筑波学院大学紀要第18集　115 ～ 120 ページ　2023年

＜研究ノート＞

A proposal for Clustering of Sunspots using Machine Learning to Explore Inter-cycle Patterns

Bhola PANTA *

Abstract

Sunspots were first observed by telescope in 1610. This makes a record of more than 400 years. Even with a long history and abundance of ground-based and space borne observatory data, fundamental questions about their formation are still shrouded in mystery. Hence, sunspots are one of the most studied phenomena in solar physics. The research on sunspots is important because they are precursors to solar flares, which can have profound impact on life here on earth. We propose a machine learning based clustering technique of the sunspot time series data to find sunspot patterns between solar cycles. This will eventually be correlated with solar flare data to understand the patterns of solar flare occurrences in future sunspot cycles. Ultimately, we expect to contribute to develop a more efficient space weather forecasting system. This research note presents a preliminary investigation of the issue in question, and a roadmap.

Key words: Sunspots, Solar flare, Machine learning, Clustering, Particle Swarm Optimization

1. Introduction

Every once in a while, we all see news about "solar flares" on TV and other news outlets. A solar flare is a large-scale explosion that occurs around a sunspot on the "surface" of the sun, called the photosphere. High-energy particles and radiation from the flare travel to the Earth. Depending on the scale of the flare, it can cause various damages. Induced currents can enter power lines and affect the power grid, communication satellites can break down or stop functioning, etc. Research on "space weather" caused by such phenomena is an active research field around the world. In the USA, National Oceanic and Atmospheric Administration (NOAA)'s Space Weather Prediction Center (SWPC) continually monitors and forecasts Earth's space environment, providing solar-terrestrial information. In Japan, the National Institute of Information and Communications Technology (NICT) provides space weather forecasts.

Sunspots are temporary spots on the photosphere. These spots are formed and grow in size, as they move across the disk of the sun. They eventually decay. Individual sunspots or groups of sunspots may have a lifespan of several days to a few months. Solar flares and other temporary

＊　筑波学院大学 経営情報学部、Tsukuba Gakuin University

events like coronal mass ejection (CME) mostly occur around sunspots. George Ellery Hale (1868–1938) discovered magnetic fields in sunspots. The sunspots appear in the magnetically active regions. The strong magnetic flux on the sunspots hinders convection from the inner core of the sun, resulting in reduced temperature, which in turn makes the spots appear darker than the surrounding areas.

In the past 100 years, we have seen notable developments in instrumentation to carry out observation, better understanding of polarized light (for spectroscopic analysis) and better modeling capabilities (e.g., MHD) to work on the theoretical aspects [10]. Thanks to those developments, we now have plenty of data of high accuracy, resolution and cadence, specially from the space-borne observatories like the well-known Solar Dynamics Observatory (SDO) and Solar and Heliospheric Observatory (SOHO). With such developments, solar astronomy has stepped in "big data", in which tools and techniques of data science are expected to play a significant role in the near future, to further unravel the mysteries of the sun.

2. Motivation

As mentioned earlier, abundance of observation data from ground based as well as space-borne observatories are available. This has created huge challenges in terms of transferring data from the observation site to the storage site, processing "raw data" to produce "science-ready data" and storing. At the same time, with significant increase in computing power, including the availability of GPUs, it is becoming increasingly possible to apply machine learning techniques to reduce human labor when processing huge volumes of data. For example, in the case of the solar radio observation data, it is said that less than 5% of the data is valuable to solar physicists [13]. Hence the importance of machine learning-enabled classification algorithms cannot be overstated: even a binary (useful/not useful) classifier would dramatically free up astronomers to do more productive tasks.

Using data science techniques, we would like to explore if there is any hidden pattern in sunspot's appearance over a sunspot cycle. Starting from the sunspot's distribution in time (date of appearance) and space (where they appear on the sun's disk) for each sunspot-cycle, we will try to analyze if there is any pattern that may be applied to another cycle. Together with the solar flare database from various observatories and analyzing the sunspot data at the flare date and time, these findings may eventually be used to develop a more efficient space weather forecasting system.

Take for example, the X9-class solar flare of Dec. 5, 2006, shown in Fig. 1. (For flare classification, readers are advised to reference appropriate literature. One example is [17]). The source sunspot of the X9-class flare is shown in Fig.2. This research will try to analyze the sunspot cluster data patterns, and analyze if correlations exist in other cycle's sunspots and flares.

Space weather forecasting is said to be several decades behind terrestrial weather forecasting. We expect to evaluate the clustering models using various metrics (e.g., silhouette index, and Dunn index) and propose efficient/reliable models. By feeding back the results to the data science and solar research community, we hope to contribute to the space weather research and forecasting field.

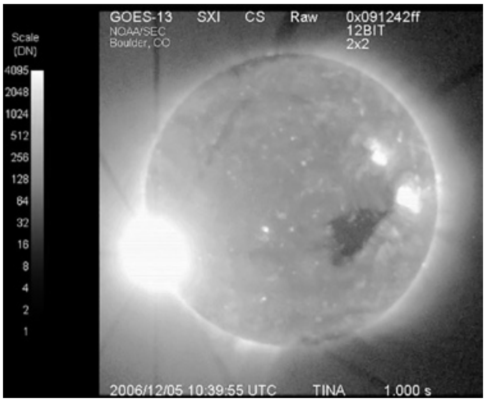

Fig. 1 The X9-class solar flare of Dec. 5, 2006, observed by the Solar X-Ray Imager aboard NOAA's GOES-13 satellite. Credit: NOAA/SWPC

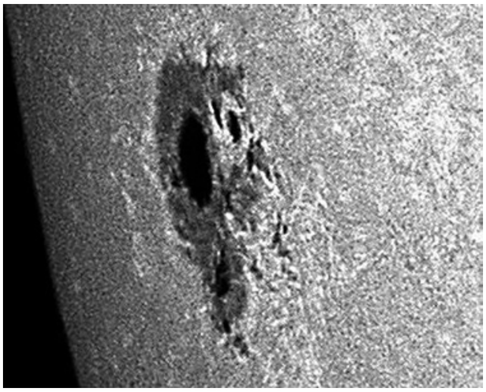

Fig. 2 Sunspot 930, the source of the powerful X9-flare on Dec. 5, 2006. Credit: John Nassar

3. Related Works

Sunspots have been studied for a long time. According to Wikipedia, the earliest record of sunspots is found in China, in a text composed before 800 BC [14]. After the invention of telescopes, astronomers started to record observations more systematically. Since the sun's magnetic field affects all aspects of solar activity [10], considerable literature exists analyzing sunspots and solar flares in the light of magnetic fields, for example, either using MHD simulations or through observed data (magnetograms or radio observations).

Data-science based exploration of astronomy data in general and solar astronomy data in particular are more recent phenomena. Since the third wave of the AI boom (2000-), we have seen a huge increase in application of machine learning techniques in sunspot data analysis, especially since the introduction of deep learning in 2006 [5,9,13]. Mandal et al. [8] reviewed sunspot area catalogs of nine archives, one of which is Debrecen Heliophysics Observatory archive. Stenning et al. [12] applied mathematical morphological analysis to extract numerical summaries of sunspot images that can be used for automatic classification. de Toma et al. [6] analyzed changes in sunspot area using photometric images over two solar cycles. Their findings were confirmed by USAF/NOAA and Debrecen data. This indicates our choice of Debrecen data is reliable. When necessary, we will reference the above results.

4. Sunspot data

Debrecen Photoheliographic Data (DPD) [4,7] is a catalog of positions and areas of sunspots compiled by using white-light full disc observations taken at Debrecen and Gyula Observing Stations in Hungary, as well as at other observatories in Austria, Belgium, Egypt, Georgia, India,

Italy, Japan, Russia, Spain, UK, Ukraine, USA and Uzbekistan. The DPD is published as an ASCII file. A snapshot of the time series of the daily data is shown in Fig. 3.
For each spot, the following data are available: date of observation (year, month, day, hour, minute and second, shown in column C1 of Fig.3), the origin of the observation (C2), the measured (projected) and the corrected areas of umbrae and the whole spot (C3 ~ C6), Julian date (C7), position angle of the northern extremity of the axis of rotation (measured eastward from the north point of the disk) (C8), and the heliographic latitude of the central point of the disk (C9). A detailed description of the data is available at the DPD site. In our clustering proposal, we will use the necessary columns, as described in the methodology section.

5. Methodology to be employed, and further Plan

In machine learning, clustering techniques are widely used to discover trends in a big dataset. Clustering is an unsupervised machine learning technique for organizing data into groups. If there are patterns in the sunspot data, an appropriate clustering technique may be able to identify when (temporal) and where (spatial) these patterns occur. For some popular clustering algorithms, readers are advised to reference appropriate literature in data mining. One excellent reference is "Introduction to Data Mining", by Tan, Seinbach, Karpatne and Kumar. (Pearson Publishing, 2018)

C1	C2	C3	C4	C5	C6	C7	C8	C9
19740101011000	MITA	0	0	0	0	2442049	999999	999999
19740102032030	MITA	45	227	23	116	2442050	1.65	-3.17
19740103005630	MITA	58	489	30	249	2442051	1.21	-3.27
19740104095230	GYUL	20	79	11	43	2442052	0.54	-3.43
19740105021530	MITA	0	14	0	8	2442053	0.21	-3.5
19740106080000	GYUL	0	13	0	10	2442054	-0.39	-3.64
19740107012300	MITA	0	0	0	0	2442055	999999	999999
19740108064616	CAPE	32	172	54	278	2442056	-1.33	-3.86
19740109073833	CAPE	45	298	31	259	2442057	-1.83	-3.97
19740110004400	MITA	235	966	186	662	2442058	-2.17	-4.05
19740111105900	GYUL	100	672	75	493	2442059	-2.85	-4.2
19740112004200	MITA	246	871	174	718	2442060	-3.13	-4.26
19740113090930	GYUL	122	892	93	669	2442061	-3.77	-4.4
19740114123800	GYUL	143	940	105	634	2442062	-4.31	-4.51
19740115105530	GYUL	141	852	88	528	2442063	-4.75	-4.61
……	……	……	……	……	……	……	……	……
……	……	……	……	……	……	……	……	……

Fig. 3 A snapshot of the daily time-series data as provided by the Debrecen Heliophysics Observatory, from 1974 to 2016. To ease data processing, the date column (C1) has been created merging individual columns for the year, month, day, hour, minute and second, in the original dataset.

For each solar cycle, we will create 3-dimensional data clusters using (1) date of observation (column C1), (2) heliographic latitude of the central point of the disk on the observation (column C9), and (3) daily sum of corrected whole spot area in millionths of the solar hemisphere (column C6).

Since we are considering using daily time series data, we will use “Time Series Clustering”. Since the DPD data has both temporal dimension (date of observation) and spatial dimension (heliographic latitude), we will perform both temporal as well as spatial clustering. Performance of k-means clustering with the dynamic time warping (DTW) metric [16] will be evaluated. We will also consider developing necessary algorithms to carry out the experiment when necessary.

In another implementation, a Particle Swarm Optimization (PSO) [1,2,3] based clustering will be considered. Since PSO is an optimization algorithm, we need to cast the clustering problem as an optimization problem. Eventually, we will consider a hybrid approach- PSO implementation in combination with classical clustering techniques like k-means, to optimize the solution.

6. Acknowledgements

A research project to cluster the sunspots using machine learning was supported by a grant from Tsukuba Gakuin University in FY2022 (Kyodo Kenkyu). This research note details a preliminary investigation to carry out that research. The author expresses his gratitude to the selection committee, and to the Kiyo committee to get this research note published.

The author would also like to express gratitude to the Debrecen Heliophysical Observatory (http://fenyi.solarobs.csfk.mta.hu/DPD/) for providing daily time series data from 1974 to 2016, and NOAA National Centers for Environmental Information for providing solar flare index data [15].

References

[1] Abraham A., Das S., Roy S., Swarm Intelligence Algorithms for Data Clustering. In: Maimon, O., Rokach, L. (eds) Soft Computing for Knowledge Discovery and Data Mining. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-69935-6_12, 2008.

[2] Alam S, Dobbie G, Riddle P, Naeem M, A., Swarm intelligence-based clustering approach for outlier detection, Proceedings of the IEEE Congress on Evolutionary Computation, DOI:10.1109/CEC.2010.5586152, Barcelona, Spain, 18-23 July 2010.

[3] Ballardini A., A tutorial on Particle Swarm Optimization Clustering, https://www.researchgate.net/publication/327496660_A_tutorial_on_Particle_Swarm_Optimization_Clustering, accessed 2022/10/1, 2016.

[4] Baranyi, T., Győri, L., Ludmány, A., On-line Tools for Solar Data Compiled at the Debrecen Observatory and their Extensions with the Greenwich Sunspot Data, Solar Physics, 291, 3081-3102, DOI: 10.1007/s11207-016-0930-1, 2016.

[5] Chola C, Benifa J., Detection and classification of sunspots via deep convolutional neural network, Global Transitions Proceedings, Volume 3, Issue 1, Pages 177-182, DOI:10.1016/j.gltp.2022.03.006, June 2022.

[6] de Toma, G. et al., Analysis of sunspot area over two solar cycle, The Astrophysical Journal, 770:89 (13pp), 2013 June 20 doi:10.1088/0004-637X/770/2/89, 2013.

[7] Győri, L., Ludmány, A., Baranyi, T., Comparative analysis of Debrecen sunspot catalogues, Monthly Notices of the Royal Astronomical Society, 465 (2), 1259-1273, DOI: 10.1093/mnras/stw2667, 2017.

[8] Mandal S., et al., Sunspot area catalog revisited: daily cross-calibrated areas since 1874, A&A, Volume 640, A78, DOI: 10.1007/978-0-387-69935-6_12, August 2020.

[9] Premebida S. et al., Prediction of Sunspot Behavior Using Machine Learning Techniques, DOI:10.21528/CBIC2019-40, Congresso Brasileiro de Inteligência Computacional, Jan 2020.

[10] Priest, E. R., Solar Physics: Overview, Published online, DOI:10.1093/acrefore/9780190871994.013.21, July 2020.

[11] Rempel M., Borrero J., Sunspots, doi:10.1093/acrefore/9780190871994.013.16, Published online: 23 February 2021.

[12] Stenning D, Kashyap V, Lee T. C. M, van Dyk D. A., Young C. A, Morphological Image Analysis and its Application to Sunspot Classification, DOI:10.1007/978-1-4614-3520-4-31, Jan 2013.

[13] Xu et al., Deep learning in Solar Astronomy, Springer, 2022.

[14] https://en.wikipedia.org/wiki/Sunspot, accessed 2022/10/21

[15] https://www.ngdc.noaa.gov/stp/solar/solarflares.html, accessed 2022/10/21

[16] https://rpubs.com/esobolewska/dtw-time-series, accessed 2022/11/24

[17] http://solarcenter.stanford.edu/SID/activities/flare.html, accessed 2022/10/21