Understanding and Modeling Human Mobility Response to California Wildfires

Lyndsey Umsted
Justin Liu
Piero Trujillo
Ellen Burrell
Laura Baracaldo Lancheros
Trevor Ruiz
lyndseyumsted@ucsb.edu
tdr@ucsb.edu
Department of Statistics and Applied Probability, University of California, Santa Barbara
Santa Barbara, CA, USA

Evgeny Noi
Enbo Zhou
Somayeh Dodge
sdodge@ucsb.edu
Department of Geography, University of California, Santa Barbara
Santa Barbara, CA, USA

ABSTRACT

Wildfires and other natural disasters have substantial impacts on human mobility, but changes in movement patterns can be difficult to study quantitatively due to sparse data and the complexity of interacting spatiotemporal processes. This project is a case study in using self-supervised learning to address this challenge. We leverage smartphone tracking data from multiple sources on activity at several Places of Interest (POIs) surrounding wildfire locations to understand the human mobility response to the Lake Fire in Los Angeles County in 2020. We use spatiotemporal clustering to group POIs based on activity over time and geography, and subsequently apply binary segmentation on aggregated activity data to infer the impactedness of each cluster. We then train a classification model to predict whether a location is impacted by the fire based on location attributes, proximity, and direction. Finally, we demonstrate the superiority of this clustering-based approach over for anomaly detection on a location-by-location basis.

KEYWORDS

wildfires, human mobility, transportation, self-supervised learning, spatiotemporal data analysis

ACM Reference Format:
Lyndsey Umsted, Justin Liu, Piero Trujillo, Ellen Burrell, Laura Baracaldo Lancheros, Trevor Ruiz, Evgeny Noi, Enbo Zhou, and Somayeh Dodge. 2023. Understanding and Modeling Human Mobility Response to California Wildfires. In *Proceedings of KDD'23 Southern California Data Science Day*. ACM, New York, NY, USA, 3 pages.

KDD'23 Southern California Data Science Day, August 7, 2023, Long Beach, CA

1 INTRODUCTION

Studying human movement allows us to understand behavioral responses to large-scale environmental changes such as natural disasters. With the increasing rate of extreme events such as wildfires, hurricanes and storms, and flooding, it is important to investigate the impact on behavior and understand urban resilience in response to these devastating events. Studies show that mobility patterns can be used as a marker to study disaster-induced changes in human behavior and dynamics of the cities, and can generate valuable insights about the resilience of communities [5]. For example, Hong et al. [3] used anonymized smartphone traces of 35% of the population in Houston, Texas to identify community resilience capacity in response to Hurricane Harvey in 2017. Ghurye et al. [2] proposed a framework to model human behavior before and during disaster using analysis of Call Record Data (CDR) to establish a Markov Chain model for normal mobility and analyze changes during a flooding event in Rwanda. These studies highlight the promising advantage of data-driven approaches using large scale mobile traces for disaster mitigation and recovery planning.

In this research, we quantify the impact of wildfires on human movement patterns using self-supervised learning. We present an analysis of the Lake Fire, which occurred in August of 2020 in the Angeles National Forest of Los Angeles County, as a case study. We utilize smartphone location tracking data aggregated geographically to visit counts across a large collection of Points of Interest (POIs). We cluster POIs and apply change-point detection methods cluster-wise to identify shifts in the data. We use the change-points to auto-label locations as impacted or not, and develop a model to predict impactedness based on geography and location attributes. We compare this clustering-based approach with a POI-by-POI analysis, and demonstrate that the clustering-based method produced both better predictions and more interpretable results.

2 MOBILITY AND WILDFIRE DATA SETS

The geospatial coordinates of the fire burn edges are publicly available from the California Department of Forestry and Fire Protection's Fire and Resource Assessment Program (FRAP) [6]. Movement patterns during the time period of interest are partially captured among proprietary data from SafeGraph and MapBox. These data are collected via mobile phone applications and other GPS devices that allow location tracking by user consent. We use two datasets capturing activity near the fire region at different spatiotemporal resolutions. The first dataset consists of counts of weekly visits to locations with high levels of human traffic such as retail stores, grocery stores, restaurants, and the like, along with location coordinates and location type. The second dataset includes a spatially aggregated daily activity index that measures the total amount of human activity within pre-defined areas.

We extract latitude and longitude, POI ID, visit counts, and type of POI (categories include grocery, gasoline, public recreation, etc.) from the first dataset and append the activity index from the second dataset after appropriate temporal aggregation and spatial alignment. We then calculate the distance of each POI from the fire's center using the haversine formula and the angular direction from the fire.

3 SELF-SUPERVISED LEARNING FOR INFERENCE OF WILDFIRE IMPACT

Our overall strategy is to auto-generate impactedness labels for each location based on shifts in the visit count and activity level data, and then train a classifier to predict impactedness based on geographic relation to the fire center and location type.

We first use spatio-temporal density-based clustering [1] to group POIs based on similarities in geography and activity level. We then identify geographically localized clusters of POIs with similar human activity between January of 2019 and January of 2021. Using these clusters, we aggregate and detrend the time series of each POI by cluster and use binary segmentation [4] to detect change-points – times associated with shifts in mean and variance – in the aggregated activity series for each cluster. If a change-point is detected during the time period of the fire, and not in the previous year, every POI within that cluster is labeled impacted. If no change-point is detected or a change-point is detected in the previous year at the same time, every POI within the cluster is labeled not impacted.

Using the labeled data, we train a logistic regression model to estimate the probability of impact based on distance and direction from the fire and type of POI (grocery, gasoline, public recreation, etc.). Distance and angle measurements are adjusted based on fire area and burn direction for a generalizable model formulation potentially applicable to the analysis of multiple fires. Adjusted distance is calculated as the distance to the fire's edge (rather than its center); angle is adjusted by 225 degrees so that the reference vector points southward in the direction of burn. Predictive accuracy was estimated using 5-fold cross validation.

4 RESULTS

Our approach classifies POI impact with an overall accuracy of 75% and an area under the ROC curve of 0.80. According to the model, locations within 20 kilometers of the fire edge and locations in the direction of the burn are most likely to be impacted. The types of locations most likely to be impacted are Historical/Nature, Public Functions, and Grocery. Compared with modeling POI impact on an individual (rather than cluster-wise) basis, and find that the cluster-based approach yields (i) predictive probabilities with more cogent spatial structure and (ii) improved classification accuracy.

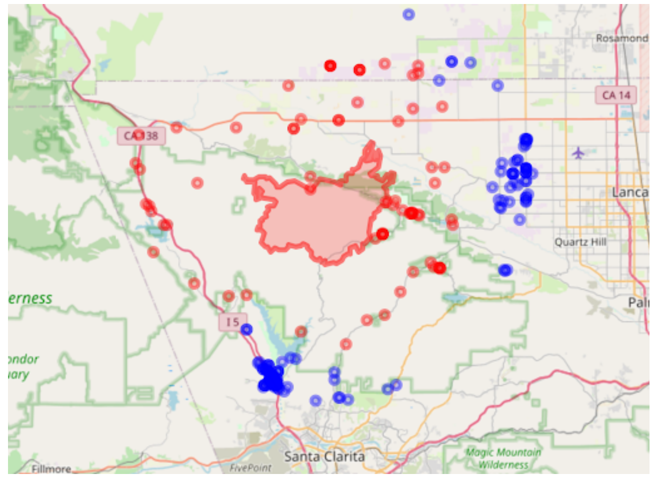

Figure 1: Predicted impact at each POI near the fire; impacted POIs (red) tend to be nearer to the fire edge than unimpacted POIs (blue), and extend farther in the direction of the fire.

5 DISCUSSION

The primary aim of this project is to quantify the human movement response to a wildfire event across several observed locations and predict potential impact based on conveniently available data from multiple sources. We take a self-supervised learning approach comprising: (i) generating impactedness labels based on spatiotemporal clustering, aggregation, and change-point detection; and (ii) training a classifier to predict impactedness. The predictions from the model allow us to understand what areas are at most risk for an influx or outflux of human activity during wildfire events, and using our preferred method we identified higher risk of impact (i) near the fire, (ii) in the direction of burn, and (iii) at location types expected to be sensitive to human mobility response.

We note that the COVID-19 pandemic substantially complicates inference of impactedness; our method does not distinguish impact due to the fire from impact due to COVID-19. The confounding of these two events in our case study likely diminishes the predictive performance of our model.

This research could be extended by augmenting the training data to span multiple fires and incorporating additional geographical features (e.g., wind direction, elevation, air quality, etc.). A more robust model trained on data from multiple fires could potentially be used to generate spatial contours of impact risk based on fire attributes and geography alone; such an extension of our work has the potential to support assessment and estimation of wildfire risk on movement flows in Californian communities, which could in turn improve disaster response planning.

ACKNOWLEDGMENTS

This project is supported by NSF Award BCS No. 2043202 and partially supported with a fellowship from NSF under Award No. 1924205. Activity index data provided by ©MapBox.

REFERENCES

[1] Derya Birant and Alp Kut. 2007. ST-DBSCAN: An algorithm for clustering spatial–temporal data. *Data & knowledge engineering* 60, 1 (2007), 208–221.

[2] Jay Ghurye, Gautier Krings, and Vanessa Frias-Martinez. 2016. A framework to model human behavior at large scale during natural disasters. In *2016 17th IEEE International Conference on Mobile Data Management (MDM)*, Vol. 1. IEEE, 18–27.

[3] Boyeong Hong, Bartosz J Bonczak, Arpit Gupta, and Constantine E Kontokosta. 2021. Measuring inequality in community resilience to natural disasters using large-scale mobility data. *Nature communications* 12, 1 (2021), 1870.

[4] Rebecca Killick and Idris Eckley. 2014. changepoint: An R package for changepoint analysis. *Journal of statistical software* 58, 3 (2014), 1–19.

[5] L Manawadu and VPIS Wijeratne. 2022. Human mobility response to natural disasters and environmental change. In *Climate Change, Disaster and Adaptations: Contextualising Human Responses to Ecological Change*. Springer, 229–242.

[6] California Department of Forestry and Fire Protection, Fire Resource Assessment Program. 2023. Historic Fire Perimeters. https://www.fire.ca.gov/what-we-do/fire-resource-assessment-program/fire-perimeters. Accessed January 2023.

Received 6 July 2023