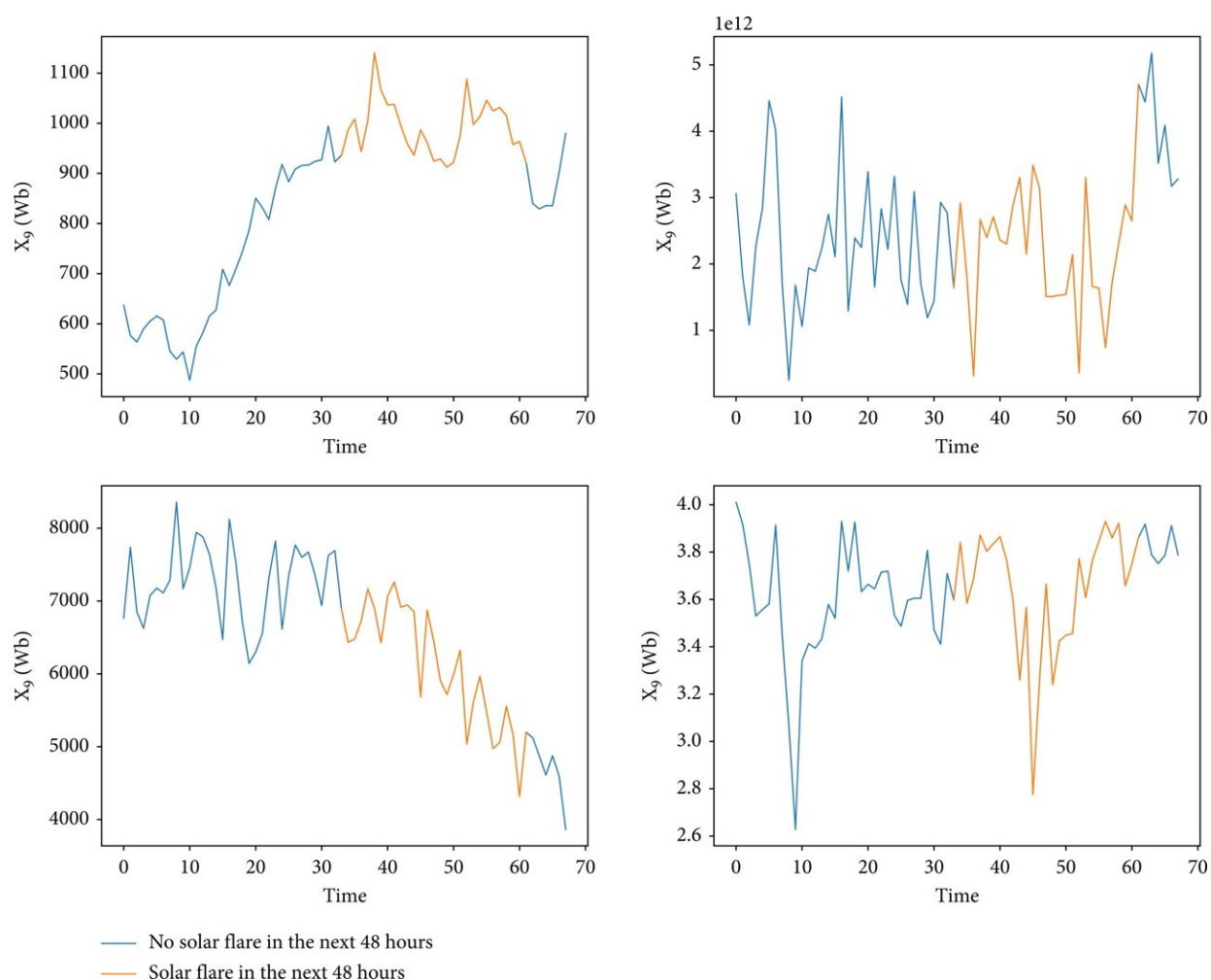


How a scientist established a two-stage solar flare early warning system

August 19 2022



The visualization of four features during the existence of an active region. The x-axis represents time and its unit is a sample, where “0” represents the start time of an active region, and the time gap between adjacent times is 1.5 h. The y-axis represents the value of a feature. The blue lines indicate that there is no solar

flare in the next 48 hours, and the yellow lines are the opposite. Credit: *Space: Science & Technology*

Solar flares are solar storm events driven by magnetic field in the solar activity area. When this flare radiation comes to the Earth's vicinity, the photo-ionization increases the electron density in the D-layer of the ionosphere, causing absorption of high-frequency radio communication, scintillation of satellite communication, and enhanced background noise interference with radar.

Statistics and experience show that the larger the flare, the more likely it is to be accompanied by other solar outbursts such as a solar proton event, and the more severe the effects on the Earth, thus affecting spaceflight, communication, navigation, power transmission and other technological systems.

Providing forecast information on the likelihood and intensity of flare outbreaks is an important element at the beginning of operational space weather forecasting. The modeling study of solar flare forecasting is a necessary part of accurate flare forecasting and has important application value. In a research paper recently published in *Space: Science & Technology*, Hong Chen from College of Science, Huazhong Agricultural University, combined the k-means clustering algorithm and several CNN models to build a warning system that can predict whether a solar flare would occur in the next 48 hours.

First, the author introduced the data used in the paper and analyzed them from the statistical point of view to provide a basis for the design of the solar flare warning system. To reduce the projection effect, the center of the active region located within $\pm 30^\circ$ of the solar disk center was selected. After that, the author labeled the data according to the solar

flare data provided by NOAA, including the start and end times of the flares, the number of the active region, the magnitude of the flares, etc.

There was a serious imbalance between the number of positive and negative samples in the dataset. To alleviate the imbalance of positive and negative samples, a principle was found to select the events which have positive samples as much as possible. The author visualized the probability density distribution of each feature in all negative samples and all positive samples. It could be easily found that the probability density distributions of the negative samples were all negatively skewed distributions and the characteristics of positive samples were generally larger than those of negative samples. Thus, it was possible to filter out events with positive samples by the feature values of each event.

Afterwards, the author built the whole pipeline with a method containing the following two steps: data preprocessing and model training. To conduct data preprocessing, K-means, an unsupervised clustering method, was used to cluster events to decrease events that only include negative samples as much as possible.

After k-means clustering, all events were divided into three categories, namely category A, category B, and category C. The author found that the ratio of positive samples in category C is 0.340633 which is much larger than that of the whole dataset. Therefore, only the data in category C were chosen as input data on the next stage of algorithm.

In the 2nd stage, the neural networks the author used were Resnet18, Resnet34 and Xception, which are commonly used in deep learning. Three-fourths of samples in category C were randomly chosen. In each event were training data for the neural network models and the rest of the samples were regarded as validation data in the process of training model.

To avoid the influence of dimension, the author also standardized the original data. The standardization method was different from those commonly used. According to the standardization calculation formula, if the label of a sample was predicted to be 1 by the neural network, this sample was regarded as a signal of solar flare which would occur in the next 48 hours. But if it is predicted to be 0, the probability of occurring solar flare in the next 48 hours would be so small that could be ignored.

Then, the author conducted experiments and discussed the results. The author first gave an introduction of experimental setting and then conducted several ablation experiments and comparisons with different models to verify the improvement of k-means clustering algorithm and boosting strategy. Besides, the author also made comparisons between the method used in the experiment and other 13 binary classification algorithms commonly used to present its prediction performance.

The experimental results showed that the prediction performance of the model which integrated several [neural networks](#) was better than the one of a single convolutional neural network. Finally, the prediction results of Resnet18, Resnet34, and Xception were combined by boosting strategy. For all networks, recall may be unchanged or even reduced greatly after clustering. However, precision was bound to increase significantly.

After clustering, although the positive sample rate would be greatly improved, from 5% to 34%, nearly 40% of the information of positive samples would also be lost. The author thought this was the main reason why recall remained unchanged or even decreased. It also meant that the number of positive samples predicted in the experiment was less than the one without clustering, but the probability that a predicted positive sample was a true positive was higher.

In contrast with the phenomenon that the prediction performance of

other binary classification methods was decreasing or even very poor after clustering, the performance of the author's method improved by more than 9% after clustering. In conclusion, the two-stage solar flare early warning system consisted of an unsupervised clustering algorithm (k-means) and several CNN models, where the former was to increase the positive sample rate, and the latter integrated the prediction results of the CNN models to improve the prediction performance.

The results of the experiment proved the effectiveness of the method.

More information: Jun Chen et al, Two-Stage Solar Flare Forecasting Based on Convolutional Neural Networks, *Space: Science & Technology* (2022). [DOI: 10.34133/2022/9761567](https://doi.org/10.34133/2022/9761567)

Provided by Beijing Institute of Technology Press Co., Ltd

Citation: How a scientist established a two-stage solar flare early warning system (2022, August 19) retrieved 27 April 2024 from <https://phys.org/news/2022-08-scientist-two-stage-solar-flare-early.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.
