Intelligent Systems Project 2 Proposal

Proposal Statement

A solar flare is a sudden increase of intensity of radiation emitted from the sun. Intense solar flares can influence many systems on Earth, including radio communication and animal behavior. Usually, the radiation emitted by a solar flare is in the x-ray range, and so satellites have been launched with the purpose of measuring the x-rays emitted from the sun to detect solar flares [1]. For my project, I intend to take x-ray flux data collected by one such satellite, the GOES15, and build a time series forecaster to predict trends in the solar cycle, and therefore predict when solar flares may occur. I plan to use a Recurrent Neural Network (RNN) to implement my time series forecaster.

Resources

To implement the RNN, I plan to use TensorFlow. Specifically, I plan to use Keras to construct, train, and evaluate the RNN. For data, I will be using a science quality dataset which contains a summary of the x-ray data collected by the GOES15 mission from 2010 to 2020 [2]. A portion of this dataset will be used for training the network, and the remainder will be used for validation of the network.

Deliverables

The deliverable of this project will be a trained and persisted RNN which will have the ability to predict trends in the flux of the x-ray radiation hitting Earth from our sun, as well as the source code used to train and persist the network. A README which contains the instructions on how to run the code, as well as a small performance report will also be included. All of these deliverables will be published in a GitHub repository.

Schedule

Friday, November 24th: Data is collected and in the desired format, source code for RNN is written. From here, I will start to debug the source code for training and evaluating the RNN.

Friday, December 1st: RNN code is debugged and the network is training. From here, I will play with the parameters of my net to try to maximize the accuracy and minimize the loss.

Friday, December 8th: Network is training correctly, loss is low, accuracy is high. From here, work on generating meaningful plots and writing the performance report.

Friday, December 15th: Network is complete and persisted, performance report and instructions are written.

Risks

One of the major risks of this project is the complexity of the dataset that I am working with. The time between each x-ray flux measurement is non-uniform, which may complicate the time series forecasting. Additionally, the dataset is fairly large for my CPU, about 70,000 rows. This makes the bulldozing process more time consuming, which makes the project's scope a little bit precarious. Another risk is my unfamiliarity with Keras. Although I worked with Keras during Project 1, I'm still not very familiar with the library, which might cause me to have lots of difficulty during the debugging process. A final risk is the strain that this project may put on my already large workload. I'm in 19 credits of strenuous physics and computer science coursework, and other assignments, exams, and projects may keep me from pouring enough time into this project to fully realize it.

References

[1] https://ngdc.noaa.gov/stp/solar/solarflares.html

[2] https://www.ncei.noaa.gov/data/goes-space-environment-monitor/access/science/xrs/goes15/xrsf-l2-flsum_science/