Wildfire Prediction Using LSTM

Team Flare

Kelly Lin, Alvin Tu, Steven Chang, Shuya Shou, Nick Ong

Introduction

Welcome!
We are Team Flare and have paired up with PricewaterhouseCoopers (PwC) for our capstone project. PwC is an international enterprise dealing in services not limited to: assurance, risk advisory, consulting, and data analytics. We have been tasked with designing a Long Short Term Memory network that focuses on predicting wildfires.

Kelly Lin

Alvin Tu

Steven Chang

Nick Ong

Shuya Shou

pwc

PwC Mentors:
Barbara Wortham, Alexander Andrianov,
Hossein Lotfi, Maggie Brickner

Background / Problem

- Climate change is making wildfires a growing problem
    - ~7.4 million acres burned annually in the U.S.
    - ~$2.4 billion in damages a year
- Danger to business infrastructure, ability to work
- Risk to individuals' properties and lives
- ***Need a method to predict wildfires to mitigate the dangers and risks they present***
- Existing methods have problems with making long-term predictions
    - Recurrent Neural Networks suffer from long term dependencies

Our Solution

- Leverage LSTM to model the long term trends between climate data and wildfire occurrences
  - LSTM is a type of recurrent neural network that is specialized to work with time series data
- Develop a pipeline for data analysis & preprocessing, model building, training, & evaluation, and performance summary
  - Tool for future ML research in the same directions
- Construct a web application to visualize model predictions

Technical Details (Data)

Data processing

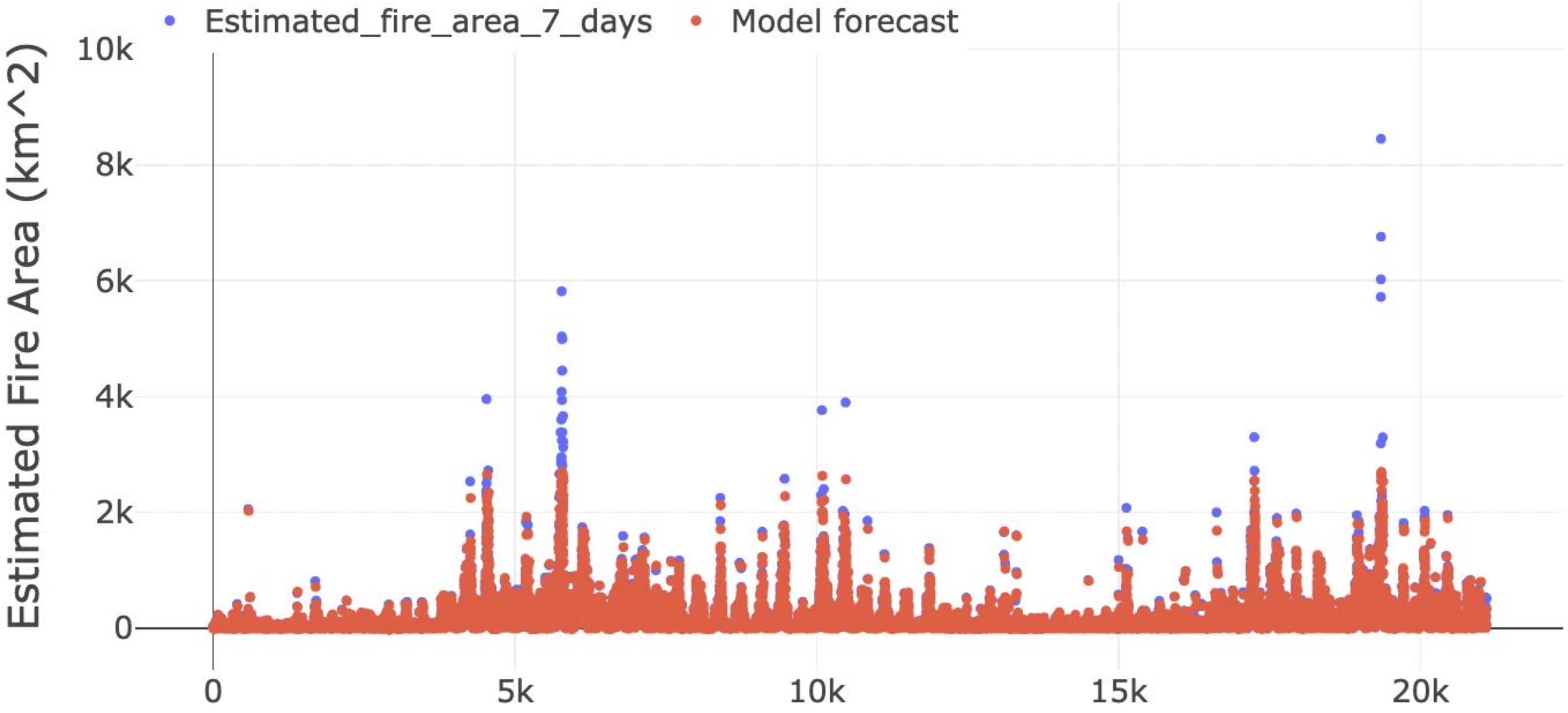
- We used the IBM dataset, which has data from the seven Australian regions
- The IBM climate dataset contains daily records of several climate variables
- The IBM historical fire dataset contains daily wildfire occurrence records
- We combined these datasets, matching the climate data with the corresponding historical wildfire data
- Categorical data such as region names are converted into numerical data for training purposes
- 26406 data points total
    - 20% set aside as the test set

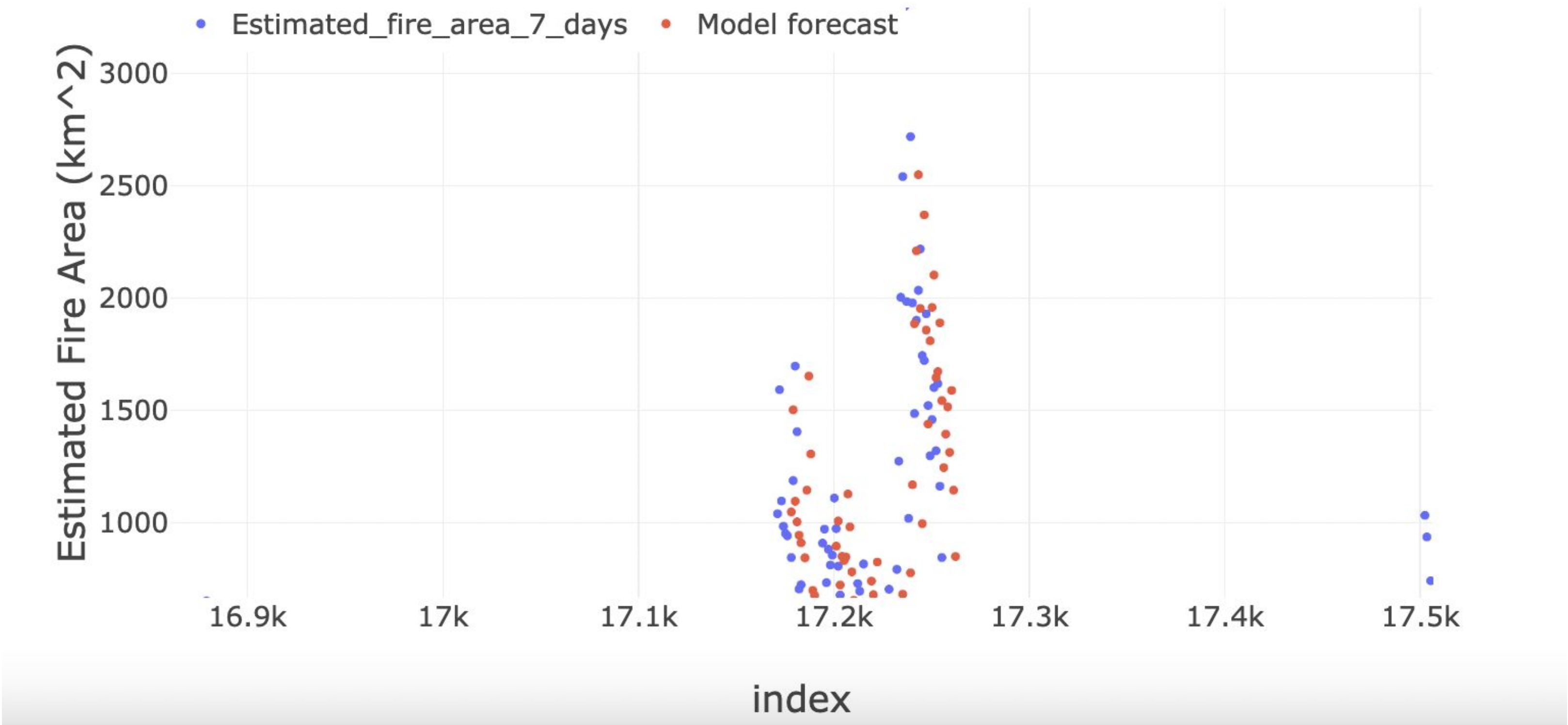
#	Column	Non-Null Count	Dtype
0	Unnamed: 0	21128 non-null	int64
1	Region	21128 non-null	object
2	Date	21128 non-null	object
3	Estimated_fire_area	21128 non-null	float64
4	Mean_estimated_fire_brightness	21128 non-null	float64
5	Mean_estimated_fire_radiative_power	21128 non-null	float64
6	Mean_confidence	21128 non-null	float64
7	Std_confidence	19311 non-null	float64
8	Var_confidence	19311 non-null	float64
9	Count	21128 non-null	int64
10	Replaced	21128 non-null	object
11	Precipitation_min()	21128 non-null	float64
12	Precipitation_max()	21128 non-null	float64
13	Precipitation_mean()	21128 non-null	float64
14	Precipitation_variance()	21128 non-null	float64
15	RelativeHumidity_min()	21114 non-null	float64
16	RelativeHumidity_max()	21114 non-null	float64
17	RelativeHumidity_mean()	21114 non-null	float64
18	RelativeHumidity_variance()	21114 non-null	float64
19	SoilWaterContent_min()	21128 non-null	float64
20	SoilWaterContent_max()	21128 non-null	float64
21	SoilWaterContent_mean()	21128 non-null	float64
22	SoilWaterContent_variance()	21128 non-null	float64
23	SolarRadiation_min()	21125 non-null	float64
24	SolarRadiation_max()	21125 non-null	float64
25	SolarRadiation_mean()	21125 non-null	float64
26	SolarRadiation_variance()	21125 non-null	float64
27	Temperature_min()	21123 non-null	float64
28	Temperature_max()	21123 non-null	float64
29	Temperature_mean()	21123 non-null	float64
30	Temperature_variance()	21123 non-null	float64
31	WindSpeed_min()	21123 non-null	float64
32	WindSpeed_max()	21123 non-null	float64
33	WindSpeed_mean()	21123 non-null	float64
34	WindSpeed_variance()	21123 non-null	float64
35	Region_int	21128 non-null	int64
36	Date_int	21128 non-null	int64

Technical Details (Model)

Model Pipeline

- Combine IBM dataset into usable format
- Process data (mean, std, min, max, variance, selected features, etc.)
- Use PyTorch dataset class
  - Finds valid index within dataset
- Build model (tune hyperparameters here and adjust model outputs)
- Run training loop
- Save model
- Use output for visualization, error calculations, etc.


Estimated_fire_area_7_days
Model forecast
Estimated Fire Area (km^2)
10k
8k
6k
4k
2k
0
0
5k
10k
15k
20k


Estimated_fire_area_7_days
Model forecast
Estimated Fire Area (km^2)
3000
2500
2000
1500
1000
16.9k
17k
17.1k
17.2k
17.3k
17.4k
17.5k
index

Technical Details (Frontend)

Web page

- Locate sample in dataset using the submitted form details
    - User inputs Region, Date, Number of days to be forecasted
- From that sample, choose samples from the previous days (14 days)
- Run collective samples through LSTM as input to generate a prediction
- Drop first sample in the collective samples and add the prediction to the end of the collective samples (generates new input)
- Repeat for the number of days the user wants to forecast

Demo

Challenges

- Domain knowledge gap
  - Spoke with Prof. Naomi Tague, Prof. Max Moritz, PhD researcher Isaac Park
  - Gained insight on environmental factors, related research/work, etc.
- Lack of suitable datasets
  - Current IBM dataset has its own flaws
  - Thanks to Max and Isaac for offering data
- Lack of hardware
  - Thanks to Prof. Tobias Höllerer for allowing us to use servers
- Understanding LSTMs
- Frontend work
  - Only one person could host the frontend on the server at a time

Goals for Next Quarter

- Improve prediction accuracy
- Visualize data on a dynamic map/graphs and revamp the UI
- Train our model with different datasets
- Implement different models (e.g. Transformer)

Questions?