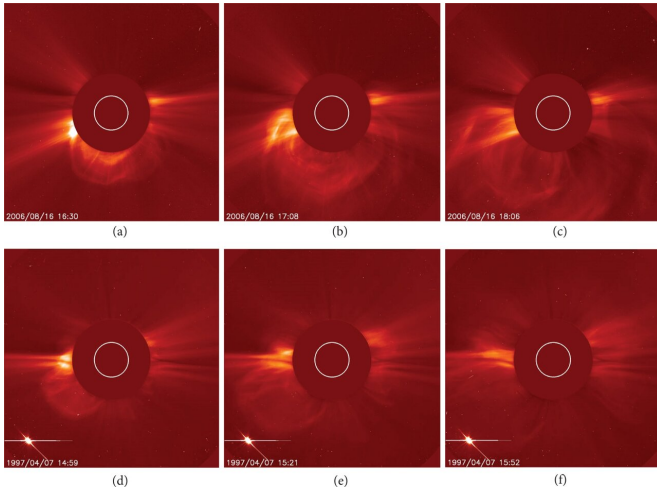


# How scientist applied the recommendation algorithm to anticipate CMEs' arrival times

11 July 2022



Top: From left to right, snapshots of the CME event occurring at 2006 August 16 16 : 30 UT. Credit: SOHO LASCO C2. Bottom: From left to right, snapshots of the CME event occurring at 1997 April 7 14 : 27 UT. Credit: *Space: Science & Technology*

Coronal mass ejections (CMEs) are eruptive solar events. They are often associated with solar flares and filaments. CMEs can cause space weather events such as geomagnetic storms, high energy electron storms, hot plasma injection, ionospheric storms, and increased density in the upper atmosphere.

Large CME events can impact communications, navigation systems, aviation activities, and even power grids. To avoid potential damage and asset loss, there is a need to accurately predict the arrival of the CMEs in two parts. Will the CME "hit" or "miss" the earth? If the prediction is "hit", then the next question is what is the expected arrival time of the CME?

In a research paper recently published in *Space: Science & Technology*, Yurong Shi from National Space Science Center, Chinese Academy of

Sciences, applied the recommendation algorithm, which could be used to recommend the similar historical CME event for forecasters, to anticipate CMEs' arrival time and proved that recommendation algorithm and logistic regression could act together to provide forecasters an option to improve the prediction results.

Firstly, data and methodology were prepared. The author selected samples from a total of 30,321 CME events which were collected from the SOHO/LASCO CME catalog, from 1996 to 2020. Oversampling was used to solve the unbalanced data and have obtained 181 positive samples (CMEs that reached the earth) and 3486 negative samples (CMEs that did not reach the earth).

Besides, 8 characteristic parameters are gathered by characteristic parameters selection, including angular width, central position angle (CPA), measurement position angle (MPA), linear velocity, initial velocity, final velocity, the velocity at 20 solar radii, mass. A complete and unified dimensionless data set of the 8 characteristic parameters was set up and ready to facilitate the development of the prediction model. Furthermore, to search for the historical event most similar to the specified CME event, the authors adopt two distances commonly used in [machine learning](#) and computer [artificial intelligence](#): cosine distance and Euclidean distance which were both proved performing well during the experiment.

Afterwards, the experiment, a controlled trial, was designed. The first stage is the data sampling. A total of 3,667 samples including 8 characteristic parameters are randomly divided into two equal subgroups. One (1,833 samples) is for weight training and the other (1,834 samples) is for the subsequent recommendation test. During [weight training](#) stage, the author used 1,466 training samples served as the training set to train weights following both the logistic regression procedure and the recommendation algorithm, while the rest as

(367 samples) the validation set.

Briefly, a total of 6 experiments are conducted to train weights, and hence, 6 sets of weight coefficients are obtained with 4 from the logistic regression algorithm and 2 from the recommendation algorithm. Two logistic regression frameworks were adopted for comparison. One was the logit function provided in the Python-based statsmodels module and referred to as "sm.logit." The other also Python-based was the LogisticRegression classifier provided in the scikit-learn (sklearn) library and referred to as "sk.LR."

Comparing all models, the sm.logit model performed the best in both the validation set and the test set. It was appropriate to choose the weights of sm.logit as the optimal weights in the following stage in this particular work. Besides, it can be seen that using the recommendation algorithms to train the weights of characteristic parameters was very time-consuming, but it was easier to obtain the weights by logistic regression. Therefore, a new attempt was to apply the weights obtained by the logistic regression to the recommendation algorithm. The feasibility of such operation was tested during the final stage, recommendation test stage.

In summary, the author first calculated the weights of the characteristic parameters of CMEs based on logistic regression and then fed them into the recommendation algorithm to provide the most similar historical events as a reference for CMEs effectiveness forecasting. It can be found that in each skill score the model applying the weights of logistic regression to the recommendation algorithm was better than that using recommendation algorithm alone, so this hybrid model was feasible. Such a treatment avoided training the recommendation weights to save time and computer resources.

At present, applying the recommendation algorithm to the prediction of CMEs is very rare in literature. The author proved that once the logistic regression model confirms the effectiveness for a CME, the recommendation [algorithm](#) can be used to recommend similar historical events. Recommending similar historical events as a vivid

reference for forecasters is a great improvement to the forecast service in contrast to the binary "yes" or "no" forecast provided by the [logistic regression](#) model only. Space weather forecasters may be able to make use of this method to execute a comparative analysis.

**More information:** Yurong Shi et al, Impacts of CMEs on Earth Based on Logistic Regression and Recommendation Algorithm, *Space: Science & Technology* (2022). [DOI: 10.34133/2022/9852185](https://doi.org/10.34133/2022/9852185)

Provided by Beijing Institute of Technology Press Co., Ltd

APA citation: How scientist applied the recommendation algorithm to anticipate CMEs' arrival times (2022, July 11) retrieved 29 November 2022 from <https://phys.org/news/2022-07-scientist-algorithm-cmes.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.*