Project Proposal
Solar Prediction

March 12, 2016

Basic Description

A solar flare is a sudden flash of brightness observed near the Sun's surface. It involves a very broad spectrum of emissions, requiring an energy release of up to 6 1025 joules of energy. The flare ejects clouds of electrons, ions, and atoms through the corona of the sun into space. These clouds typically reach Earth a day or two after the event.
Flares occur when sped up charged particles, mainly electrons, interact with the plasma medium. On the Sun, magnetic reconnection may happen on solar arcades a series of closely occurring loops of magnetic lines of force. These lines of force quickly reconnect into a low arcade of loops leaving a helix of magnetic field unconnected to the rest of the arcade. The sudden release of energy in this reconnection is the origin of the particle acceleration. Solar flares typically erupt from what are known as the active regions on the Sun where magnetic fields are much stronger on average.
Current methods of flare prediction are problematic, and there is no certain indication that an active region on the Sun will produce a flare. However, many properties of sunspots and active regions correlate with flaring. For example, magnetically complex regions (based on line-of-sight magnetic field) called delta spots produce the largest flares.
Two classification systems exist for sunspots: McIntosh and Mt. Wilson. McIntosh classification depends on the size, shape and spot density of sunspots, while the Mt. Wilson classification is based on the distribution of magnetic polarities within spot groups. For this project we will use McIntosh classification. Thus based on the

Data

As of now we have found the following datasets:- https://archive.ics.uci.edu/ml/datasets/Solar+Flare Each class attribute counts the number of solar flares of a certain class that occur in a 24 hour period. The data has the attributes:

1. Code for class (modified Zurich class)
2. Code for largest spot size
3. Code for spot distribution
4. Activity
5. Evolution
6. Previous 24 hour flare activity code
7. Historically-complex
8. Did region become historically complex on this pass across the sun's disk
9. Area
10. Area of the largest spot

The data to predict is

1. Number of C-class flares production by this region in the following 24 hours (common flares)

2. Number of M-class flares production by this region in the following 24 hours (moderate flares)

3. X-class flares production by this region in the following 24 hours (Severe Flares)

However, this data is just to predict the number of flares on the day based on the solar activity.
To make the data more realistic, we will be using http://sci2s.ugr.es/keel/dataset.php?cod=98 as our dataset.

Links

Our key paper we would be referring to is AUTOMATIC SHORT-TERM SOLAR FLARE PREDICTION USING MACHINE LEARNING AND SUNSPOT ASSOCIATIONS by R. QAHWAJI and T. COLAK.
http://www.stce.be/sipworkshop2014/presnpost/s3pres/PresentationMLSF2_DAmbrosio.pdf
This paper has described the methodology to be done. We would be following this paper closely, along with making our own modifications.

Other than the above paper, we would also be referring to this paper for more clarity on the topic
http://www.stce.be/sipworkshop2014/presnpost/s3pres/PresentationMLSF2_DAmbrosio.pdf

Methods

In the research paper we are referring to, Cascade-Correlation Neural Networks (CCNN), Support Vector Machines (SVM) and Radial Basis Function Networks (RBFN) are used and compared for flare prediction. CCNN and RBFN are used because of their efficient performance in applications involving classification and time-series prediction. We would be applying SVM on 2 variables at a time for prediciton if time permits we would expand it to multidimensional SVM. Although the research clearly shows that the SVM clearly outdid Neural Networks, we wish to get our hands dirty and try it for ourselves as well.

Plan

Plan for 21st Submission: We intend to submit the features abstraction which would be basically from the first database and if time permits we would start with few simple algorithm on the real data.

Plan for final Submission: We would submit the final working code for the following dataset, the test data and the training data will be well separated and the final outcomes. The algorithms will be implemented as close to the research paper to predict the class of the Solar Flares.

Group Members

1. Harshal Mahajan (140050003)

2. Jayanth Shankar Chennareddy (140050039)

3. Vishal Babu Bhavani (140050049)

4. Chaitanya Rajesh Banala (140050073)