

Final Project Plan

Raymond Blaha and Paul Hwang

November 28, 2023

1 Introduction / Question

Can we predict wildfire risk zones using historical wildfire, vegetation, and climate data in the California region?

2 Data Collection

Since I am collecting data from two different sources one challenge will be to combine them into one cohesive dataset. Current the order of events in which data collection will be occurring:

1. 1.0 MODIS: I created an Earthdata account to get access to the most updated data.
2. 1.1 Explored the NASA Earthdata Search. Where I filtered the MODIS14A.
3. 1.2 Used the `chmod 777 download.txt` in local terminal to start the download process.
4. 1.3 Used login information in local terminal to download the data.
5. 2.0 USGS: Went in and downloaded the data.
6. 2.1 Merged the MODIS and USGS data via geographical location.
7. 2.2 Download the climate data.
8. 2.3 Extract the temporal climate and perform merge with the already joined MODIS and GAP data.
9. 2.4 Split the data into training, validation, and testing.
10. 3.0 Formulate the model to have a seamless execution and flow from each model to the next.
11. 3.1 Load the three models to test the model on unseen data.
12. 3.2 Visualize the results.
13. 3.3 Compare the results to the labeled ground truth data with fire labels.

3 Data Exploration

3.1 MODIS

After downloading the satellite data which took 10 hours. I inspected the data and the metadata of the HD5 files to get a better understanding of how we could merge the data.

3.2 USGS

After downloading the data, the most important and relevant data is the *raster_values* and the spatial data. Since the vegetation data does not really change over time. The best approach to merge was to convert the USGS data from WGS 84 referencing the spatial instances of vegetation to latitude and longitude. For every instance or non instance of the MODIS we can combine the two datasets. This allows us to keep the temporal instances of fires and non fires, while having vegetation data. Merging took approx 30 hours to complete. For the sake of computational efficiency, we extracted the 2D spatial data from the USGS data. So it still can be read into the CNN while keeping the temporal instances of the MODIS data.

3.3 Climate Data

Since the data is temporal. We can merge the data by the temporal instances, while trying to keep in mind location.

4 Data Preprocessing

We will need to preprocess certain columns in which we want to input into the models. First instances will be some of the metadata from MODIS about fires, time, the raster values from USGS, and the climate data.

5 Modeling

Regarding the autoencoder, I will first have the autoencoder find features in the satellite data. Once that is done, I will then have the CNN analyze the features the 2D identify spatial patterns that will be key for predicting wildfire risk zones. Finally, the LSTM will give us prospective on the temporal climate instances which will give the model more context to predict wildfire risk zones.

6 Visualization of Results

Results will be presented in a simple format, such as risk zone maps based on the model's analysis.

Another aspect that I would like to explore is comparing the model's predictions to the ground truth data. To see how accurate the model is and or if there are any inefficiencies in the model. Then make the proper adjustments to the model if needed.