Team Mahogany — Wildfire Spread Predictions

Wildfires cause billions of dollars in damage each year in the United States alone, burning millions of acres of land, and releasing enormous amounts of CO_2, and causing thousands of deaths. In order to reduce these significant costs, we want to predict where ongoing wildfires will spread. Stakeholders could use this information to optimally allocate resources and direct first-responders where to begin suppression and government mandated evacuation efforts. Using data from wildfires in the United States from 2012 to 2020[1.], the goal of our project is to answer: How accurately can we predict where wildfires will spread?

Data Collection

Preprocessing consisted of assembling the dataset from 6 different satellite-based sources via Google Earth Engine. This can be done manually, or the complete preprocessed dataset can be downloaded from Kaggle.

The full dataset consists of 18,545 wildfires in the United States from 2012–2020. For each fire, the dataset consists of eleven environmental features describing the terrain, weather, and an additional feature defining where the fire was on two consecutive days. We randomly crop a square from each region, and then use all of the features, along with the fire boundary from the first day to predict where the fire will spread on the next day.

The dataset itself limits how accurately we are able to predict where fires spread. A significant amount of data is either missing, due to smoke, cloud, or tree cover; or not on fire. As only ~1% of the total area is on fire, the fires we want to predict are occasionally cropped out during preprocessing, making it very difficult to predict where they are going to spread.

The dataset also contains no information about fire suppression efforts, like dispersing water/retardant and setting up firebreaks, limiting where the fire would naturally spread, which reduces prediction accuracy.

Modeling Approach

We chose 3 models to predict where wildfires spread:

- Logistic Regression: as a baseline,
- Random Forest: to have a comparison of the baseline,
- Convolutional Neural Network: can learn spatial hierarchies.

Both Logistic Regression and Random Forest yielded improved precision over the original paper; however, they lacked spatial awareness of the fire spread. Both models only predicted that the fires would remain in the same location. These models disagreed on which features were most important, other than the previous fire boundary.

Using a Convolutional Neural Network allows us to spatial information to predict the location of the fire. Although it takes a long time to train, it produced state-of-the-art recall and AUC (PR) metrics for this dataset. During inference, stakeholders can run the model almost instantaneously, as soon as a mask defining the fire boundary can be manually produced. Unfortunately, this dataset only collects fire information once a week.

Conclusions and Future Directions

Wildfires are an inevitable part of millions of peoples' lives, as they result in thousands of deaths and billions of dollars in damage each year. Predicting wildfire spread allows stakeholders like government officials to direct evacuations and rescue teams to allocate resources and optimally engage in fire suppression techniques.

Of our models, Random Forest had the best precision rate at 39.72% while the Convolutional Neural Network had the best recall and AUC (PR) at 42.41% and 29.72%, respectively. These values are state-of-the-art!

In the future, these models could be enhanced by incorporating ongoing suppression techniques, and applying the models internationally to predict wildfire spread in those regions. We would like to find a way to account for missing data, and improve performance to the point where a final model could be reliably used by the public.

1. F. Huot, R. L. Hu, N. Goyal, T. Sankar, M. Ihme, and Y.-F. Chen,, "Next Day Wildfire Spread: A Machine Learning Data Set to Predict Wildfire Spreading from Remote-Sensing Data", IEEE Transactions on Geoscience And Remote Sensing, 60, 2022.