Using Unsupervised Machine Learning to Explore New Classification of Sunspot Active Regions

Student: Sara Housseal (Millersville University)
Mentors: Tom Berger (SWx TREC),
Varad Deshmukh (SWx TREC, CU Boulder Computer Science)

Abstract

For as long as active regions have been getting classified, it has been done manually by humans, by studying their features and labeling them according to the McIntosh and/or Hale classification systems. There has also been previous work done to automate McIntosh and Hale classification using supervised machine learning but it's time to see if computers can perform unsupervised machine learning to classify active regions in a new quantitative way. We have performed Principle Component Analysis (PCA) on the Solar Dynamics Observatory (SDO) Helioseismic and Magnetic Imager (HMI) Space-weather HMI Active Region Patch (SHARP) images of all sunspot active regions from 2010 to the present. We performed a PCA on 100,000 randomly selected active region images to create 40 PCA "eigenvector images" of active regions. The 40 eigenvector images show that the PCA picked up on the complex magnetic multipole of sunspot active regions and can be narrowed down to 16 eigenvector images to allow compact classification. While work will be continued, the end goal is to create a new classification system for sunspot active regions based on the evolution, magnetic field strength, and the PCA weights of different polarity features of the active region. The results confirm that PCA analysis can identify key features of active region structures and characterize them in a compact eigenvector form. Applications of this method include improved prediction of solar flares based on quantitative PCA structure classification.