Predicting Global Climate Activity Through Machine Learning

Myles Barney, Braydon Jones

10/10

Natural disasters around the world in 2022 caused an estimated $313 billion in damages, destroyed the homes of 3.4 million people, and caused around 31,000 deaths.[1] Each year brings new challenges associated with climate change, population growth and increased reliance on global supply chains. But the problem of predicting early when and where these disasters will occur has continued to elude modern scientists. We propose a new approach to predicting these events by analyzing historical trends and data using a machine learning model, making predictions based on a larger input set by using self-attention or a recurrent neural network. Data will be fed into the model in a sequential manner, allowing the model to learn and respond to more data during training. We hope that this model will be able to provide more timely warnings to people living in areas that are vulnerable to natural disasters – giving them a chance to prepare or leave long before the sirens begin to sound.

When thinking through the features to include in our data set, we thought of two approaches. The first approach is to have the model take a day's worth of weather data and, using the context of the surrounding days, be able to predict a natural disaster for a given context. This would mean that the dataset would have a relatively small number of features but a larger number of rows of data. The second approach is to have the model take a full year—or two or three years—of data as input to train being able to predict a natural disaster based on a single row of years' worth of data. The problem we ran into with this approach is that the feature set was too large—a whole year's worth of data. This would then lead to a smaller dataset because we could only give the model as many rows as there are years of historical data. Because of this, we chose to go with the first approach, greatly reducing our dataset's feature set. The features are thus on a day-by-day basis as follows: Date (milliseconds since epoch), Latitude, Longitude, Barometric pressure, Temperature Min and Max, and inches of precipitation. The output is the severity, location, and date of the hurricane that is predicted to occur.

Date	Lat	Long	Pressure	TMin	TMax	Precipitation	**HSeverity**	**Lat**	**Long**
03092023	39.75	-111.2	30.06	10	30	2.2	3.4	40.3	90.1

There is a plethora of historical climate data available online. We will use datasets which contain daily readings of each feature so that we can feed the data in sequentially. Several such datasets can be found at the National Centers for Environmental Information.[2] We will download large datasets from different websites and break them into their separate regions so they can be fed into the model in their proper order. The data would need to be combined, cleaned, and integrated.

---

[1] "The Invisible Toll of Natural Disasters". Undrr.org. 2022. https://www.undrr.org/explainer/the-invisible-toll-of-disasters-2022

[2] "Climate Data Online: Dataset Discovery". Ncei.noaa.gov. https://www.ncei.noaa.gov/cdo-web/datasets