Harnessing Social Sensing: Real-time Wildfire Detection through Social Media Analysis and Geotagging Inference

Yachay.ai

Alina Pak[1], Olivia Li[2]

[1]Yachay.ai, [2]Spyro Labs

Spyro Labs

Overview

Rapid identification of wildfires is a critical tool for disaster management. However, current methodologies, such as satellite and camera-based data are insufficient for this task, due to the high latency and costs of deployment for hardware tools. Previous work in using social networks for Crisis Informatics lack real-time capabilities and heavily rely on geolocation metadata that is drastically disappearing with new policy changes1.

We propose ViralFire, an ensemble model that uses social media posts to detect wildfires in real-time and infer their locations. Operating in 50 languages, ViralFire boasts a median location error of 15 km—significantly outperforming other text-based geotagging models by 2,000 km—and delivers alerts 4 hours faster than local systems.

Data

- TWWORLD[2] containing 12 million geotagged tweets
- CrisisMMD Twitter Dataset[3]
- Kaggle Datasets
- CrisisLex[4]

Machine Learning

Event Detection

1. Topics detection on the disaster-related corpus: topic modeling methods, such as Latent Dirichlet Allocation and Latent Semantic Analysis, were applied to the corpus to identify potential signals of interest.
2. Binary classification model was trained on derived signals to identify social media posts of interest

Geolocation

ByT5-based text-based geolocation model to identify the longitude and latitude of the user's location.

For noise resilience, the model employs hierarchical agglomerate clusterization to perform classification rather than regression.

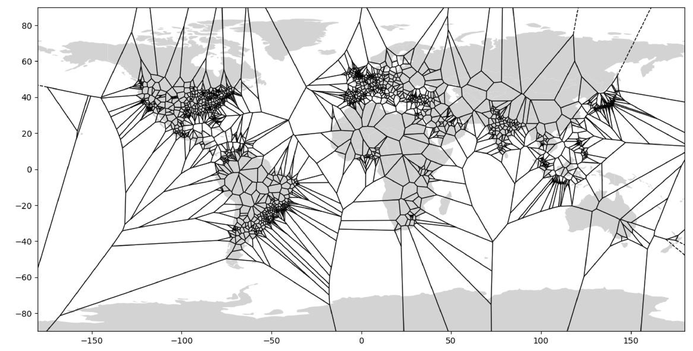

Fig. 1 Hierarchical agglomerative clustering

Case Studies

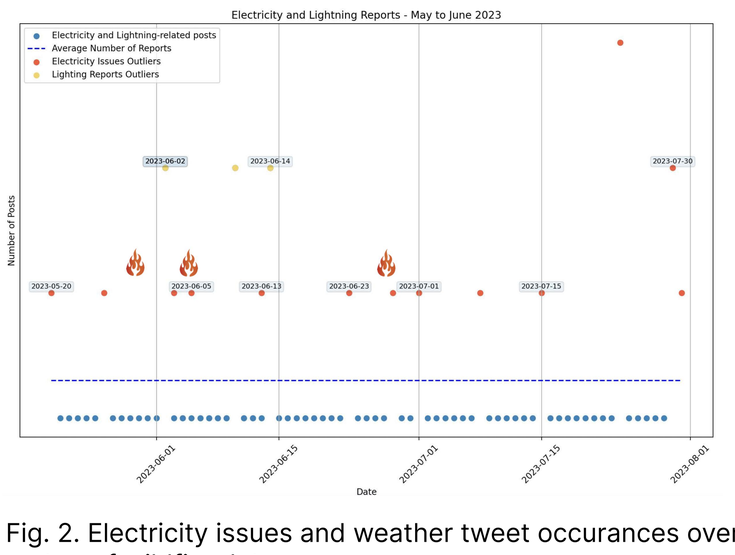

Fig. 2. Electricity issues and weather tweet occurances overlaid on top of wildfire dates

Fresno Fires 2023

By analyzing temporal spikes in specific signals, ViralFire can offer a dynamic tool for identifying emerging risks and preventive measures. CalFire listed the official causes of the Fresno 2023 fires as unknown.

An analysis of user-generated content before and during the time period of the fire saw a noticeable increase of mentions about electricity issues, hot, dry conditions, and lightning strike coinciding with the Pistol, Crane and Panoche fires.

SCU Lightning Complex & LNU Lightning Complex 2020

CAL FIRE reported the SCU Fire on August 16, 2020, at approximately 9:25am. ViralFire was able to capture early tweets relating to smoke, fire, and lightning as early as 5am.

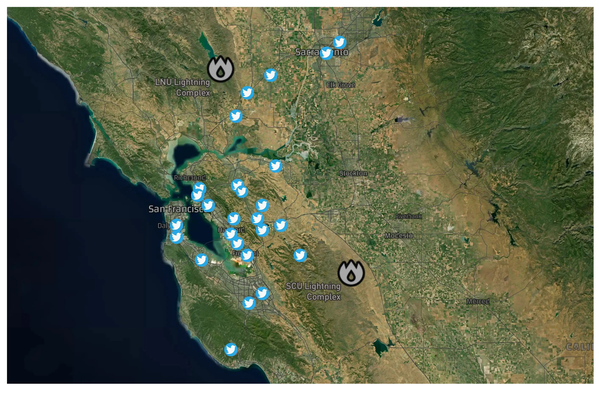

Fig. 3. Geotagged tweets and their relation to the SCU and LNU Lightning Complex Wildfires

References

1. Reuter, C., Hughes, A. L., & Kaufhold, M. A. (2018). Social media in crisis management: An evaluation and analysis of crisis informatics research. International Journal of Human–Computer Interaction, 34(4), 280-294. Chicago
2. Han, B., Cook, P., & Baldwin, T. (2014). Text-based twitter user geolocation prediction. Journal of Artificial Intelligence Research, 49, 451-500. Chicago
3. https://crisisnlp.qcri.org/crisismmd
4. https://crisislex.org/data-collections.html#CrisisLexT26