

Coffee Break Experiment #2

Charles Hutchinson, Sam Kalman

Due 2022-11-20

Introduction

One of North America’s most noticeable symptoms of climate change has been a dramatic increase in the volume and intensity of wildfires, particularly in the western regions. August and September were previously defined by clear skies and fair weather, but fire and smoke have become a recurrent factor in this area. According to the National Interagency Fire Center (NIFC), more acreage burned nationally in each of the last seven years than any single year but one in the first ten years this data was collected (1983-1993, ex. 1988).¹ We sought to analyze if and how public sentiment has shifted over time with respect to these cataclysmic events, with the hopes of gaining some insight into public perceptions of climate change.

Data

We analyzed public sentiment towards wildfires by pulling publicly available tweets through Twitter’s API. We chose to pull tweets made around the time of four of the largest, most recent wildfires in California. These events were the 2021 Dixie Fire, the 2020 Bay Area Fires, the 2018 Camp Fire, and the 2017 Tubbs Fire.

It was difficult to find relevant data. Tweets that were well-tagged (e.g., including #DixieFire for the Dixie Fire) tended to be associated with news organizations, which were overwhelmingly neutral and rarely “organic” reactions. However, loosening the search parameters inevitably lowered the data quality (e.g., searching for “smoke” mostly returned tweets referencing marijuana usage). It was especially challenging to develop searches with consistent methodology across separate wildfires, while still getting useful data.

We found our best results using the following approach:

- Use the commonly accepted fire name and “fire” as search terms. For example, the search terms for the 2021 Dixie Fire would simply be “dixie fire”.
- Limit the search results to English. Creating sentiment dictionaries in other languages was out of scope for this project.
- Remove retweets.
- Limit the tweets to a month time frame. Some fires lasted longer than this, so we picked the interval in which the smoke was most impactful.

The size of each fire’s corpus was unfortunately still far from uniform, but we found this to be a reasonable compromise among a set of unappealing options.

Table 1: Relevant Tweets per Recent Wildfire.

Fire	Tweets
Tubbs Fire	3530
Camp Fire	50469
Bay Area Fires	4680
Dixie Fire	26139

¹See <https://www.nifc.gov/fire-information/statistics/wildfires>.

Methods

To better understand how perception of climate change has changed over time, a sentiment analysis algorithm from the `syuzhet`² package was applied to the 84,818 tweets across the 4 major fires of interest. The `syuzhet` package returns sentiment scores in the degree of positively/negativity using a sentiment dictionary developed in the Nebraska Literacy Labs. All links and web addresses were removed from the tweets as well as twitter handles, indicating that the tweet was mentioning or replying to another user. Each tweet was given a sentiment score. Scores above 0 were considered positive, below 0 were negative, and equal to 0 were considered a neutral tweet. Any tweets from twitter accounts with the words “news” or “meteorology” in their biographies were filtered out. We found that most news organization tweets contained neutral, non-opinionated information about the respective wildfire. To control for twitter bots adding unnecessary noise to the data, the users needed to have at least 1 follower and the tweet must be interacted with (like/retweet/comment) at least once.

Results

In Figure 1 below, the sentiment classification breakdown is plotted for each of the 4 wildfires. To account for the tweet volume difference between the four fires, percentages of sentiment type is calculated based on the total number of tweets pertaining to a respective fire. The fire are in chronological order, beginning with Tubbs Fire and ending with the Dixie Fire.

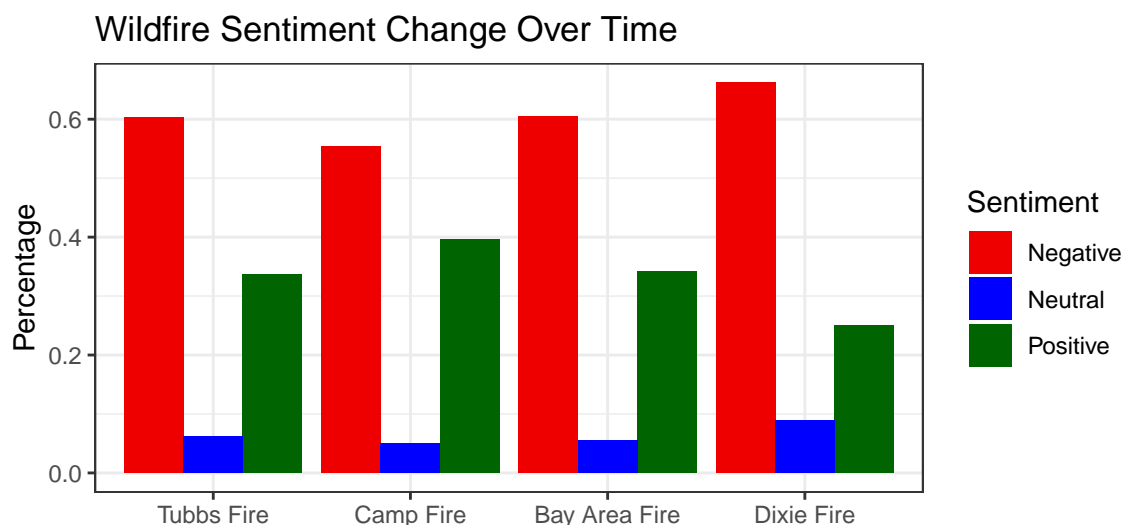


Figure 1: Shift in Wildfire Sentiment Over Time.

Even though the tweet volume is not consistent throughout time, figure 1 gives us preliminary evidence that there may be a shift in tweet sentiment as we observe more and more of these natural disasters. Beginning with Camp Fire, the amount of negative tweets has been steadily increasing for each wildfire.

Discussion

²See <https://www.rdocumentation.org/packages/syuzhet/versions/1.0.6>