solar activity, including 2001 and the years 2012–2014[1]. We selected intensity temporal profiles at a frequency of 9.4 GHz for recognition. The choice of frequency was determined by two reasons. The first one is the emission of the most solar flares at this frequency, which forms in the region of an optically thin source and is therefore free from additional effects. The other reason is that the Siberian Radioheliograph observes at the same frequency, and the model trained on NoRP

[1] https://solar.nro.nao.ac.jp/norp/html/event/

data can be applied to SRH data at this frequency. Before putting the data into the model, they were preprocessed. We normalized every temporal profile to its maximum intensity. The moment of the maximum was used as a zero point on the time scale. Therefore, the time during the rise phase has negative values, while the time during the decay phase has positive values. To get the same time scale for events with different durations, time was normalized using units defined by the time it took for the intensity to decrease to half of the maximum value for each time profile. The described methodology is used to reconstruct an average profile in Davenport et al. (2014). Each array contains 35 elements, which corresponds to the number of features. For training and testing, we used a dataset consisting of 100 arrays (see examples in Fig. 1). 43 profiles have a complex structure that may contain multiple peaks and/or noise that dominates the signal (class 0). The remaining 57 profiles have a classic structure (class 1).

Support Vector Machine (SVM) is a supervised machine learning algorithm used for both classification and regression. The main objective of the SVM algorithm is to find an optimal hyperplane in a high-dimensional space that can separate data points into different classes in the feature space. The hyperplane tries to maximize the margin between the closest points (support vectors) of different classes.

When the number of input features is greater than three, SVM uses kernels to transform the hyperplane into a nonlinear one. There are linear, polynomial, and Radial Basis Function (RFB) kernel functions. RFB has a gamma hyperparameter, which determines the impact of support vectors on the data. The C hyperparameter provides a correct classification of training examples against maximization of the decision function's margin. We used the GridSearch method to determine the optimal parameters for the model with different kernel functions. We optimized the model by using GridSearch to select hyperparameters based on the accuracy achieved on the test set. The optimal hyperparameters for the models are presented in the first row of Table 1. The scikit-learn library for Python was used to train the model [2]. SVM is typically used for binary classification problems and in cases where the data is not linearly separable and the cross-correlation method may not be applicable to the recognition problem.

3 Results of classification and evaluation metrics

Another way to recognize the shape of a temporal profile is to use cross-correlation between the training profiles and the analyzed one. This method is sensitive to stationary and linear relationships in the dataset. The number of correctly recognized

[2] https://scikit-learn.org

events using the correlation coefficient method for level 0.8 was 75 while compared to 90 events obtained using the SVM method. We used metrics such as precision, recall and F1 measure to evaluate the model's performance (see the example Wang et al. 2024).Precision measures how many observations predicted to be positive are actually positive. The recall measures how many positive observations we classified as positive, out of the total number of observations. F1 is the harmonic mean between precision and recall. The model is more accurate the closer F1 is to 1. The results of testing models with different kernel functions are presented in Table 1. It can be seen that the most accurate information was obtained from the RFB kernel with the gamma hyperparameter equal to 1. However, the obtained results have to be tested on a larger dataset over a wider frequency range.

Table 1. The result of testing SVM models for recognizing time profiles

	Linear (C=1, γ=1)			RBF (C=10, γ=1)			Poly (C=0.1, γ=0.1)		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Simple flare	0.73	0.92	0.81	0.79	0.85	0.81	0.67	0.50	0.77
Complex flare	0.88	0.62	0.73	0.82	0.75	0.78	0.85	0.92	0.63

Acknowledgements. Authors thank the team operating at the Nobeyama for the Nobeyama Radio Polarimeter database.

Funding

The research was funded by the Russian Science Foundation grant No. 24-22-00315, (https://rscf.ru/project/24-22-00315/).

References

Asensio Ramos A., Cheung M.C.M., Chifu I., et al., 2023, Living Reviews in Solar Physics, 20, 1, id. 4

Davenport J.R.A., Hawley S.L., Hebb L., et al., 2014, Astrophysical Journal, 797, 2, id. 122

Torii C., Tsukiji Y., Kobayashi S., et al., 1979, Proceedings of the Research Institute of Atmospherics, Nagoya University, 26, p. 129

Wang Y.H., Feng S.W., Du Q.F., et al., 2024, Solar Physics, 299, 4, id. 54

Xu L., Yan Y.-H., Yu X.-X., et al., 2019, Research in Astronomy and Astrophysics, 19, 9, id. 135