AUSTRIA'S LEADING RESEARCH CENTER
FOR DATA-DRIVEN BUSINESS AND BIG DATA ANALYTICS

Automatic Detection and Classification of ICMEs in Solar Wind Data

H. Rüdisser[1,2], A. Windisch[1,4], U.V. Amerstorfer[3], T. Amerstorfer[3], C. Möstl[3], R.L. Bailey[5]

[1]Know Center, Graz, Austria; [2]University of Graz, Graz, Austria;

[3]Space Research Institute, Austrian Academy of Sciences, Graz, Austria;

[4]Department of Physics, Washington University in St. Louis, MO 63130, USA;

[5]Conrad Observatory, ZAMG, Vienna, Austria

Abstract

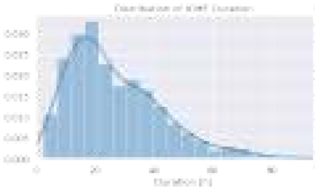
Interplanetary coronal mass ejections (ICMEs) are one of the main drivers for space weather disturbances. In the past, different machine learning approaches have been used to automatically detect events in existing time series resulting from solar wind in situ data. However, classification, early detection and ultimately forecasting still remain challenges when facing the large amount of data from different instruments. We attempt to further enhance existing convolutional neural network (CNN) models through extending their possibilities to process data from multiple spacecraft and to include a post processing step commonly used in the area of computer vision. Additionally, we make an effort to extend the previously binary classification problem to a multiclass classification, to also include corotating interaction regions (CIRs) into the range of detectable phenomena. Ultimately, we aspire to explore the suitability of several other methods used in time series forecasting, in order to pave the way for the elaboration of an early warning system.

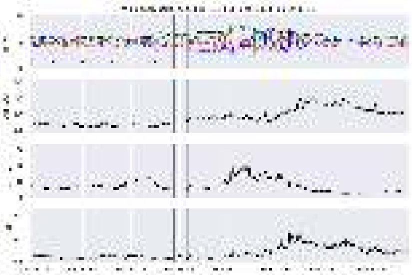
Introduction

The Classification Problem

The first step on the way to an early warning system is the reliable detection of ICMEs in an existing time series. While CNNs are often used to discover objects or patterns in images or data series, there are two main problems when facing our specific task:

- Duration variability
- Rather ambiguous definition of start and end



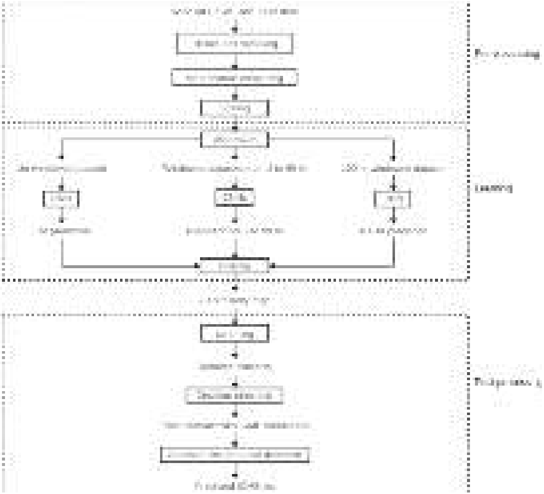


Distribution of ICME Duration and an example event, showing start and end of the ICME and start of the magnetic obstacle

Reimplementation

Architecture and Pipeline

Nguyen et al. 2019 proposed a deep learning approach consisting of several CNNs which were trained to predict a similarity parameter. At pre processing, the model takes 33 input variables from in-situ solar wind data, performs resampling and scaling and computes additional features. This method reasonably overcomes the problem of having highly varying durations in the range of ICMEs by ultimately combining the output of 100 separate CNNs which have been trained on different time lengths. A post processing step based on peak detection is then used to convert the 2D similarity map to a list of predicted ICMEs.

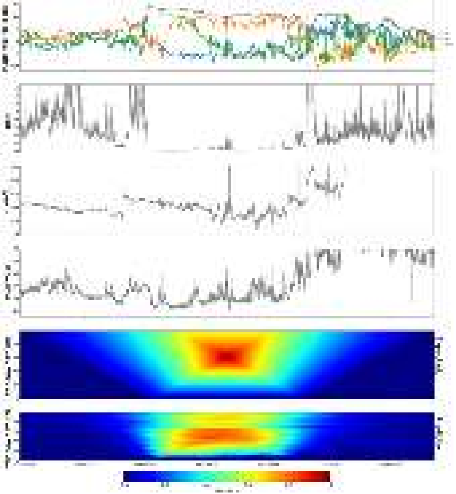


Pipeline of the Nguyen et al. 2019 model, divided into pre processing, learning and post processing (Nguyen et al. [2019])

Reimplementation

Performance

The model had previously been tested on WIND data and achieved a maximum recall of 84% $\pm$ 4.5% and a maximum precision of 84% $\pm$ 2.6%. The robustness to missing features could be established, even though the performance thus lacked in accuracy.

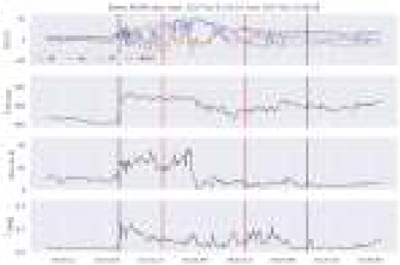


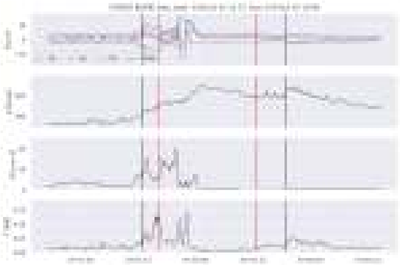
Observation during ICME from WIND spacecraft. Solid vertical lines delimitate the ICME. Dashed vertical line indicates beginning of sheath. From top to bottom are representations of magnetic field amplitude and components, plasma parameter, solar wind velocity, thermal velocity, similarity the ICMEs have with sliding windows of size 1 to 100, similarity predicted by the model. (Nguyen et al. [2019])

Reimplementation

New Data

We reimplemented the pipeline and additionally tested it on STEREO A and STEREO B data, which contain less variables than WIND data. The model was still able to achieve similar recall on all three data sets but the precision went down to around 30 %.
However, keeping in mind the ultimate goal, early recognition is an extremely important feature. The preciseness achieved by the model when delivering start and end time is limited, probably due to the fact, that the post processing is mainly based on arbitrary thresholds.





Events from STEREO A (left) and STEREO B (right). Black vertical lines delimitate true event, red vertical lines delimitate predicted event.

Reimplementation

Thoughts

Even though the reimplemented model performs reasonably well in terms of detection rate, we still observed some critical features.

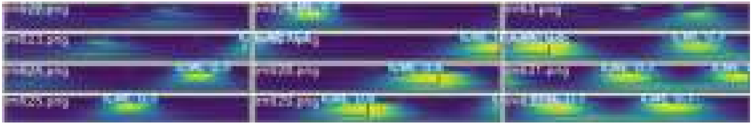
- Training 100 different models is computationally expensive and takes around 55h with GPU support
- Start and endtime could not be precisely predicted

Another interesting result is the fact, that simultaneously training the model on WIND, STEREO A and STEREO B data did not improve the performance as expected. This is a clear indicator, that the three data sets in fact are very different from each other and enormously raise the level of variability in events.

Post Processing

YOLOv5

YOLO (Ultralytics 2021) stands for You Only Look Once and is one of the most effective object detection algorithms currently existing in the computer vision community. It detects instances in images and outputs corresponding bounding boxes. We currently attempt to exchange the threshold based post processing step by a YOLO based image detection that converts the similarity map to a catalogue of ICMEs.



The YOLO algorithm processes the similarity map created by the Nguyen et al. 2019 model and outputs bounding boxes for detected events.

Conclusion

Achievements so far

- Reimplementation of the Nguyen et al. 2019 model and testing it on WIND, STEREO A and STEREO B data
- Inclusion of a YOLOv5 post processing step

Next steps

- Improvement of the post processing step
- Alignement of the datasets from different spacecraft

Through combining the presented models we hope to further improve detection rate and accuracy for providing start and endtime independent from the data source or specific event.

References

Gautier Nguyen, Nicolas Aunai, Dominique Fontaine, Erwan Le Pennec, Joris Van den Bossche, Alexis Jeandet, Brice Bakkali, Louis Vignoli, and Bruno Regaldo-Saint Blancard. *Automatic Detection of Interplanetary Coronal Mass Ejections from In Situ Data: A Deep Learning Approach*. The Astrophysical Journal, Vol. 874, No. 2, April 2019, 2019.

Ultralytics. Yolov5. `https://github.com/ultralytics/yolov5`, 2021.

Europlanet 2024 RI has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871149.

A. Windisch acknowledges support by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics under Award Number # DE-FG-02-05ER41375.

T. Amerstorfer thanks the Austrian Science Fund (FWF): P31265-N27

C. Möstl and R. Bailey thank the Austrian Science Fund (FWF): P31659-N27.