CIF21 DIBBs: Systematic Data-Driven Analysis and Tools for Spatio-temporal Solar Astronomy Data

PI: Dr. Rafal Angryk (GSU/Computer Science(CS)), **Co-PIs:** Dr. Petrus Martens (GSU/Solar Physics(SP)), Dr. Katherine Reeves (SAO/SP), **Post-Docs:** Dr. M. Schuh (CS), Dr. S. Hazra (SP), **Students**: B. Aydin, D. Kempton, T. Gholston and W.G. Johnson

Abstract

We present the overview of interdisciplinary data-driven research conducted by the Data Mining Lab at Georgia State University under NSF's CIF21 DIBBs award. Our works focus on novel analyses of multiple types of solar astronomy data coming from space and ground observatories, including: SDO (AIA & HMI), and GONG H-Alpha Network (BBSO) instruments. We present fundamental steps in our process of spatio-temporal patterns mining, which integrate massive image (raster), and object (vector) data repositories of solar astronomy.

Objectives

- To report our progress on Solar Astronomy Big Data Mining research and development
- To stimulate new advances through reaching out to the community, sharing our cleaned data benchmarks and open-source software
- To find new research collaborations and projects related to Solar Data Mining and Spatio-temporal Big Data Analytics. Please contact us at: rangryk@gsu.edu

Background

Our interdisciplinary Data Mining Lab was founded in 2005 by Prof. Petrus Martens (**Solar Physics**), and Prof. Rafal Angryk (**Computer Science**) at Montana State University. In 2013 the Lab moved to Georgia State University. We were awarded DIBBs CIF21 grant in 2014.

dmlab.cs.gsu.edu

Solar Astronomy Big Data

Goal: Knowledge discovery from Big Data and verification and combination of Automated Event Recognition Modules.

Event reports in HEK:

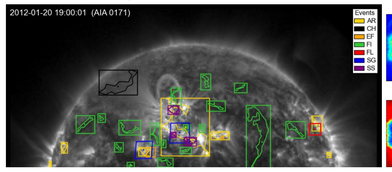
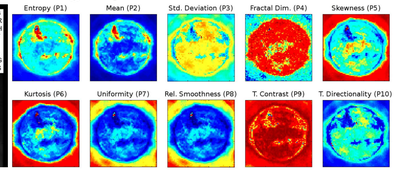
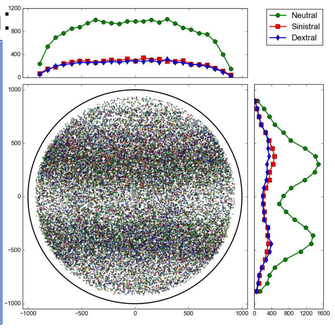


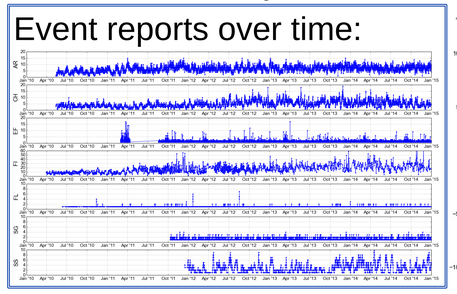
Image params in AIA:



Filament chirality distribution:



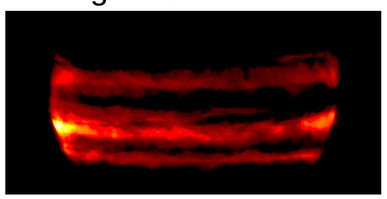
Event reports over time:



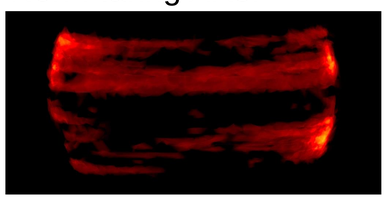
Event Tracking

Goal: Generation of the largest and most reliable dataset benchmark of continuous spatio-temporal solar event instances to enable new large-scale analyses of solar astronomy data by data scientists.

Flaring ARs:

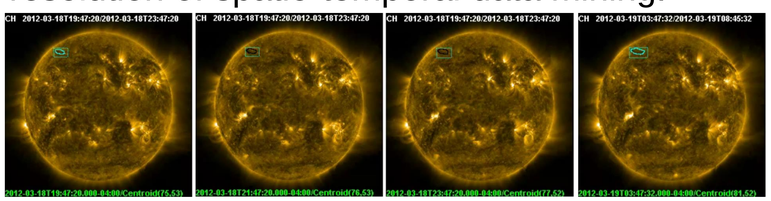


Non-Flaring ARs:



Shape Interpolation

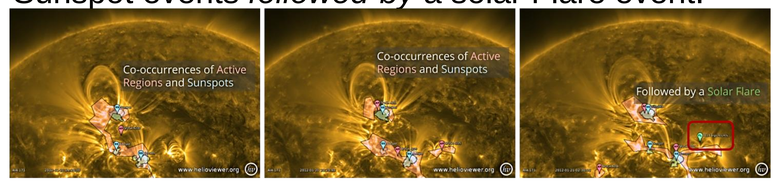
Goal: To improve accuracy and increase resolution of spatio-temporal data mining.



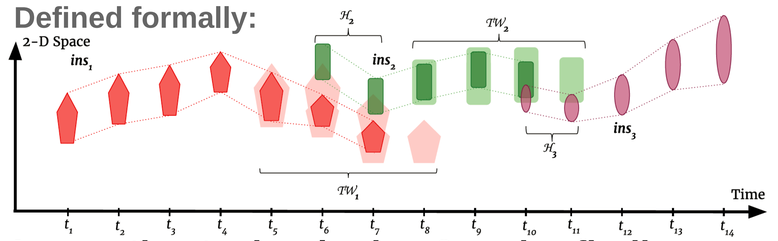
Spatio-Temporal Patterns

Goal: Discovering spatio-temporal sequences of solar events by mining relevant subsets of event types that frequently occur one after another. This knowledge is crucial towards the predictions of solar events and space weather.

E.g.: The co-occurrence of Active Region and Sunspot events *followed-by* a solar Flare event.



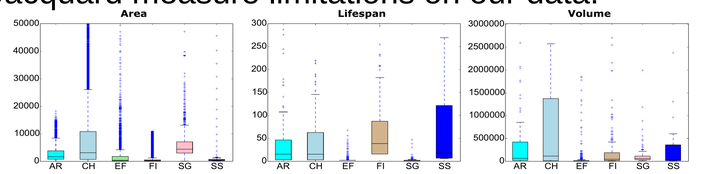
Defined formally:



Innovative technologies & major findings:

- Development of measures that are appropriate for highly dynamic and variable types of spatio-temporal events
- Implementation of scalable algorithms that can analyze solar astronomy big data
- Discovery of novel, accurate, interesting, and actionable/useful knowledge

Jacquard measure limitations on our data:



Our first spatio-temporal sequence patterns of solar events (January 2012):