

INSERT

Charles Hutchinson, Sam Kalman

December 12, 2022

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 its 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 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” returned a large sample of 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, possibly due to unequal media coverage, but we found this to be a reasonable compromise among a set of unappealing options. We found both Camp Fire and Dixie Fire generated far more content on Twitter under our listed constraints when compared to Tubbs Fire and the Bay Area Fires. Table 1 shows the volume of tweets per fire.

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

Table 1: Relevant Tweets per Recent Wildfire.

Fire	Tweets	Words
Tubbs Fire	3530	51187
Camp Fire	50469	1021221
Bay Area Fires	4680	129082
Dixie Fire	26139	577785

Our twitter license included access to a range of metadata. We chose to analyze tweet level information (specifically Likes, Retweet, Quote Tweets, and Date), as well as certain user level information (specifically Followers, Following, Biography, and Twitter Name).

Filtering for tweets that explicitly mentioned the fire by name significantly limited the total number of results. This proved to be particularly debilitating when working on location-based analyses, since only a subset of that limited population reported their location while tweeting. To alleviate this problem, we chose to use much more generic search terms. For the purposes of this paper, we found the search term “fire” tended to return the most relevant results (as compared to similarly broad terms like “smoke” or “burn”), though this was not a systematic decision. Nonetheless, it is hard to argue that “fire” is not an appropriate search term to use when conducting sentiment analysis on wildfires.

We continued to use the same date windows for each fire. As before, we only pulled English tweets, and filtered out any retweets. We asked the Twitter API to filter for tweets in the United States, then manually filtered those tweets down to those made in the contiguous lower 48 with longitude and latitude filters. Twitter provides locations in a bounding box, not on a single point, so we assumed that each tweet was made in the center of the bounding box. This was somewhat problematic when dealing with state-size bounding boxes, but the vast majority of locations were given on the city level, with much smaller bounding boxes.

Table 2: Location-Enabled Tweet Counts

Fire	Tweets	Words
Dixie	36353	749509
Bay Area	71608	1468934
Camp	84284	1746368
Tubbs	76839	1165880

Methods

We used two sentiment analysis algorithms of varying complexity to classify the aforementioned tweets. The first algorithm we used was from the `syuzhet` package² in R. The `syuzhet` package returns sentiment scores on a spectrum of positivity/negativity using a sentiment dictionary developed in the Nebraska Literacy Labs. We deemed any sentiment score that was greater than zero to be positive, less than zero, negative, and exactly equal to zero, neutral. Unfortunately, the `syuzhet` package wasn’t designed or trained on short, declarative statements such as tweets.

To overcome this issue, we decided to use Cardiff’s `roBERTa`-base model for sentiment analysis.³ This model was trained on roughly 124 million tweets from January 2018 to December 2021 and used the `TweetEval` benchmark for fine tuning the sentiment analysis. The algorithm outputs probabilities that a tweet is positive, neutral, or negative. For the purposes of our analysis, we simplified the output to a categorical variable representing the sentiment with the highest probability. (Strength of sentiment was not captured.)

BERT models are trained on both language modeling and next sentence prediction. It is important to note that BERT models generally do not use words as tokens - they use a combination of word-level, segment-level,

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

³<https://huggingface.co/cardiffnlp/twitter-roBERTa-sentiment-latest>

and position-level tokens to encode the input text. Word-level tokens are individual words or sub-words in the input text, segment-level tokens are used to indicate whether a given word is part of the first or second sentence in a pair of sentences (in the case of sentence-pair tasks), and position-level tokens are used to indicate the position of each word within the input text. This may convey some of the complexity of the contextual embedding performed in the model. Regardless, we note that the word counts in Tables 1 and 2 are not accurate token counts for Cardiff’s roBERTa model.

We found that most tweets from news organization contained neutral, non-opinionated information, which wasn’t useful for our research question. Therefore, any tweet that came from a user with the word “meteorology” in their bio was not considered in our analysis. Additionally, in an attempt to filter out bots adding unnecessary noise to the data, we only used tweets from users with at least one follower, and the tweet itself needed at least one external interaction (a like/retweet/comment). This was done consistently before running both of the discussed sentiment algorithms. It’s worth noting that the authors of Cardiff’s roBERTa model (Loureriro et al., 2022) address the bot issue by filtering out the top 1% of users by volume, but we did not have a similarly unbiased data set to determine an appropriate posting cutoff.

To track to sentiment shift through time, we chronologically ordered the fires (Tubbs Fire, Camp Fire, Bay Area Fire, Dixie Fire) and calculated the proportion of each sentiment for each fire. We used proportions because of the unequal distribution of tweets across the fires.

We first converted from longitude/latitude coordinates to WGS84 coordinates, then projected these coordinates using a Lambert Conformal Conic projection (LCC). This process was completed using the `sf` package. [ADD MORE HERE]

Results

In Figure 1, we plot the sentiment classification proportions for each of the four wildfires. The fires are in chronological order, beginning with Tubbs Fire and ending with the Dixie Fire. The error bars reflect a 95% confidence interval for each proportion.

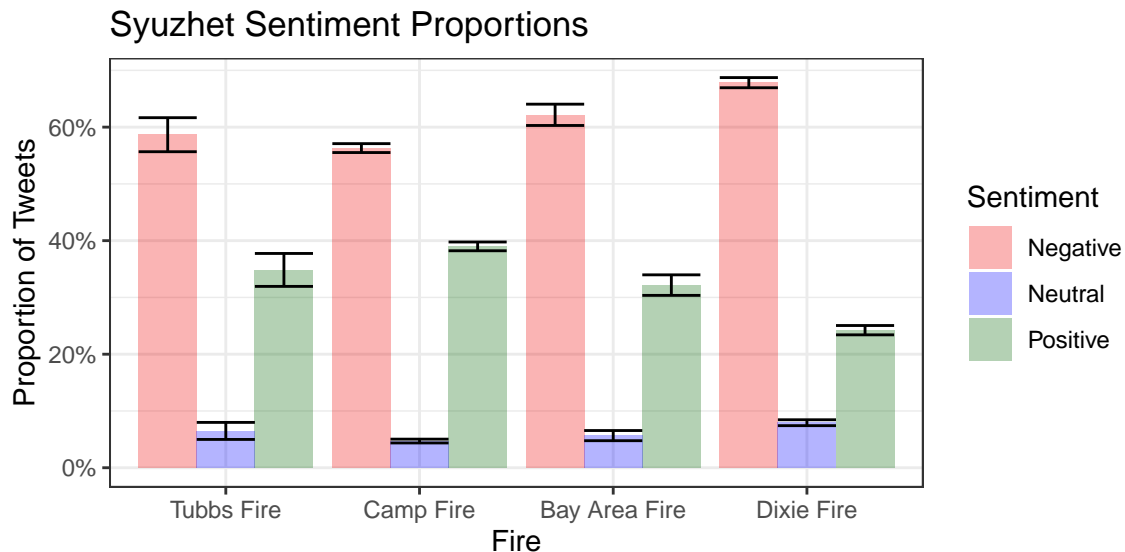


Figure 1: Shift in Wildfire Sentiment Over Time.

There are not many neutral tweets, though our definition of neutrality with the `syuzhet` scoring is quite stringent. The majority of tweets are classified as negative, and the proportion of negativity appears to be increasing through time.

We repeat the same analysis using scores from Cardiff’s roBERTa model in Figure 2.

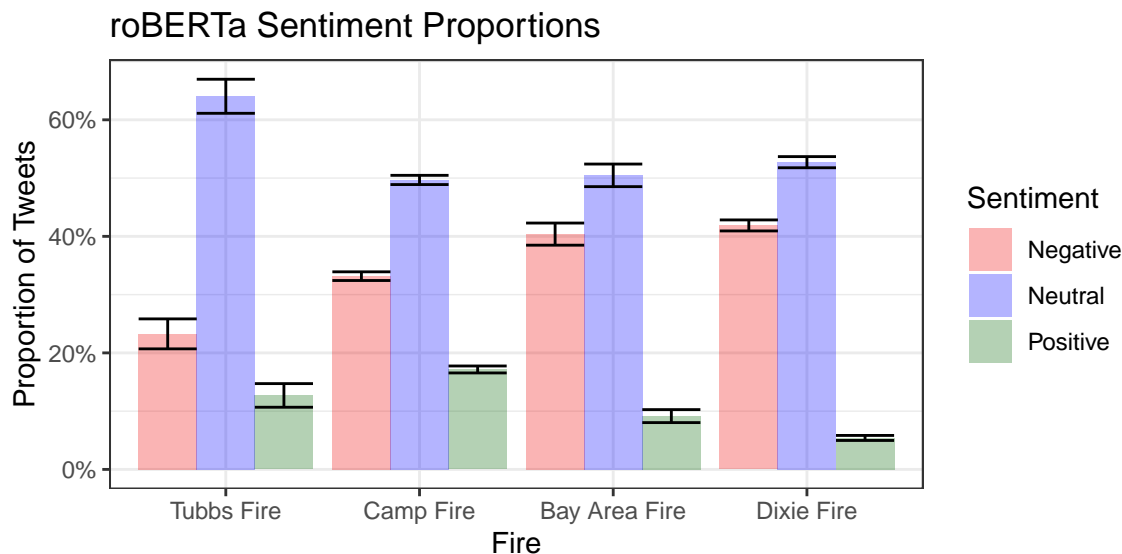


Figure 2: Shift in Wildfire Sentiment Over Time.

The roBERTa model classifies a much larger proportion of the tweets as neutral. In previous iterations of this analysis, it was difficult for us to remove all news related tweets. We found that this iteration was very effective at putting “inorganic” reactions from news organizations in this neutral category. However, the roBERTa model classified less negativity for the Tubbs fire than the syuzhet package and we see a more consistent increase in negativity, plateauing at the Bay Area Fire and Dixie Fire. It was difficult for us to remove all news related tweets, roBERTa model

Let’s examine differences in specific tweet classification between the syuzhet and roBERTa model. Specifically, for the Tubbs fire where we saw less negativity classified by the roBERTa model. It was difficult for us to remove all news related tweets. roBERTa model classified a lot of these as neutral when the syuzhet package would incorrectly assign the tweet as slightly positive or slightly negative.

Examples of neutral for roBERTa and negative for syuzhet

- Where Did the Tubbs Fire Begin? Investigators Scour the Ashes <https://t.co/9z2QjBXi77>
- The cause of the massive Tubbs fire is still unknown – but officials think they may be zeroing in: <https://t.co/0k9w4nFgpE>.

These are examples of truly neutral tweets from news organizations. However, the syuzhet package tends to not assign many neutral tweets. These were incorrectly classified as negative by syuzhet, possibly due to words like “scour” or “unknown”, but correctly seen as neutral from roBERTa.

Examples of negative for roBERTa and positive for syuzhet

- The personal treasures former Raiders great Cliff Branch lost in Santa Rosa fires—a truly sad tale by Bob Padecky. <https://t.co/gIT0AulURq>
- My parents lost everything in the Tubbs Fire yesterday. Anything will help. Please share as much as you can. Thank you so much. #tubbsfire <https://t.co/xFAexYshSL>
- Many Sonic staff in Santa Rosa and surrounding area are affected by the Tubbs Fire. Best wishes to all, and be safe! It is devastating. <https://t.co/pVqAFisxKd>

The first tweet used the word “great” to describe the football player but the overall sentiment should have been negative due to Cliff losing his possessions in the fire. Many tweets consisting of “Thank you” or “Be

safe” were seen as positive from the syuzhet model, when the overall message of the tweet was in a negative context. This subtle distinction was correctly picked up by the roBERTa model.

Examples of positive for roBERTa and negative for syuzhet

- This Friday’s midnight showing the Rocky Horror Picture Show is now a benefit for Tubbs Fire victims: <https://t.co/Su7tOzcnfm>
- All I can say is wow. Props to Berkeley #firefighters. Crazy video of firefighters stunned by scope of Tubbs Fire <https://t.co/8kZwh6g5uy>

These two examples highlight how the syuzhet package was tricked into a negative classification when the tweet should have been seen as positive. The describing of the movie has some words associated with negativity but the overall message of tweet was a benefit for Tubbs Fire victim, this was correctly picked up on by roBERTa. The second tweet is a user thanking the firefighters, the fact that she was stunned by the scope of the fire made syuzhet believe the tweet was negative. However, she tweeted this to give props to the firefighters. These examples highlight how the roBERTa model was more effective at distinguishing subtle nuances in tweets, resulting in more accurate classifications of sentiment.

Important to note that Tubbs was much shorter, and the other summer fires were usually coincident with a broader wildfire season. [ADD MORE HERE]

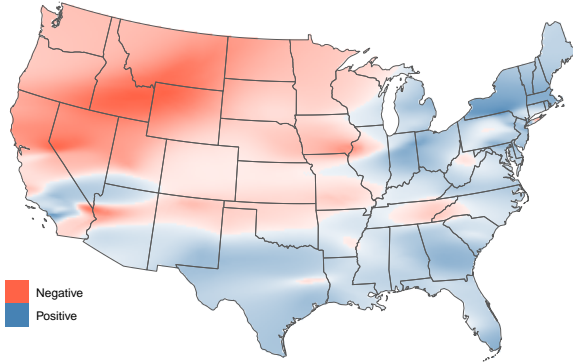
Discussion

Even though the tweet volume was not consistent throughout time, Figure 1 gives us preliminary evidence that there may have been a shift in tweet sentiment towards wildfires in the last five years. Beginning with Camp Fire, the amount of negative tweets significantly increased for each wildfire.

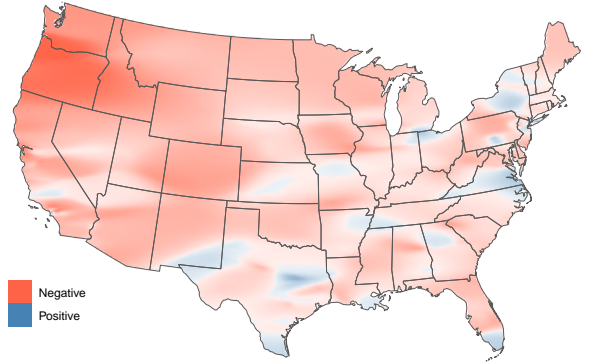
As mentioned previously, data quality was the limiting factor for this analysis. It was difficult to find a set of relevant tweets using a consistent methodology for each fire. Finding a balance between organic and relevant reactions ended up requiring more judgment than we would have liked.

It was surprising that we could not find more tweets for the Bay Area and Tubbs fires, given their proximity to major urban centers. We originally suspected that the Camp Fire tweet set contained many false positives, but it appears that fire generated more engagement than expected due to the almost total devastation (and subsequent press coverage) of the towns of Paradise, Concow, Magalia, and Butte Creek Canyon. The Dixie Fire was similarly destructive. As the largest ever single wildfire in the state’s history, part of its impact included popular Lassen National Park, the coverage of which likely increased engagement on Twitter.

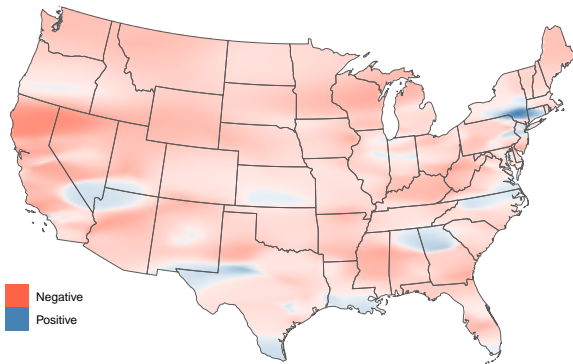
Our next area of research will be studying location-based elements of the wildfire tweets. Data quality issues were compounded in our initial attempts to collect this data, since only a subset of Twitter users enable this functionality. However, knowing the location of each tweet will allow us to better understand to which extent these changes in sentiment were localized.



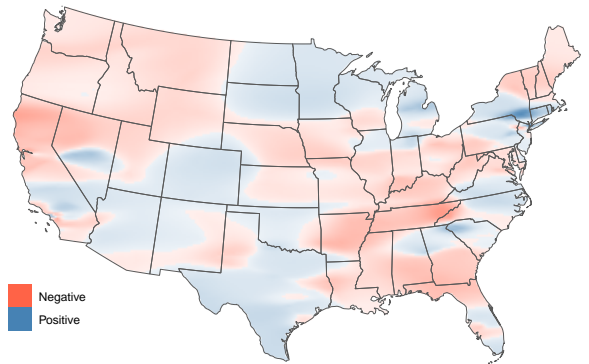
(a) Dixie (2021)



(b) Bay Area (2020)



(c) Camp (2018)



(d) Tubbs (2017)

Figure 3: Tweet Sentiment (roBERTa) by Location