

## Association rules for solar flare attributes through hierarchical clustering

Jéssica F. Pereira\*, Ana E. A. da Silva.

### Abstract

This project aims to explore a set of association rules through hierarchical clustering as a way to preprocess the rules. The goal is to have a limited set of association rules based on the creation of groups which would indicate promising rules. The data used in the clustering and the construction of the rules are parameters of solar flares. For the construction of a proper data set of solar flares, the following actions were done i) choose a set of data; ii) preprocess the data in order to transform categorical data into binary data; iii) create groups by hierarchical clustering; iv) select groups to create association rules; v) apply the rules using an association algorithm; vi) validate the association rules.

### Key words:

Association rules, Hierarchical clustering, Solar flares.

### Introduction

Solar activities are phenomena that occur in the Sun in higher intensities in certain phases of the solar cycle. The solar cycle has an approximate duration of eleven years and has a pattern of behavior that presents greater activity in a given period - which is called solar maximum - and less activity in the rest of the cycle - called solar minimum [1].

This work focuses on the study of solar flares and the relation among their parameters. Solar flares, when intense, can trigger events which can affect the Earth and our daily activities [2]. The effects caused by these phenomena interfere with technological systems, such as mobile signals, satellites and global geographic positioning devices [3]. Therefore, the main idea is to apply an association rule algorithm in order to discover association rules, which can point out the relationship among parameters present in higher solar flares, such as M and X flares.

### Results and Discussion

There are several solar activity phenomena which are represented by various indexes which are coded in a numerical form. These indexes are classified into different ways and represent numerical characteristics of solar activity, which are observed and measured directly such as: magnetic classification of sunspots in active regions, area of sunspots, number of sunspots per active region; solar flare measured by X-ray flux, and radio flux. Values for these parameters can be found in the databases of the Space Weather Prediction Center [4]. An active region on the Sun is an area with a strong magnetic field. Most solar events like solar flares, are originated from active regions. For this study, the attributes presented were used as data for association rules algorithm and hierarchical clustering. All the data related to active regions were provided by SWPC, and Chart 1 represent the dataset.

Chart 1. Transactional dataset.

| Date       | Region ID | Spot Area | Magnetic classification | Solar flare X-ray flux | Radio flux |
|------------|-----------|-----------|-------------------------|------------------------|------------|
| 2000/03/01 | 8898      | 850       | beta-gamma              | $3,8 \cdot 10^{-5}$    | 240        |

The experiments were divided into three main periods according to the solar activity: 1997 to 2016 – all data from SWPC; data related to 23<sup>th</sup> solar cycle; and data related to 24<sup>th</sup> solar cycle. Interesting rules according to support and confidence metrics are present in the Chart 2. For association rules, data were categorized according to intensity.

Chart 2. Rules from experiments with the 23<sup>th</sup> and 24<sup>th</sup> solar cycles

|   | Support | Confidence | Association rules                                     |
|---|---------|------------|---|
| 1 | 1,95%   | 43,33%     | area-huge, beta-gamma-delta, radio-veryhigh -> xray-X |
| 2 | 1,20%   | 100%       | area-medium, beta-gamma-delta, radio-medium -> xray-M |
| 3 | 2,09%   | 62,50%     | beta-delta, radio-veryhigh -> xray-X                  |
| 4 | 35,14%  | 72,41%     | beta-gamma-delta -> xray-M                            |

Rules 1 and 2 represent most interesting rules found on experiments related to 23<sup>th</sup> solar cycle, and rules 3 and 4 related to 24<sup>th</sup> solar cycle. Hierarchical clustering is not present on these experiments as it proved to be inappropriate to this type of data.

### Conclusions

It was possible to discover interesting relationships between solar explosion attributes and its probability according to active region's configuration.

### Acknowledgement

This study was sponsored by Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP.

<sup>1</sup> YOUNG, C. Alex. Solar Activity. 2010. <<http://www.thesuntoday.org/the-sun/solar-activity/>>. Accessed in: 20 jan. 2018.

<sup>2</sup> COLAK, T.; QAHWAJI, R. Automated Solar Activity Prediction: A hybrid computer platform using machine learning and solar imaging for automated prediction of solar flares. Space Weather. v. 7, jun. 2009.

<sup>3</sup> FOX, Karen C. Impacts of Strong Solar Flares. 2013. <[https://www.nasa.gov/mission\\_pages/sunearth/news/flare-impacts.html#.WAGF2uArLIV](https://www.nasa.gov/mission_pages/sunearth/news/flare-impacts.html#.WAGF2uArLIV)>. Accessed in: 12 out. 2016.

<sup>4</sup> SPACE Weather Prediction Center: NOAA / NWS Space Weather Prediction Center. 2015. <<http://www.swpc.noaa.gov/>>. Accessed in: 13 out. 2016.