Understanding and Modeling Human Mobility Response to California Wildfires

Abstract

Human movement varies in response to environmental changes, particularly during natural disasters. Our research examined the impact of the Lake Fire, a 2020 wildfire in Los Angles County, on visits to several nearby points of interest (POIs). We devised two methods for determining whether a POI was impacted: (i) using changepoint detection algorithms on clusters of POIs; and (ii) using the total number of anomalies over several mobility metrics to determine impact. We then attempted to predict impact based on geographical location in relation to the fire and the type of POI (e.g. grocery, healthcare, transportation). We found that a logistic regression model trained on labels obtained using changepoint detection applied to POI clusters yielded the best predictive performance. The model suggested that POIs near the fire and/or categorized as historical/nature or public functions were more likely to be impacted.

Statement role

I contributed primarily to the anomaly detection method for identifying impacted POIs. In addition, I acted as the notetaker of the group. I took detailed notes of our biweekly meetings, one with our advisors and one with just the group members.

Project overview

This project looked at the impact of California wildfires on human movement patterns, specifically focusing on Lake Fire in Los Angeles County from August to September 2020. We utilized data from Safegraph and Mapbox containing information on movement patterns at places of interest (POIs) near Lake Fire.

One of the main questions that our team sought out to answer was “What types of POIs tend to be impacted by wildfires?” Since we didn’t have any labels indicating which POIs were impacted, we created our own using two different methods. Method I consisted of clustering groups of POIs with similar characteristics (e.g., location, movement patterns), then labeling each cluster as impacted if a changepoint occurred during the fire. Method II looked at each POI individually, adding up the total number of anomalies across three mobility metrics and labeling the POI as impacted if that total was greater than a predetermined threshold.

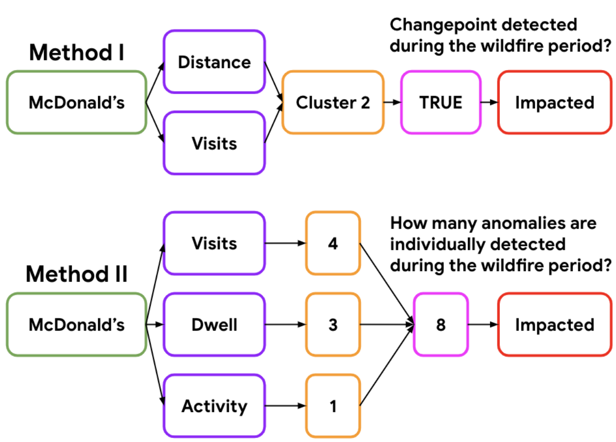

Figure 1: Methods I (top) and II (bottom) for labeling the POIs.

We then trained several machine learning models on these labels using geographical location (e.g., distance and angle from fire) and the type (e.g. grocery, healthcare, transportation) of each POI as predictors. Using a logistic regression model on the labels obtained from Method I yielded the best performance, achieving a test accuracy of 0.75 and a ROC-AUC value of 0.80. Upon looking at the predictions, this model was more likely to predict POIs closer to the fire and/or categorized as historical/nature or public functions as being impacted.

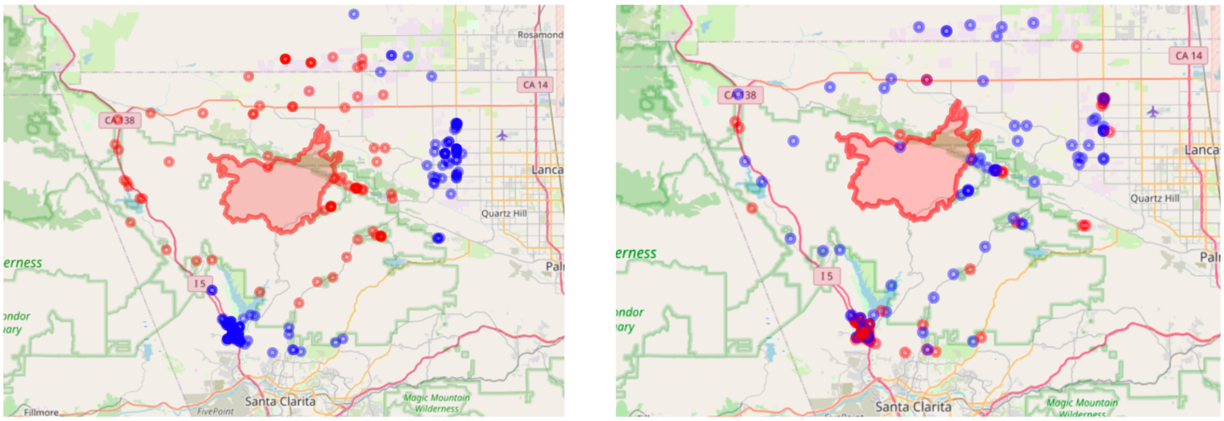

Figure 2: *Spatial plots of the logistic regression model's predictions with Method I (left) and the random forest model's predictions with Method II (right). The red region shows where Lake Fire occurred, and the red and blue circles indicate the impacted and unimpacted POIs, respectively.*

<u>Project outcomes</u>

For our meetings with the advisors, we made slides every week with our progress (including any data visualizations). We created a poster summarizing our project findings over 2 quarters and presented it at UCSB's 2023 Data Science Capstone Showcase.

<u>Personal contributions</u>

For the exploratory part of the analysis, I created 2 Shiny apps in R – one using Leaflet to visualize how the visit counts at each POI change over space and time (see the video here) and the other using anomaly detection to see which individual POIs tended to have an abnormally higher or lower number of visits during the wildfire.

I worked primarily on the anomaly detection pipeline of the project. I experimented with different thresholds to label the POIs as impacted or not using anomaly detection. Based on these labels, I then created several classification models using cross-validation.

As for my role, I took down notes of what was being said during our biweekly meetings and shared them with my group members for convenient reference. To better coordinate tasks, I also created an accountability sheet in Google Sheets where every member documented what tasks they completed on each day, along with the amount of time that they took. This allowed me see what progress we were making as a group and helped with distributing tasks among all of us.