Cognitive Computation (2024) 16:1531–1553
https://doi.org/10.1007/s12559-023-10195-8

Neurosymbolic AI for Mining Public Opinions about Wildfires

Cuc Duong[1,2] · Vethavikashini Chithrra Raghuram[3] · Amos Lee[4] · Rui Mao[1] · Gianmarco Mengaldo[4] · Erik Cambria[1]

Received: 10 October 2022 / Accepted: 19 August 2023 / Published online: 14 September 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Wildfires are among the most threatening hazards to life, property, well-being, and the environment. Studying public opinions about wildfires can help monitor the perception of the impacted communities. Nevertheless, wildfire research is relatively limited compared to other climate-related hazards. This article presents our data mining work on public opinions about wildfires in Australia from 2014 to 2021. Three key aspects are analyzed: the topic of concern, sentiment polarization, and perceived emotions. We propose a data filtering approach to acquire golden samples to train a supervised model for emotion quantification to achieve the last target. The results show that the new model produces a more accurate emotion estimation than the existing lexicon approach. Through data analysis, we find that people have seen wildfires as one of the impacts of climate change; trends of tweets can reflect the damage of wildfires in real life.

Keywords Neurosymbolic AI · Sentiment analysis · Wildfires

Introduction

Human-induced global warming is causing a significant increase in the number and intensity of extreme weather events [1]. Among these, extreme drought conditions and heatwaves are of particular interest. Indeed, these weather patterns may create conditions prone to wildfires, one of the most threatening hazards to life [2], property [3], well-being [4], and the environment [5]. However, wildfires have been relatively understudied in the literature compared to other climate-related hazards [6].

Studying public opinions about wildfires can help monitor the perception of the impacted communities, which is beneficial for implementing effective adaptation strategies [6]. Analyzing public perceptions with big data has been an important research method in cognitive computation [7, 8]. The recent advances in Artificial Intelligence (AI) in this field have been a significant source of inspiration for this project. Using machine learning algorithms can help identify patterns in public opinion and detect changes in sentiment and emotion over time [9, 10]. Furthermore, natural language processing techniques can provide a deeper understanding of public opinions by extracting the underlying topics and themes.

Hence, in this pilot study, we analyze (i) the topic of concern, (ii) sentiment polarization (positive or negative opinion about the entity), and (iii) perceived emotions regarding wildfires in Australia in the period 2014

Cuc Duong and Vethavikashini Chithrra Raghuram contributed equally to this work.

✉ Rui Mao
rui.mao@ntu.edu.sg

Cuc Duong
duongthi001@e.ntu.edu.sg

Vethavikashini Chithrra Raghuram
v.vetha@iitg.ac.in

Amos Lee
amoslee@u.nus.edu

Gianmarco Mengaldo
mpegim@nus.edu.sg

Erik Cambria
cambria@ntu.edu.sg

[1] School of Computer Science and Engineering, Nanyang Technological University, 50 Nanyang Ave 639798, Singapore

[2] Interdisciplinary Graduate Program, Nanyang Environment & Water Research Institute, 1 Cleantech Loop, Singapore

[3] Department of Physics, Indian Institute of Technology, Guwahati 781039, India

[4] Department of Mechanical Engineering, National University of Singapore, 21 Lower Kent Ridge Road 117575, Singapore

Table 1 The most frequent hashtags for each year

Rank	2014	2015	2016	2017
1	Bangor	Koalamittens	**climatechange**	Australia
2	WAM	Sampson_Flat	WaroonaFire	**climatechange**
3	**climate**	WAM	Waroona	heatwave
4	EWM	Adelaide	**ClimateChange**	Heatwave
5	WhitemanPark	EWM	emissions	NSWfires
6	Australia	Australia	biodiversity	australia
7	BANGOR	SAFires	Australia	buildings
8	vicfires	**climate**	WAfires	Himawari
9	em2au	alert	SFDRR	Tasmania
10	heatwave	fire	News	damaged
Rank	**2018**	**2019**	**2020**	**2021**
1	auspol	auspol	Australia	Morrison
2	qldpol	Australia	auspol	Adelaide
3	**ClimateChange**	**ClimateEmergency**	australia	Australia
4	nswpol	**ClimateChange**	**ClimateEmergency**	PapuaNewGuinea
5	Drought	NSWfires	NSWfires	endangered
6	Australia	**climatechange**	arson	FarNorthQueensland
7	StopAdani	bushfiresNSW	**climatechange**	auspol
8	**climatechange**	NSW	**ClimateChange**	australia
9	heatwave	**ClimateCrisis**	AustraliaFires	politics
10	Sentinel	qldpol	animals	conservative

– 2021. We leverage Twitter,[1] which provides free access to their entire archive to acquire raw opinions for academic research.

Since data on Twitter are unstructured, there are three main challenges in extracting the information listed above. First, in the topic of concern task, an unsupervised approach such as the Latent Dirichlet Allocation (LDA) [11] is a well-known method for topic clustering as it is easy to implement and does not require labeled data. However, recent works showed that this method's outcomes are rarely meaningful [12]. To this end, a keyword-based approach is more helpful for finding popular topics. Hence, we present the results of this task via a hashtag[2] ranking approach (Table 1).

Second, quantifying sentiment polarity and emotion from the text component of a tweet[3] remains challenging. Methods to tackle these two tasks come from the natural language processing field, where symbolic approaches such as lexicon are convenient to apply but inefficient (see Table 2). Supervised learning is a promising methodology to address the tasks [13]; however, there is no open wildfire-related sentiment and emotion dataset. In this work, we propose to use ClimateTweet [12], a closely related dataset to wildfires, to quantify sentiment polarity. Table 3 shows that the dataset contains relevant tweets helpful in training a sentiment detection model. For the emotion quantification task, we propose using the neurosymbolic AI approach, integrating symbolic knowledge and neural networks to achieve more robust outcomes. In detail, multiple models (i.e., symbolic, commonsense, and neural networks) would be combined to extract highly credible samples to train a supervised model. Our experimental results show that the supervised model significantly improves task outcomes compared to the existing lexicon approach.

Third, the location information is not always mentioned in a social media post, which causes difficulty in geographical analysis. "Data and Method" section shows that approximately 20% of Twitter's data concerning wildfires have detectable locations. In this work, we overcome this problem by combining the known-location data with the most frequent topics to make unbiased conclusions.

In summary, the main contributions of this work are:

1. Trained a supervised learning model for the sentiment detection task
2. Implemented a hybrid approach to extract golden samples to train another supervised learning model for the emotion quantification task.
3. Identified popular topics concerning wildfires in Australia
4. Visualized the public's sentiment and emotion trends from 2014–2021

[1] https://twitter.com/

[2] hashtags are user-emphasized keywords in Twitter.

[3] tweet is a post on Twitter.

Table 2 Ablation studies on NRC lexicon to show that some samples are incorrectly predicted

No.	Sentences	NRC pol.	NRC emo.	BERT-FC	Sentic-Net
1	'Great work by our fires {Name} advises threat posed by bushfire at {Place} in the {Place} has reduced {url}'	Neg (0.33)	Anger (0.33)	Pos	Pos
2	'My beautiful country is still burning bushfires across {Place} so sad and scary for those living in the country'	Pos (0.5)	Joy (0.5)	Neg	Neg

After reviewing the related works in "Related Works" section, we will elaborate on our findings and contributions in "Data and Method" section.

Related Works

Wildfires Opinion Mining

Significant work on tweets' analysis focused on hazards has been done recently. Kirilenko and Stepchenkova [14] studied tweets about climate change in 2012 and 2013 and found that some events that ignited an increase in tweeting activity relating to climate change were Hurricane Sandy, COP-18, and United Nations conferences. Kirilenko et al. [15] showed that the local temperature anomalies can also spike up discussion on Twitter. However, there are few studies concerning wildfires. Hence, we conduct a trending analysis for Australian wildfires in this study. Similar to [16], we aim to investigate the trends by volume analysis, topic modeling, and sentiment analysis. In addition, the emotion analysis is performed to add an extra dimension to understanding the public's opinions. Willson et al. [17] explored the themes and nature of sentiment of Twitter content associated with the Australian wildfires from 2019–2020. Differently, our work covers a more extended range and more statistics to investigate the emotional aspects of the opinions.

Sentiment and Emotion Analysis

Recognizing and quantifying sentiment and emotion from text are active research branches in natural language processing. Over the last two decades, the research developments could be classified into two main directions: supervised and unsupervised learning. In unsupervised learning, lexicon-based analysis is the leading approach. The method employs rules and vocabularies with sentiment polarity and emotion scores to determine the text's overall sentiment and emotion scores. The prompt-based method is another technological trend in unsupervised affective computing [18–20], where lexicon knowledge is used for label-word mappings. Popular English lexicons, e.g., WordNet Affect [21], SentiWordNet [22], SenticNet [23], and NRC lexicon [24], have been widely employed. We choose the NRC lexicon as the benchmark for our proposed emotion quantification model. It has the largest vocabulary size (14182 words) among these lexicons, and its scores are human-labeled with crowd-sourcing.

One major drawback of the lexicon-based analysis is that the entire textual content may not match the scores computed from individual words. The first example from Table 2 shows that even though the word '*threat*' was mentioned in the sentence, the more prominent polarity/emotion is *joy/positive* rather than *anger/negative*, as predicted by the NRC lexicon. To tackle this issue, we propose to use a supervised learning approach to capture the meaning of the entire text before assigning the polarity and emotion scores for the text.

Supervised learning (including transfer learning) requires a labeled dataset to train (or qualify) a neural network to predict the polarity and emotion scores. Open pre-trained word embeddings such as Word2Vec [25], GloVe [26], and BERT [27] have ignited developments in many applications [28, 29]. A relatively small domain-specific labeled dataset can be used with these embeddings to create an efficient sentiment [30] and emotional predictor [31]. However,

Table 3 ClimateTweets contains fire-related tweets

No.	Sentences
1	'{Place} was on fire, terrible black summer bushfires supercharged by climate change'
2	'New @{Name} study predicts danger of wildfires to increase in {Place} in a few decades' time.'
3	'Researchers say that the tiny particles released in #wildfire smoke are up to 10 times more harmful to humans than particles released from other sources, such as car exhaust.'
4	'{Place} is on the verge of a dramatic and devastating fire season due to #climatechange {emoji:frowning face}'
5	'Significant #heatwave temperatures for {Place} over this weekend. #staycool keep hydrated, leave water for pets and wildlife.'

such a labeled dataset does not exist in the wildfire domain. We use the ClimateTweets dataset, which contains many wildfire-related samples with polarity labels, for the sentiment analysis task to overcome this issue. We propose a hybrid approach to filter noises from the NRC lexicon results for the emotion analysis by employing the pre-trained network SenticNet [23], a commonsense-based neurosymbolic AI framework, and our pre-trained polarity model. We call the outcomes the golden samples since they are more reliable than the previous population. The hybrid approach has demonstrated that the emotion model trained by these golden samples performs better than the NRC lexicon.

Data and Method

Data

The data for this analysis were taken from the Twitter engine using the Academic Research **A**pplication **P**rogramming **I**nterface[4] (API). Twitter was chosen as the data source because it has been frequently used for opinion mining in different domains [32–34]. More than 2.5 million tweets were scrapped backward from 2014 to 2021 using its full archive search API. The number of tweets before 2014 is insignificant. Hence, our analysis starts from 2014 onward. Different keyword combinations of the word bushfires and wildfires were used to query the tweets, such as *#bushfire, #bushfires, #wildfire,* and *#wildfires*, case-insensitive. For each tweet, the unique identification number (id), date, hashtags, and annotated location (if available) were collected in addition to the text for analyzing the trends over the years.

Approximately 20% of the tweets have been annotated with a location. This location information was extracted from the text content of the tweet and referred to as the target location. For example, in this sentence, 'CFS advises a bushfire in the Southern Flinders Ranges may threaten your safety,' the Southern Flinders Ranges is the target location of the *bushfire*. Some tweets mention more than one location. These cases were excluded from all location analyses.

The location information is unstructured. It can be a continent (e.g., Europe), country (e.g., Australia), state or city (e.g., California), or other administration levels. In this work, we converted these unstructured data into countries using the Google geo-decoding API[5] search engine. The country component of the result was assigned to that location. Hence, tweets with a location that does not have a specific country would be excluded from all location analyses. After inferring the country information, we retrieved around 470,000 tweets (17.5%) with valid country information, of which around 107,000 were from Australia.

Hashtag Analysis

In this analysis, the discussed topics were ranked to visualize the most popular topics and see how they evolved over the years. We employed hashtag counting as a tool for the topic ranking year-wise. Hashtags were taken as an indicator because they are usually relevant to the main topic. As a result, Table 1 presents the most trending hashtags. Political keywords such as *Australia, auspol, and *pol* dominate the table from 2017. Starting from the year 2016, the keyword *climate change* and its other form (e.g., *climate emergency*, *climate crisis*) ascended to the top of the table. The term *'climate'* was mentioned frequently in 2015; however, the pivotal time of turning the public's attention toward climate change was one year later. Besides, the keywords *heatwave and drought* have been frequently mentioned since 2017. These two observations show that people have connected wildfires to drought and climate change more and more since 2017.

The relationship between wildfires and climate change has been investigated since the 1990 s. Moritz et al. [35] showed that three main factors that cause wildfires are the weather, the availability of vegetation, and the ignition source. For many years, the third factor (e.g., lightning, accidental or deliberate fires) has been the main focus in classifying and preventing wildfires [36]. The role of climate change in the natural ignition source of wildfires was modeled quite early but lacked concrete evidence [37, 38]. On the other hand, the increasing temperature has become more visible to the community, and the hot weather (i.e., the first factor) naturally intensifies the risk and severity of wildfires. This temperature change can influence people to connect wildfires to climate change.

Why this relationship was widely accepted and presented on Twitter from 2016–2017 still needs to be determined. One hypothesis is that people witnessed the increasing damage of wildfires and the temperature concurrently; hence, they associated climate change with the cause of this acceleration. This hypothesis is slightly supported by the line graph of Fig. 1 (described in "Volume Analysis" section). The graph shows that the burned areas in 2016 were not worse than in previous years, but the number of damaged buildings surged, which might have alarmed Australians about the increasing risk of wildfires. Another hypothesis is that news media helped spread research and statistics related to climate change and wildfires, which convinced the community about this causal relation. With the increasing temperature, the community started to accept the connection between

[4] https://developer.twitter.com/en/products/twitter-api/academic-research.

[5] https://developers.google.com/maps/documentation/geocoding/start

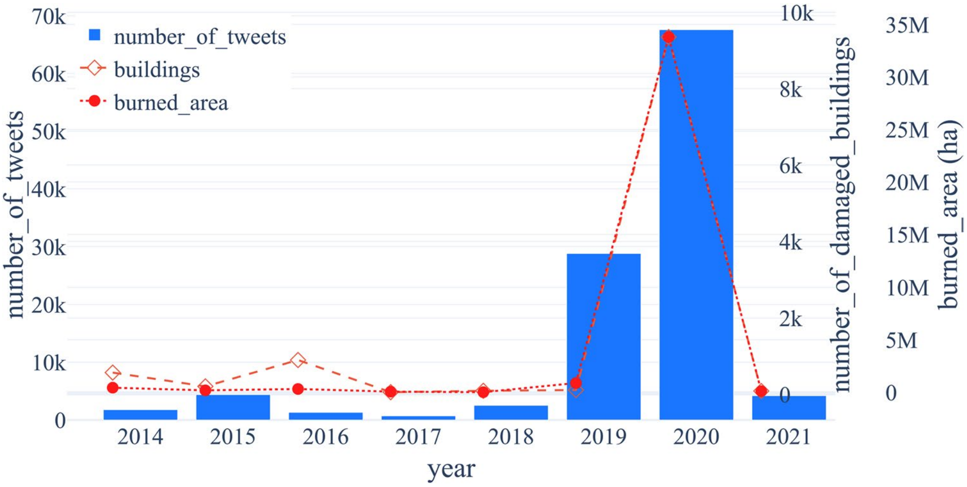


Fig. 1 The trend of the tweets from Australia. The red curves plot the number of damaged buildings (diamond dashed curve) and burned areas (circle dotted curve) due to wildfires. Both have some degree of correlation to the number of tweets

climate change and wildfires. Researchers Linnenluecke and Marrone [39] proved that more newspaper articles covering wildfires and climate change in 2017 than in the previous year. However, this hypothesis needs more evidence from the content of the tweets, which is limited by the current technique of hashtag counting.

Volume Analysis

Figure 1 overlays the number of tweets (bar graph) and the number of damaged buildings (line graph) due to wildfires in Australia. A peak from the season of 2019–2020 could be observed, leading us to further investigation.

The number of damaged buildings presented in the line graph of Fig. 1 was compiled by Wikipedia contributors based on the reported data from Australian local news [40–47]. Since the data were aggregated season-wise (e.g., the first data point is from the 2013–2014 season), each data point was placed between two consecutive years in this graph. The line and bar graphs are highly correlated, especially during the peak season of 2019–2020. The correlation shows that people use social media to express opinions about hazardous events such as wildfires when observing them.

This finding agrees well with similar observations reported in other types of events. For example, [15] showed that the temperature changes were reflected by the number of tweets about the weather on social media. On the other hand, in this work, the number of damaged buildings, a critical figure representing the cost of the events, was plotted to prove that the social media reactions could also echo the economic loss of wildfires. Next, the following sections will present our polarity prediction and emotion quantification models to gain more insights into people's opinions about wildfires.

Sentiment Analysis

Observing people's sentiments while conversing and tweeting about wildfires is an essential mark representing how they feel about this hazard. Hence, we aimed to detect the sentiments of the tweets. We focused on the two main polarities for this purpose: positive and negative. Recently, [12] have shown that the polarity detection task in tweets about climate-related domains is more accurate with the supervised learning approach using a domain-specific dataset than a general predictor such as the Stanford Corenlp API [48]. Therefore, we trained a polarity detection model for the wildfire domain in this work.

ClimateTweets [12] is an appropriate dataset for this task. It contains climate change-related tweets, which are often related to wildfires. Table 3 displays typical samples from ClimateTweets. Although ClimateTweets has three sentiment levels (i.e., positive, neutral, and negative), only the positive and negative samples were used in this work. This selection aims to increase the contrast between the labels, enhancing the model's prediction accuracy.

Moreover, this polarity model is the intermediate step in producing the golden samples to train the emotion model ("Emotion Analysis" section). One critical goal of selecting the golden samples was to filter out the excessive neutral samples (i.e., samples with nearly zero scores in all emotions and polarities) from the population produced by the NRC lexicon. The population has more than 80% neutral samples, as shown in Fig. 25. Hence, not choosing neutral samples while training the polarity model could help reduce their portion in the golden set. Figure 26 shows that the emotion model's training set is less skewed than the original population. Though, there are still plenty of presented neutral samples. Hence, not selecting neutral samples while training the polarity model does not harm the golden set in terms of lacking neutral sentences.

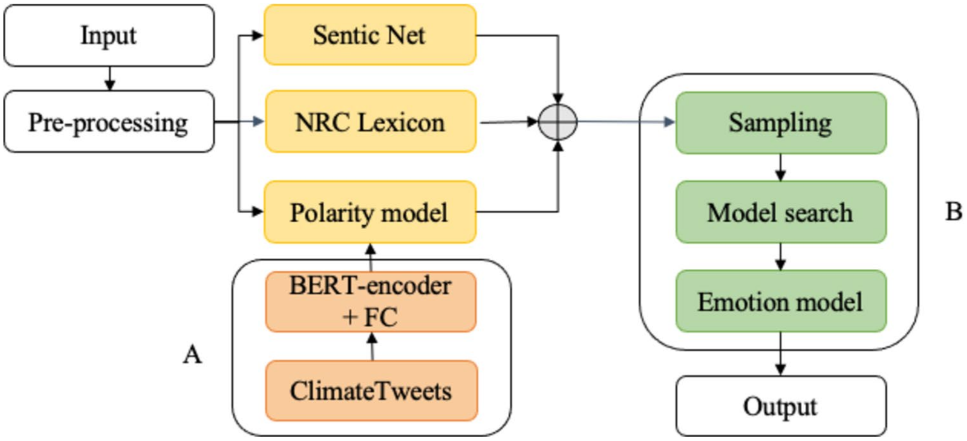


Fig. 2 Emotion model training flow. Block A: train the polarity model using ClimateTweets dataset. Block B: train the emotion model using golden tweets filtered by SenticNet polarity API, NRC lexicon, and the polarity model

Block A of Fig. 2 illustrates the training procedure of the polarity prediction model. The scraped tweets and the samples from the ClimateTweets dataset were preprocessed before passing into Block A. The preprocessing step removed all the special characters: URLs and the "RT" keyword (i.e., Re-Tweet). In Block A, the model contains a BERT-Bidirectional Encoder Representation [27] and a fully connected (FC) layer with the softmax activation function. The dropout rate was kept at 0.4, with a 2e-5 learning rate during training. The model achieved an overall validation accuracy of approximately 91%. Later, the trained polarity model was used for the tweets' polarity labeling, and the confidence scores were recorded. Here, the softmax probabilities were taken as the representations for the confidence scores. This block serves as the polarity detection model.

Emotion Analysis

In addition to the polarity detection, analyzing the emotions plays a pivotal role and provides us with a bigger picture of studying the impacts of this hazard. At first, we followed a lexicon-based method using the NRC lexicon [24]. The NRC lexicon comprises ten labels: eight emotions (i.e., joy, anticipation, trust, fear, anger, sadness, surprise, disgust) and two polarities (i.e., positive and negative). We employed the NRXLex,[6] based on the NRC lexicon, to produce ten values for the above ten labels. These values are continuous numbers between 0 and 1, quantifying the intensity of the emotion or sentiment in the tweet.

Nevertheless, using only the NRC lexicon likely leads to false predictions if the text contains contradictory words, as presented in Table 2. For example, "great" and "threat" in the first sample or "beautiful" and "sad" in the second sample might confuse the NRCLex to determine the correct emotion and sentiment for the text. On the other hand, BERT-FC ("Sentiment Analysis" section) and SenticNet [23] are supervised learning models that deduce the sentiment by using all the words of the text. Hence, the two models predict these sentences correctly, justifying the strength of the supervised model approach. Therefore, we trained a supervised learning model for the emotion prediction task instead of solely using the NRC lexicon.

Methodology

Figure 2 describes the flow of the training process. The raw data were passed through the pre-processing block, then directed into three models: the SenticNet sentiment polarity model, our BERT-FC polarity model (block A, described in "Sentiment Analysis" section, and the NRC lexicon model. The intersection of the outcomes from these three models formed the pool of trusted samples, or golden samples, for emotion labels. The rules of the intersection are:

- BERT-FC and SenticNet agreement: valid samples are samples with the same polarity predicted by BERT-FC and SenticNet (i.e., both are positive or both are negative) or samples with more than 0.8 BERT-FC confidence score. The confidence score is the probability of a label predicted by the model.
- BERT-FC and NRCLex agreement: valid samples are samples with higher than zero polarity scores on the corresponding BERT-FC label (i.e., samples that BERT-FC predicts as positive (or negative) and the NRCLex's positive score (or negative score) is higher than zero.

The samples passed through the three models would provide more credible labels (i.e., emotion and polarity scores)

[6] https://github.com/metalcorebear/NRCLex.

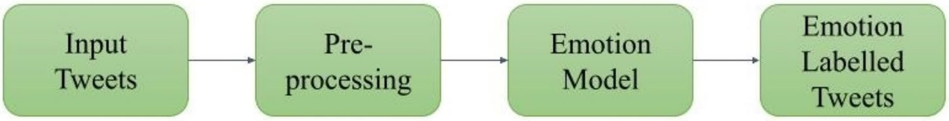


Fig. 3 Emotion prediction flow

than the original population from the NRC lexicon. In order to balance the training dataset, a downsampling step was carried out. The final set has around 10,000 samples, each having eight emotions and two polarities. All emotion and polarity scores range from 0 to 1.

After downsampling, the samples were used to search for the best emotion quantification model. Autokeras [49], a Neural Architecture Search engine, was employed to find the best neural network for this training data. A 10-head regression model was built. The best model was found by performing several operations, such as adding new layers, expanding the dimensions, and adding new connections. The number of epochs was set to 100. The search space trials were set to 20, which enabled Autokeras to explore the 20 most-suited models and choose the best-fitting one for our dataset. Table 4 displays the model architecture that was chosen and employed to run on the training data. In detail, the model comprises a text vectorization layer followed by an embedding layer. The dropout layer is to reduce over-fitting. The two consecutive convolutional layers separated by max pooling are inserted with dimensions 62 x 32 and 29 x 32, respectively. The final output is flattened and is passed through a dense layer. The activation function used is ReLU. This trained model was used to label the emotions for all the tweets. Figure 3 describes the production flow.

Evaluation Metrics

The given approach was evaluated using two main metric sets:

Table 4 Emotion quantification model architecture

Input (1)
Text Vectorization (64)
Embedding (64, 128)
Dropout (64, 128)
Separable Conv 1D (62, 32)
Max Pooling 1D (31, 32)
Separable Conv 1D (29, 32)
Max Pooling 1D (14, 32)
Flatten (448)
Dense (32)
ReLU (32)
Dense (1) x 10

- Polarity model (BERT-FC): accuracy and categorical cross-entropy loss. Let y be the binary indicator and p represent the predicted probability. Then, the loss is given by:

$$Loss = -(y \log(p) + (1 - y) \log(1 - p)) \tag{1}$$

- Emotion model: accuracy and the Mean Squared Error (MSE). Let Y_{tr} and Y_{pr} be the true and predicted emotion, then $MSE(Y_{tr}, Y_{pr})$ is given by:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (Y_{tr}(i) - Y_{pr}(i))^2 \tag{2}$$

where N refers to the number of samples in the test set.

Results

This section elaborates on the performance of the trained emotion quantification model. The predicted values follow the golden scores closely with the average MSE of 0.0055 and Pearson correlation coefficient of 0.89. The MSE(s) and Pearson correlation coefficients across all ten output features are recorded in Table 5.

To further evaluate the effectiveness of the new emotion model, we built a manual dataset containing 200 samples, each labeled by three independent annotators. The annotators were asked the question: "How much *emotion/sentiment* do you see from this sentence (from any mentioned subject)?". Then, the annotators would score one of the four levels: *not at all, slightly, moderately, and very much*. The scores were averaged and linearly normalized to the range of [0, 1] before being compared with the NRC lexicon method and the new emotion model. Throughout this analysis, we will use the Mean Absolute Error (MAE) to assess the differences among annotators and methods (Eq. 3).

Table 5 MSE test scores and Pearson correlation coefficients of all emotions and sentiments

Features	Positive	Anticipation	Surprise	Trust	Joy
MSE	0.072	0.048	0.046	0.056	0.036
Corr-coeff	0.87	0.88	0.92	0.88	0.89
Features	Negative	Sadness	Anger	Fear	Disgust
MSE	0.065	0.051	0.045	0.055	0.034
Corr-coeff	0.86	0.91	0.90	0.89	0.86

Table 6 MAE errors between human labels versus the two models. *EM. denotes Emotion model*

Features	Positive	Anticipation	Surprise	Trust	Joy
NRC	0.215	0.178	0.0714	0.120	0.0950
EM	**0.171**	0.171	0.0653	0.101	0.0939
Features	Negative	Sadness	Anger	Fear	Disgust
NRC	0.241	0.115	0.0861	0.197	0.0541
EM	**0.221**	0.126	0.0855	0.184	0.0619

$$MAE = \frac{1}{N}\sum_{i=1}^{N} abs(Y_{tr}(i) - Y_{pr}(i)) \tag{3}$$

where N refers to the number of samples in the manual labeled set.

Table 6 summarizes the comparison results using MAE errors between the manual scores of the two models: the NRC lexicon and the new emotion model. Data show that the new emotion model has better MAE errors in most features, except for *sadness and disgust*. For example, the new model has improved by 20% and 8% in the positive and negative polarities, respectively.

Qualitative Analysis

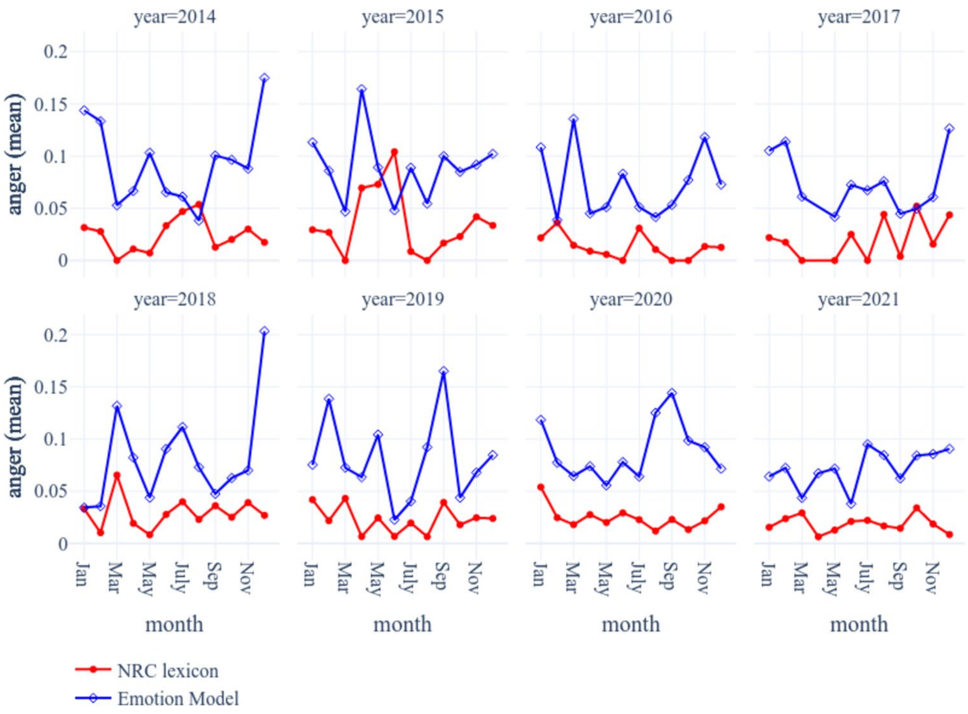
A. Mean Scores The model was applied to the entire population to evaluate the ten features of each tweet. Figure 4 compares the outcomes of the proposed model to the original NRC lexicon regarding anger emotion. The figure plots the average scores from emotion or polarity over twelve months of the year. The average score was chosen to eliminate the bias toward the peak number of tweets in the season 2019–2020. The figure unveils one significant shortcoming of the NRC lexicon: the limited size of the lexicon could only cover some cases, leading to nearly zero scores across many months. On the other hand, the trained emotion quantification model overcame this challenge and returned more non-zeros scores. The trend provided by this model shows more realistic results than the NRC lexicon since it seems unrealistic that the community's emotion would suddenly reduce to zero.

The proposed model can also produce the same trend as the NRC lexicon in critical years, such as the 2019–2020 season in Fig. 4. These years received the highest samples; hence the NRC lexicon produced fewer zero scores and more credible trends than other years. The trends from the emotion model agree well with the NRC lexicon, especially the peaks in September and December of 2020 are more apparent.

Similar characteristics can be observed in other negative emotions and polarity (see the Appendix section for more plots (Figs. 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 23, and 24)). One exception is the fear emotion plotted in Fig. 5. The NRC lexicon does not provide many zero points in this figure as in the sadness graph. Hence, both models produce closer results than the previous one.

On the other hand, the difference between the two models for positive emotions and polarity is smaller than on the negative side. Figure 6 shows the positive

Fig. 4 Anger mean scores predicted using the NRC lexicon and the emotion model



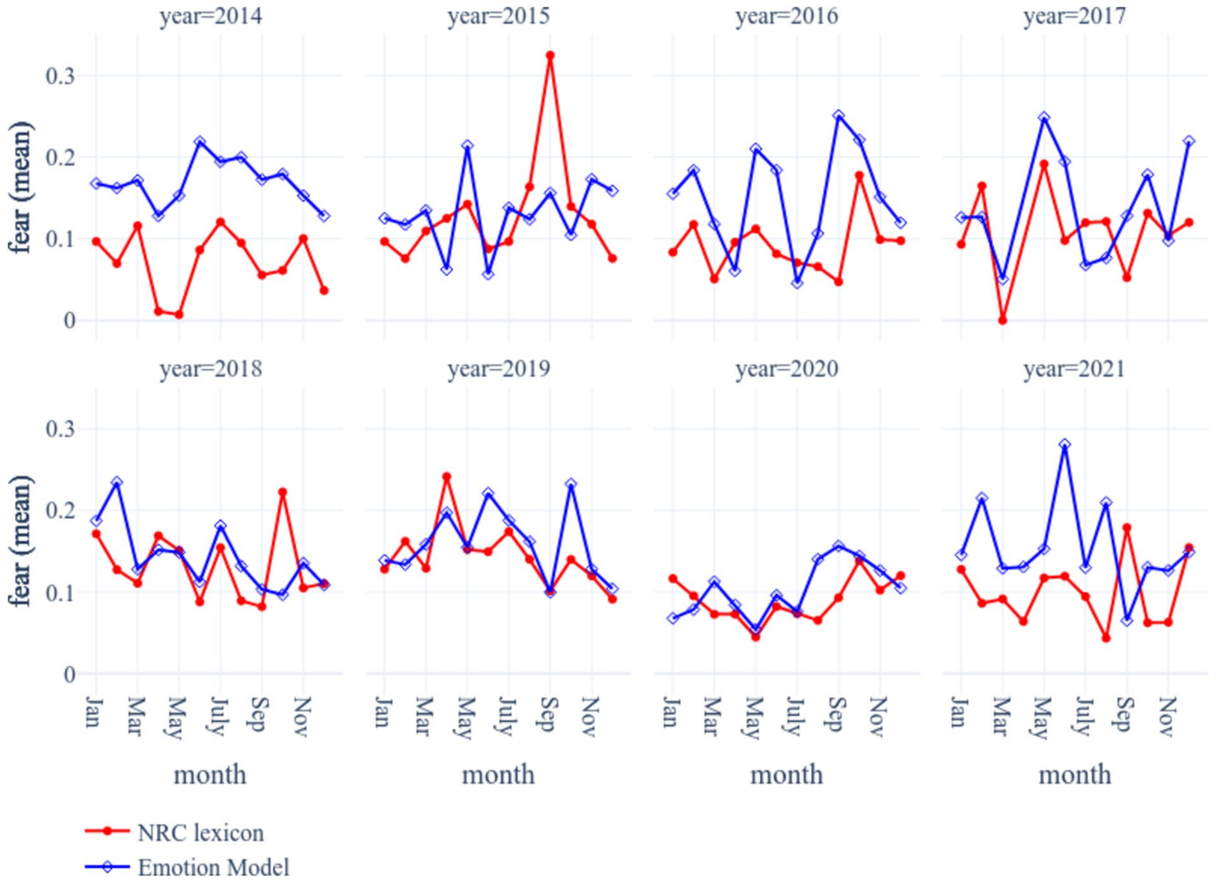
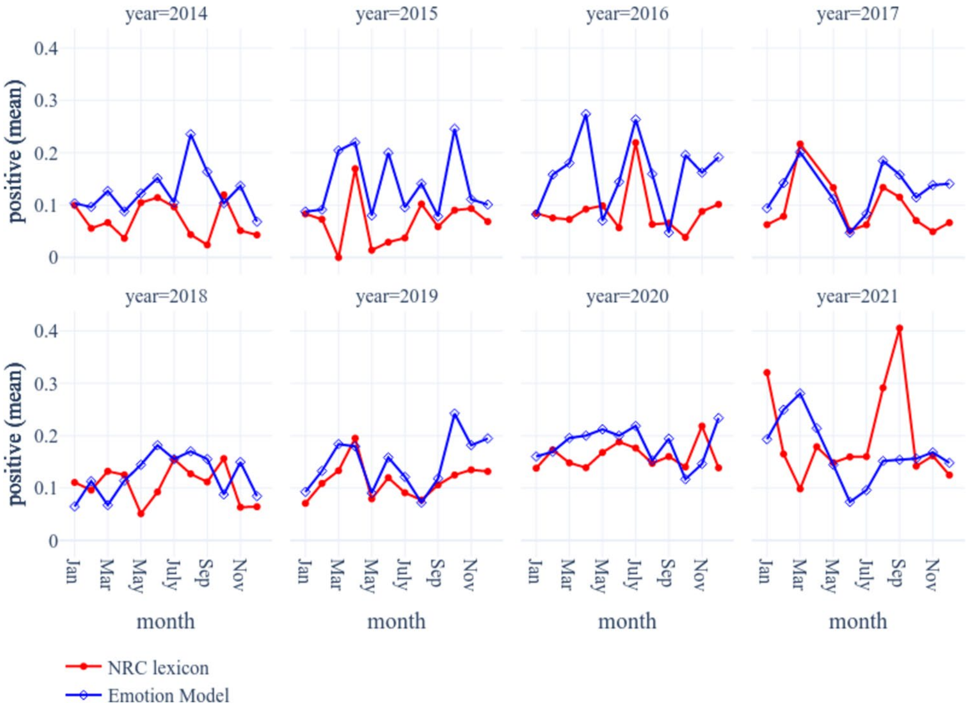


Fig. 5 Fear mean scores predicted using the NRC lexicon and the Emotion model

sentiment trend over the years produced by the models. Compared to Fig. 4, there is no significant gap between the red and blue curves. Also, a few zero points are observed in Fig. 6. Hence, this could be an indicator to identify cases where the NRC lexicon could be a reliable estimator.

Fig. 6 Positive scores predicted using the NRC lexicon and the Emotion model



B. Volume Analysis The mean score analysis of the above section shows some advantages of the proposed emotion model over the NRC lexicon. However, the mean score cannot determine the relationship between tweets and actual wildfire events, as in Fig. 1. Hence, in this section, we present another set of graphs overlaying the number of tweets and the actual wildfire events to demonstrate the usefulness of the emotion model.

Figure 7 illustrates actual wildfire events and the predicted tweets with the anger emotion. The figure has three graphs: 1. the dashed red line graph shows the number of tweets classified as being anger by the NRC lexicon, 2. the solid blue line graph shows the same entity by the emotion model, and 3. the blue bar graph shows the number of retrieved wildfire events. A wildfire event could last from several days to several months. Since not all events have recorded duration, the bar graph only counts for the start date. It would be more meaningful if the bar graph represent the events' damaging scales. Unfortunately, many events do not have enough details for us to summarize. Nevertheless, the tweet volume with anger presented in Fig. 7 reflects the number of actual events.

Earlier in this paper, we mentioned that the eight emotions and two polarities predicted by either the emotion model or the NRC lexicon have continuous values in the range of [0,1]. Therefore, to count the number of tweets with anger, we set a threshold of 0.5 to classify whether the tweet has anger emotion. The same threshold was used for all features and models. The threshold significantly affects the outcomes, influencing the comparison between the two models. Figures 8 and 9 demonstrate the effect of the threshold.

The tweets predicted by the two models have some prominent differences. First, Fig. 7 shows that the emotion model can extract more tweets with anger emotion than the NRC lexicon, such as in 2014, 2015, 2017, 2018, 2019, and 2021. The trend of the emotion model

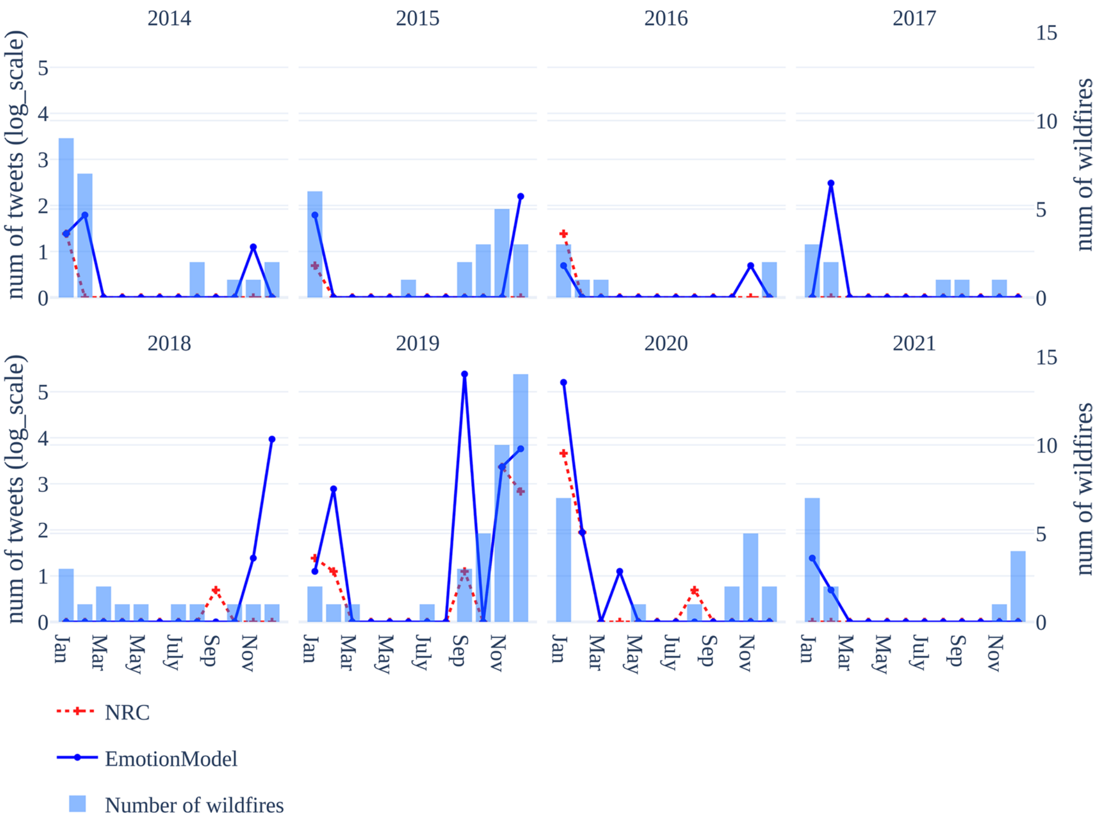


Fig. 7 Anger tweets predicted by the NRC lexicon and the Emotion model. The left y-axis is the shared axis for the two line graphs. Due to the significant difference between the number of tweets in the 2019–2020 season versus the rest, the logarithm with base-10 was used for the left y-axis to enhance the readability of the graphs. The right y-axis, in linear scale, is for the bar graph. The figure has eight sub-plots corresponding to eight years of this study

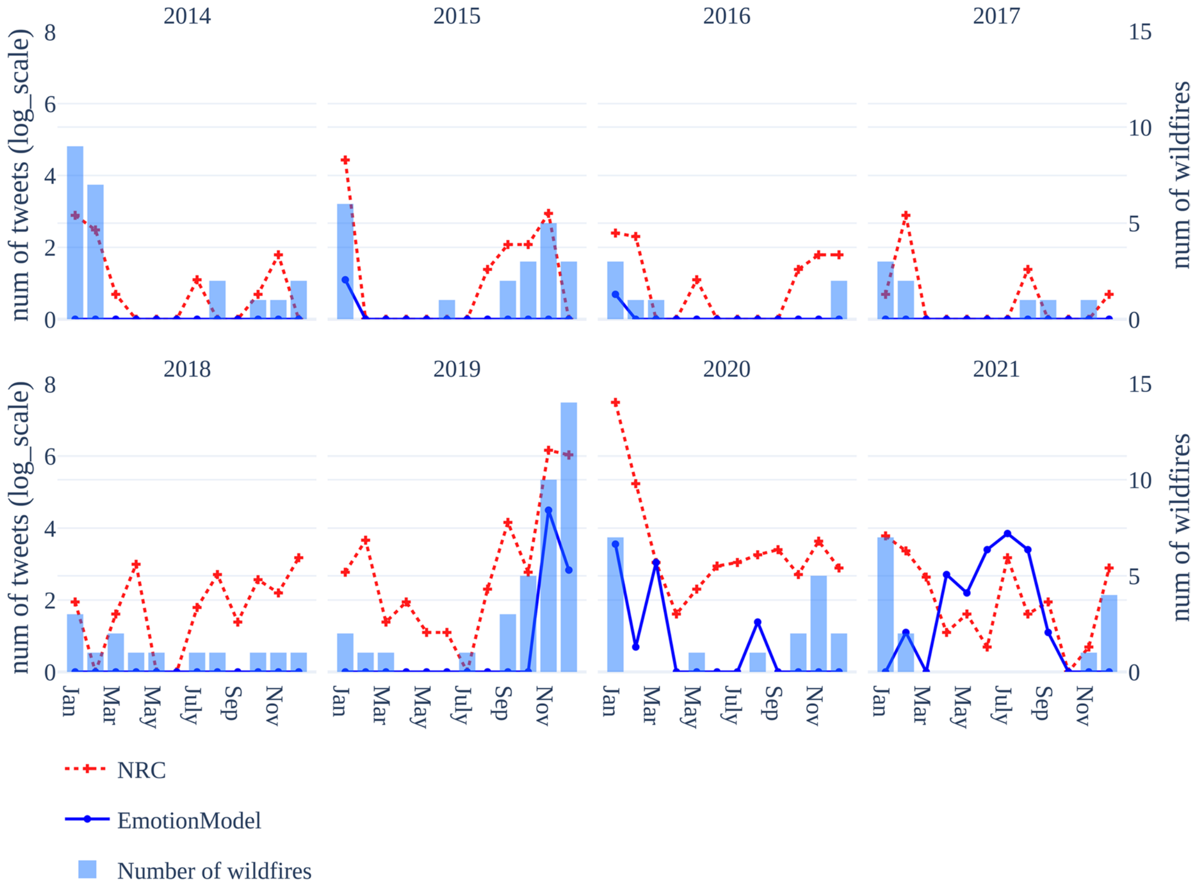
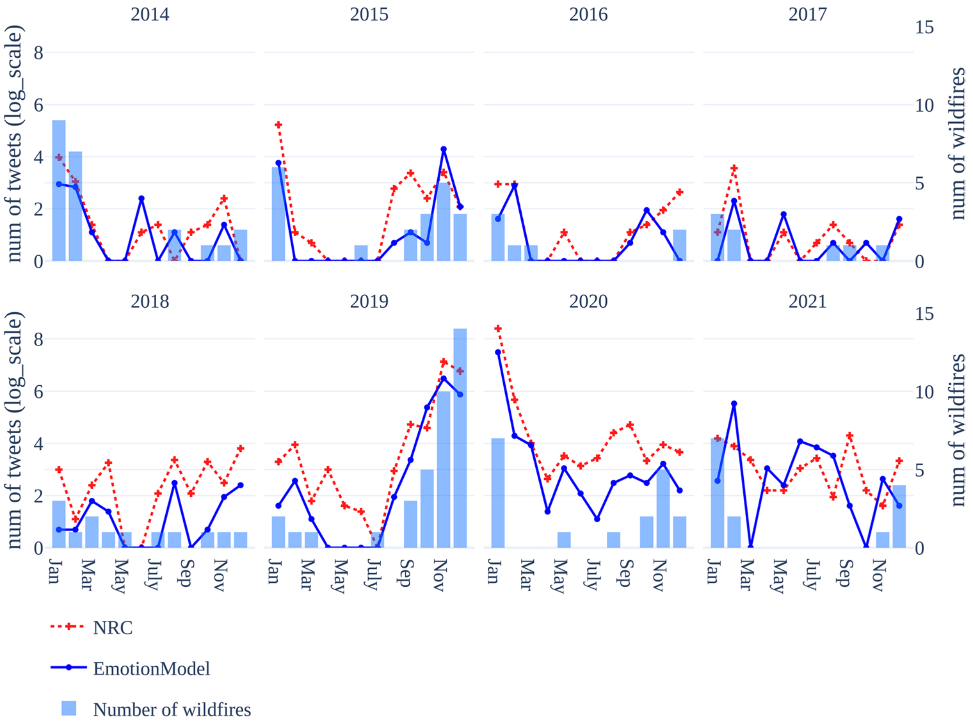


Fig. 8 Fear tweets predicted by the NRC lexicon and the Emotion model

Fig. 9 Adjusted fear tweets (threshold = 0.4) predicted by the NRC lexicon and the Emotion model



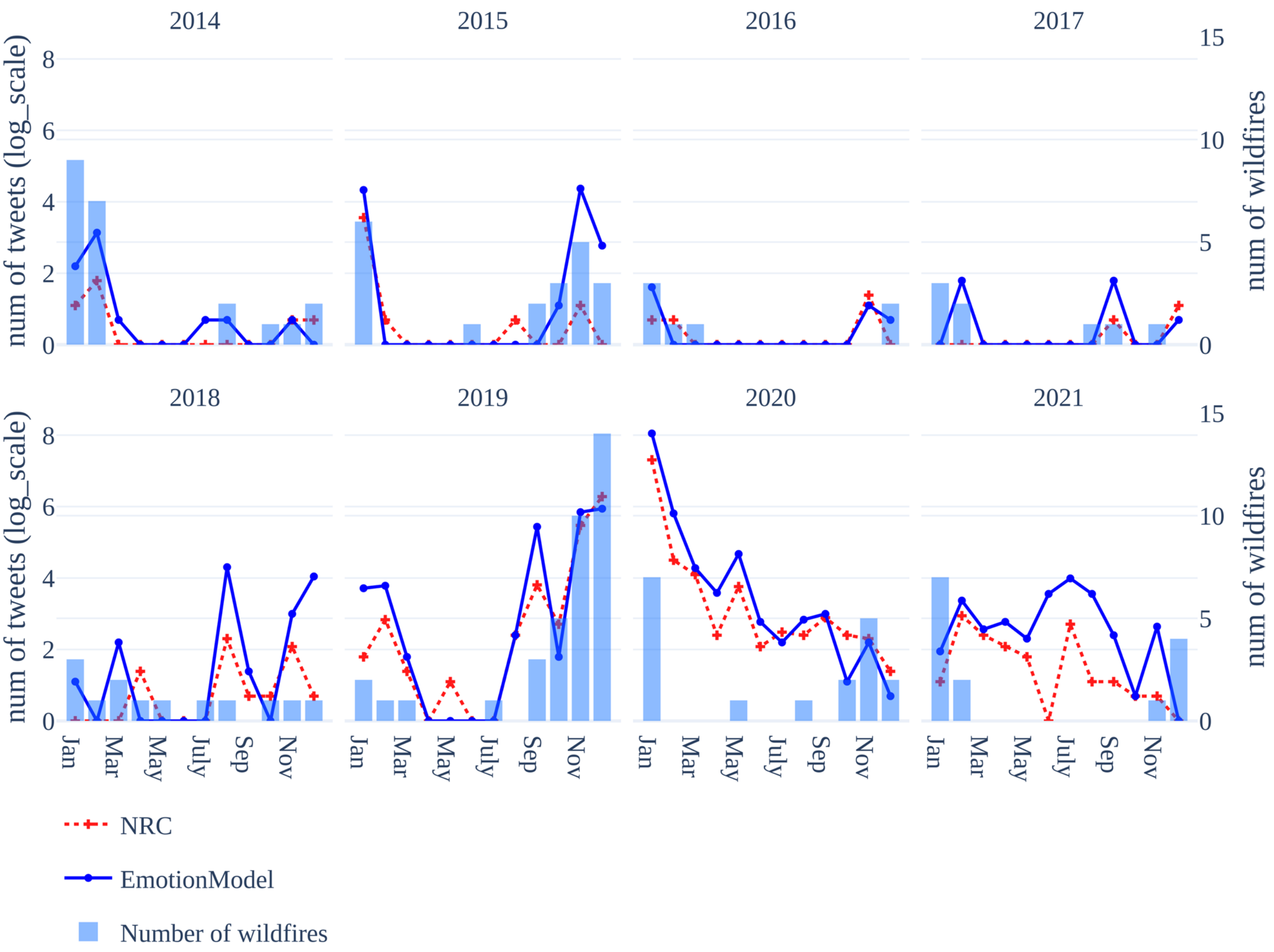


Fig. 10 Negative tweets predicted by the NRC lexicon and the Emotion model

also closely matches the time of the wildfires. A similar observation happens in the sadness emotion (Fig. 19). However, Fig. 8 shows that the NRC lexicon could extract more tweets with fear. Adjusting the threshold value, we found that lowering the threshold to 0.4 can help the emotion model matches the NRC lexicon in the fear emotion, as in Fig. 9. When placing Figs. 7 and 8 side-by-side, the total tweets seem to match each other. One hypothesis is that the NRC lexicon leans toward fear more than anger, while the emotion model does the opposite.

Nevertheless, Fig. 10 shows that the total number of negative tweets from the two models is close and follows the trend of the wildfire events. The tweets from the emotion model are higher than those from the NRC. Since the y-axis is in the logarithm scale, this gap is significant.

Second, for positive emotions such as joy, the NRC lexicon produces more tweets than the emotion model (i.e., Fig. 11). Statistically, tweets with joy are part of the population. When the population increases, the number of these tweets also rises. However, intuitively, there should be a few tweets with joy during fire season. Hence, the emotion model quantifies the joy emotion better in this case. Nonetheless, a deeper investigation must be conducted to make a trusted conclusion.

Finally, we found that people still had negative feelings about fires after the 'Black Summer' of the 2019–2020 season. The number of tweets with fear remains high after March 2020 (Figs. 8 and 9). The negative polarity plot also reflects the trend (Fig. 10). The trend shows that the 'Black Summer' consequences are not just the burned areas or damaged properties but also the mentality.

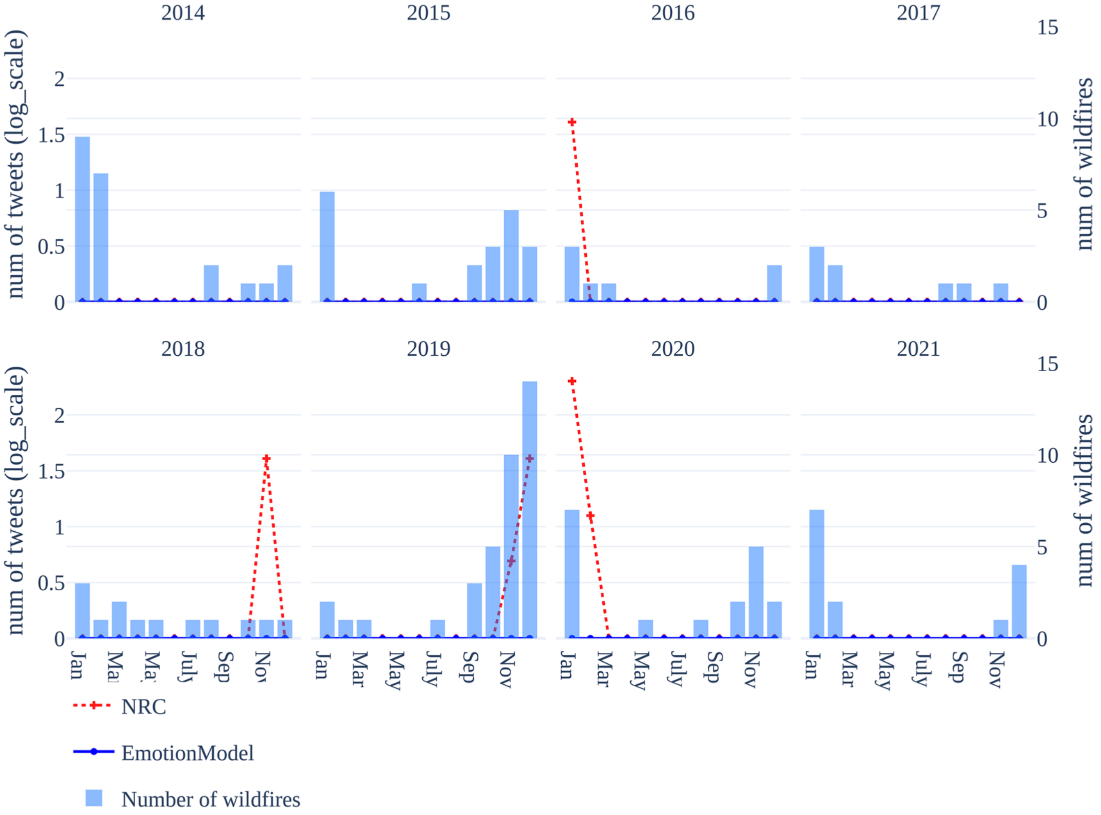


Fig. 11 Joy tweets predicted by the NRC lexicon and the Emotion model

Conclusion

Wildfires are costly and impactful hazardous events; however, research in this domain is limited to traditional survey methods and social media data analysis. Taking advantage of the available archived data from Twitter, we investigated people's opinions about wildfires from 2014 to 2021. For the sentiment and emotion quantification tasks, we developed a hybrid approach to extract golden samples from the NRC lexicon's estimation to train the model. The method is vital to building an emotion-labeled dataset in the wildfire domain.

The proposed emotion model generally estimates emotional and polarity scores better than the NRC lexicon. Data show that it has better MAE errors than the benchmark in most features, except for *sadness and disgust*. When applying the model to visualize public perception from 2014 to 2021, the graphs produced by this model showed more realistic results than the benchmark ones. These graphs also helped us understand that the consequences of the 'Black Summer' are not just related to burned areas, and damaged properties, but it also affected people's emotions and mindset towards climate change and wildfires.

During the investigation, we discovered two significant events. First, people have seen wildfires as one of the impacts of climate change in the years 2016–2017. Research has found some relationships between wildfires and climate change; however, the source for the acceptance trend that people presented on Twitter was unclear. Two hypotheses were discussed regarding this uptrend: 1) people witnessed the increasing damage of wildfires and the temperature concurrently, and 2) the news media helped spread research and statistics related to climate change and wildfires. Nevertheless, more research must be conducted to make a concrete conclusion.

Second, the trends of tweets, a social media product, can reflect the damage of wildfires in Australia, including the burned areas and the number of damaged buildings. The finding agrees with similar conclusions in other domains that data from social media are helpful to echo real-life changes. In future work, we will leverage explainable AI [50] to investigate public opinion evolution by time series in the climate change domain.

Appendix

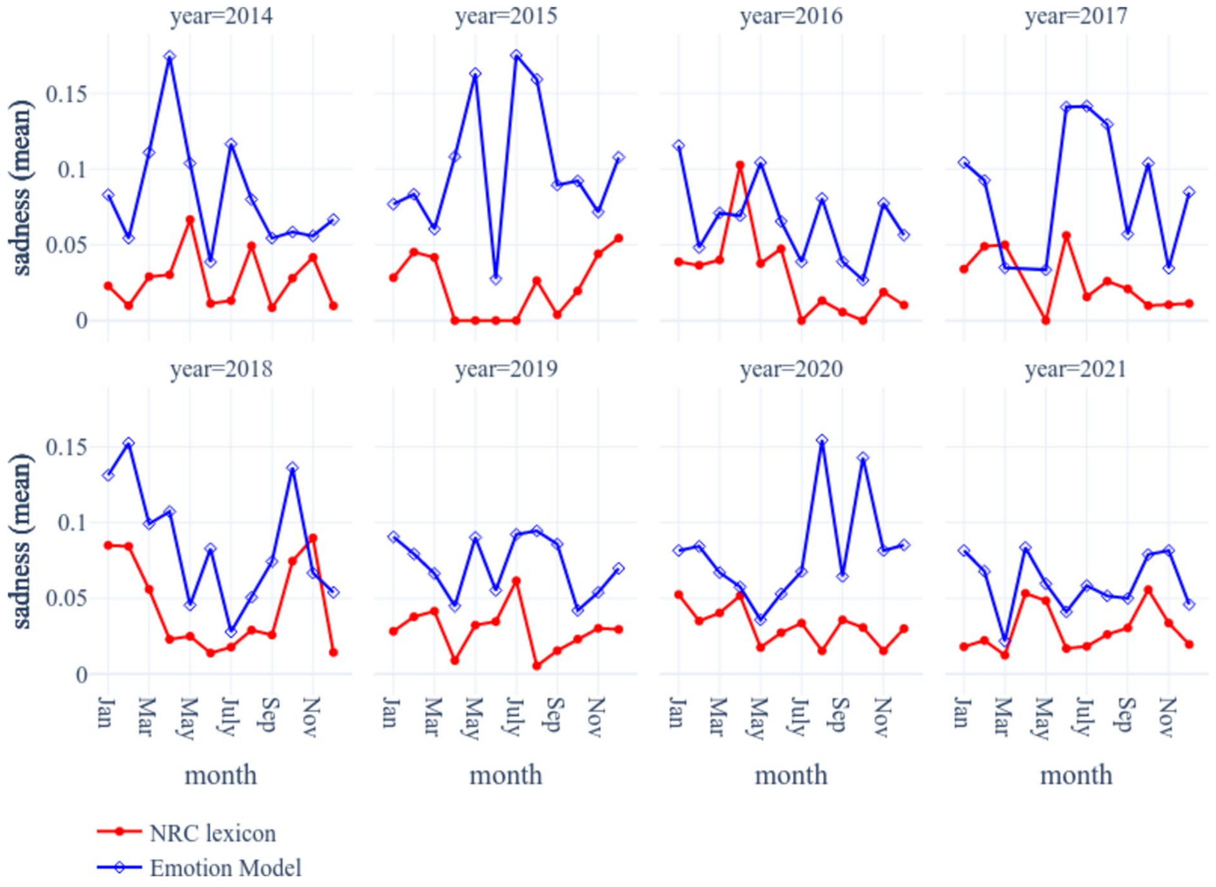
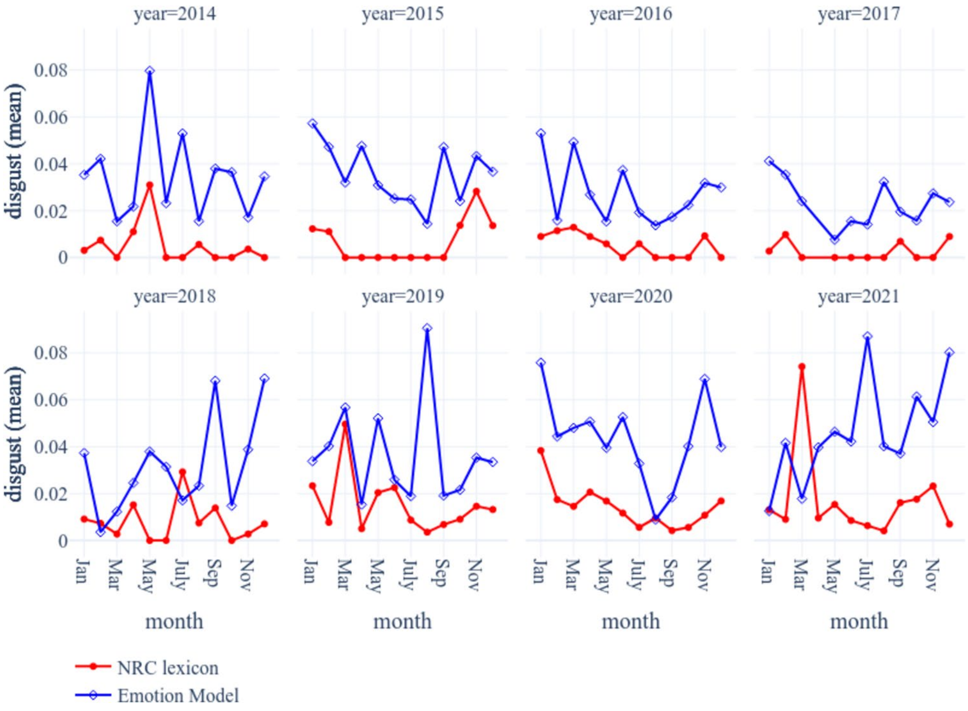


Fig. 12 Sadness scores

Fig. 13 Disgust scores



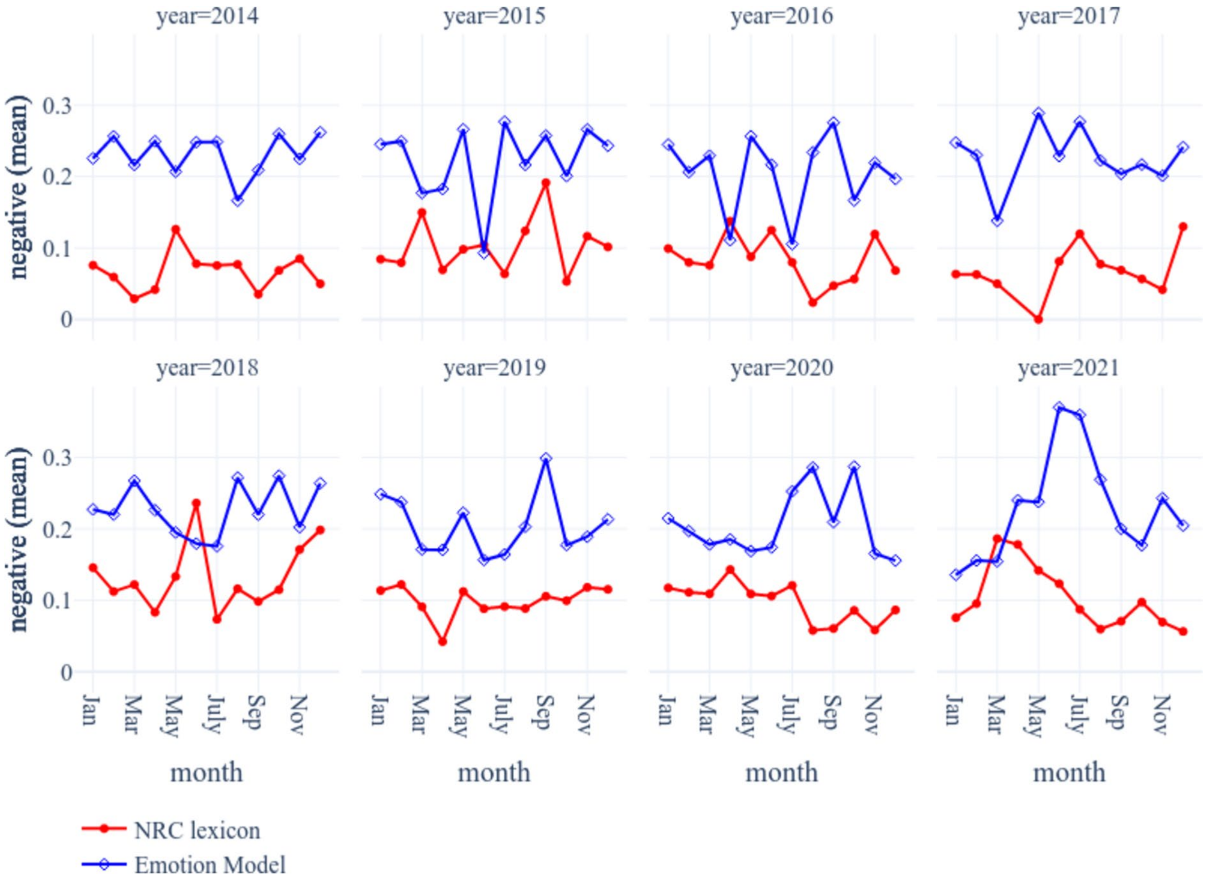


Fig. 14 Negative scores

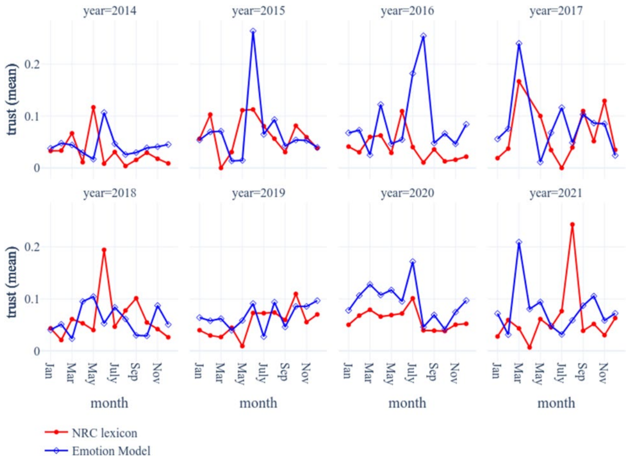


Fig. 15 Trust scores

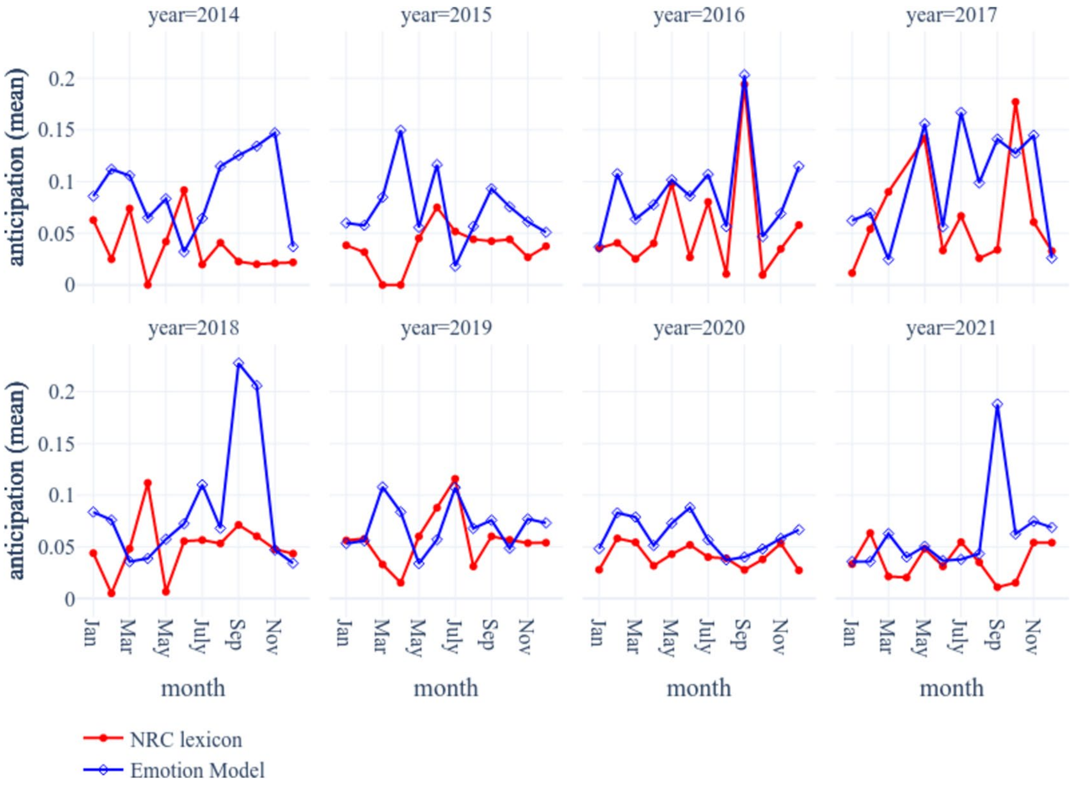
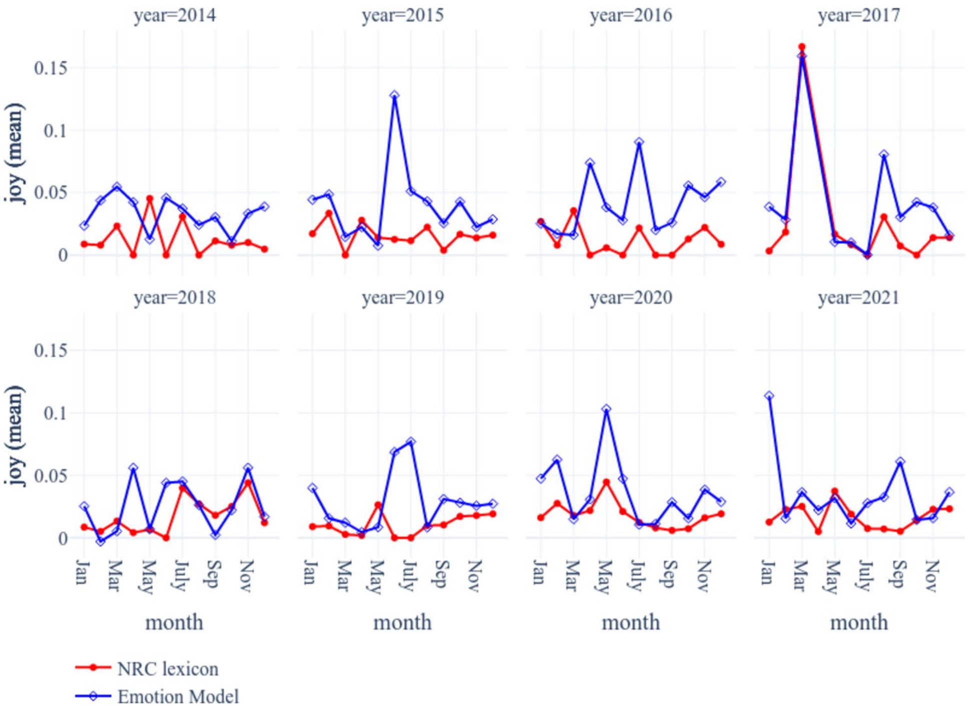


Fig. 16 Anticipation scores

Fig. 17 Joy scores



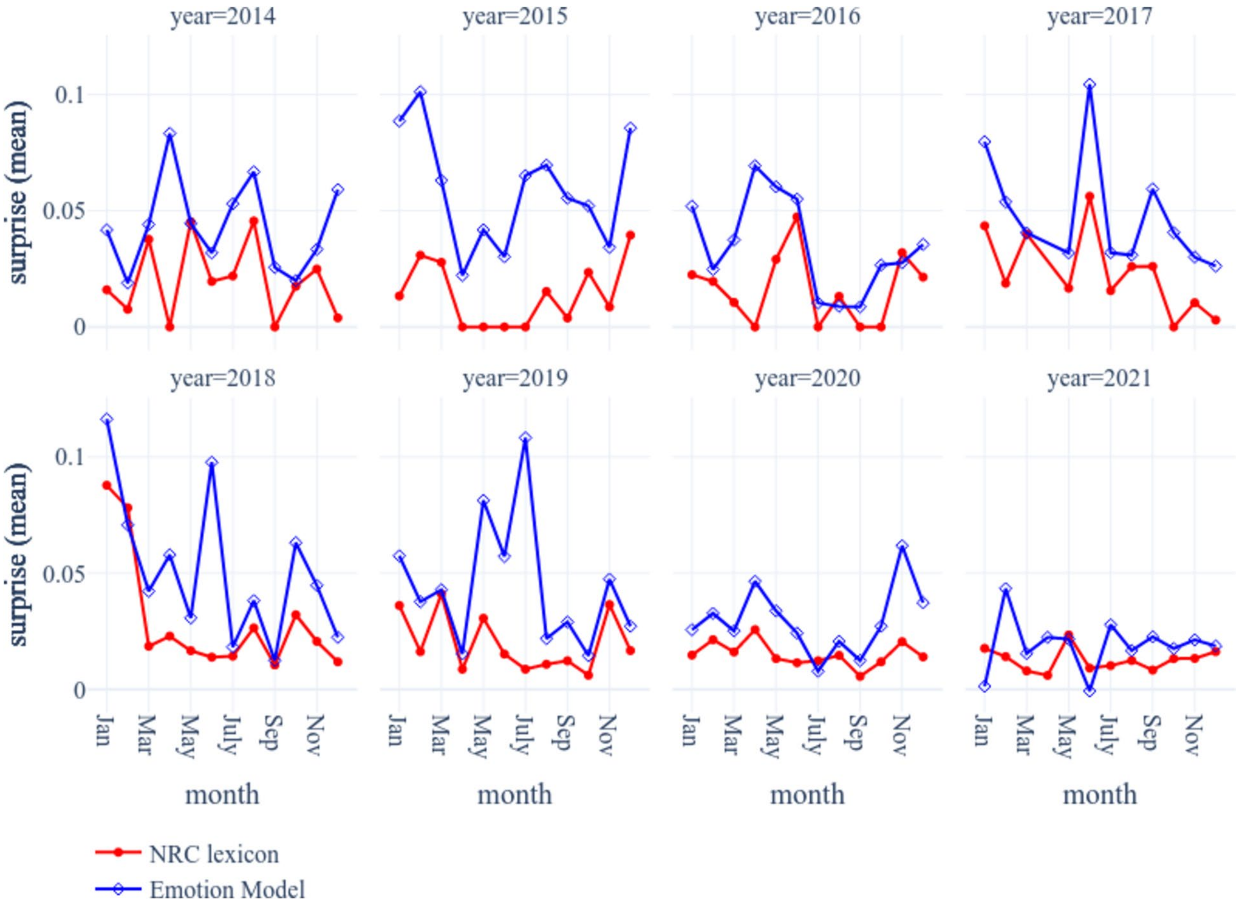


Fig. 18 Surprise scores

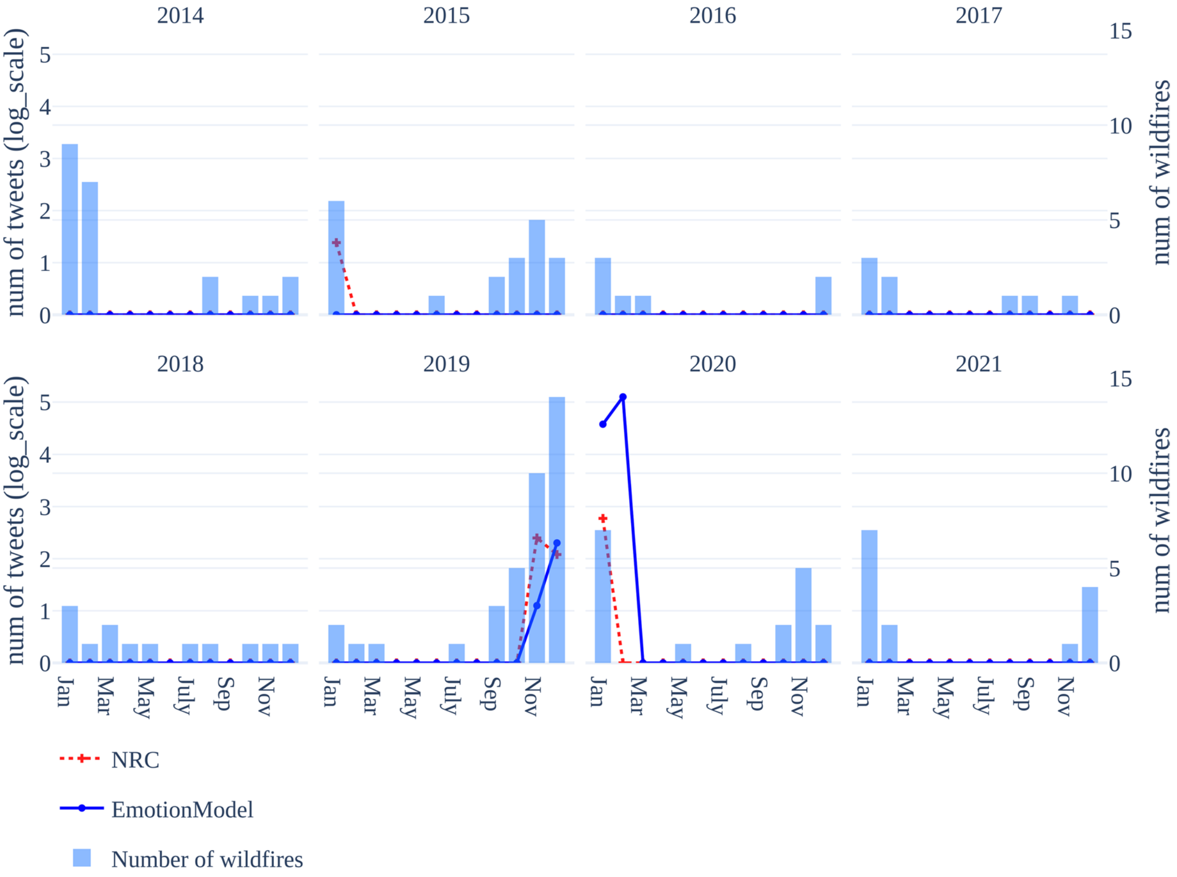


Fig. 19 Sadness tweets predicted by the NRC lexicon and the Emotion model

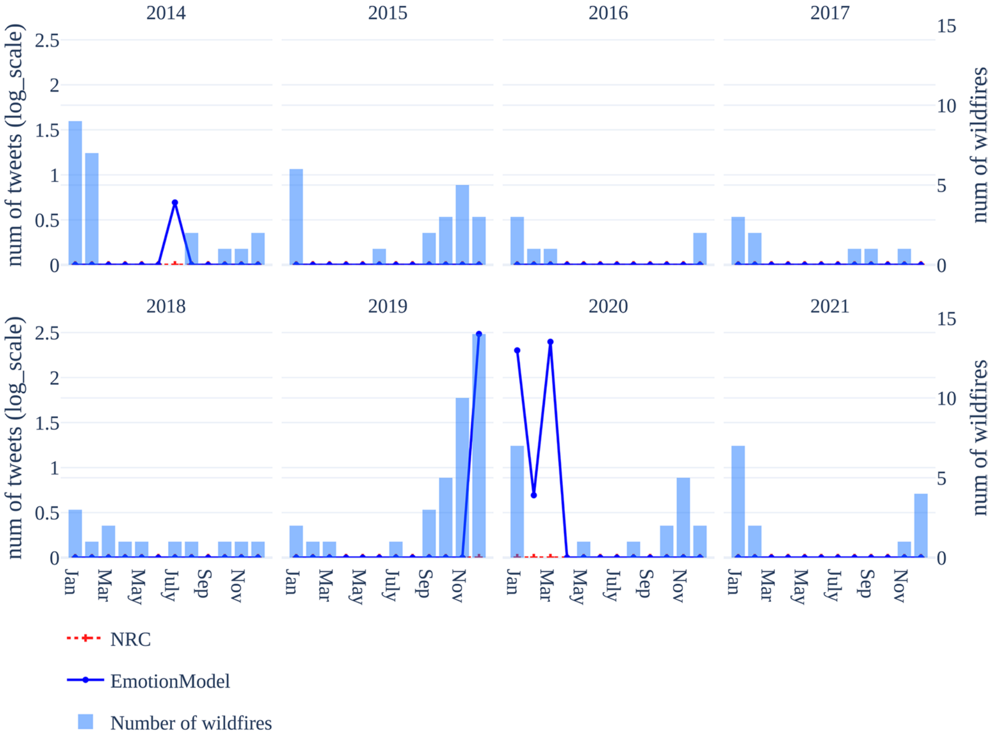


Fig. 20 Disgust tweets predicted by the NRC lexicon and the Emotion model

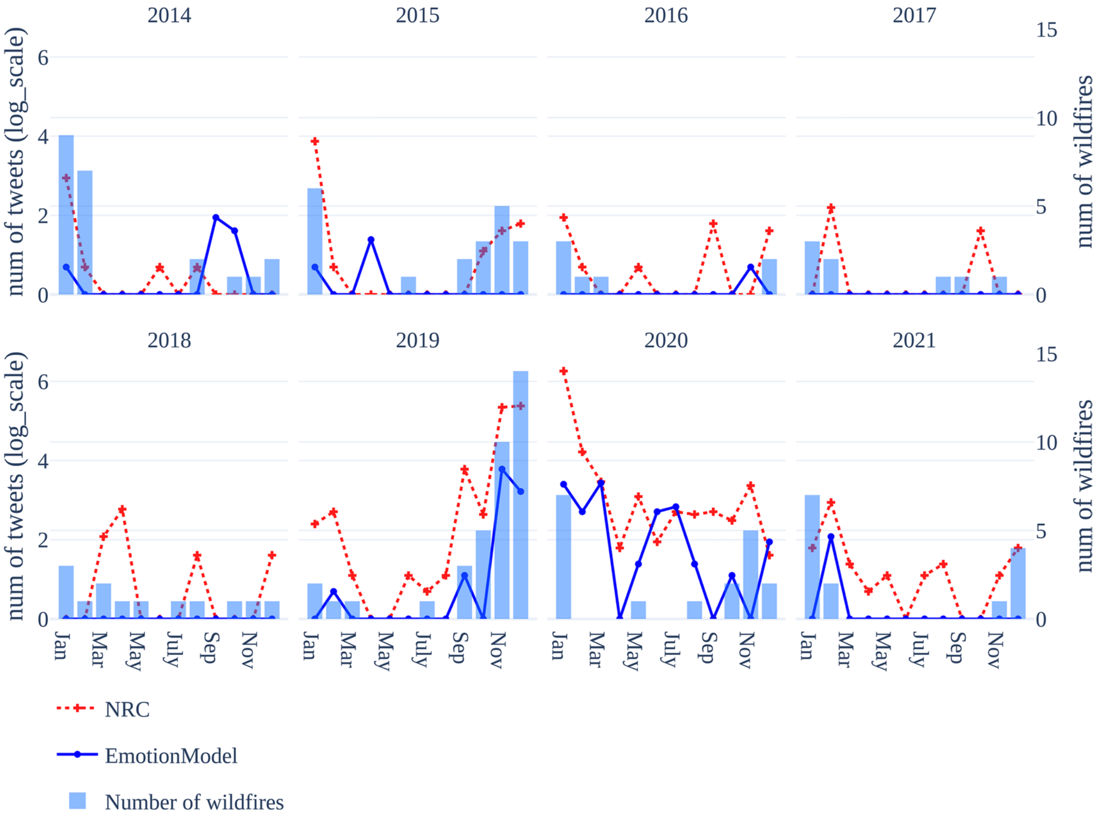


Fig. 21 Anticipation tweets predicted by the NRC lexicon and the Emotion model

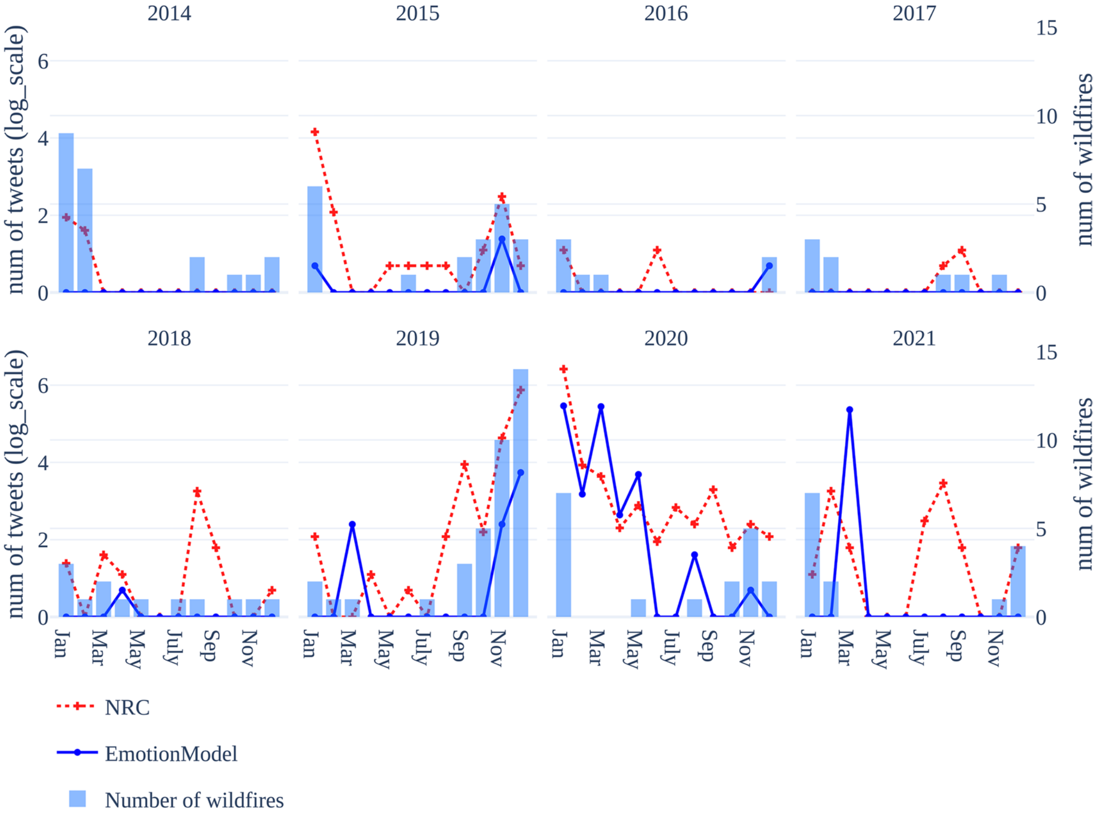


Fig. 22 Trust tweets predicted by the NRC lexicon and the Emotion model

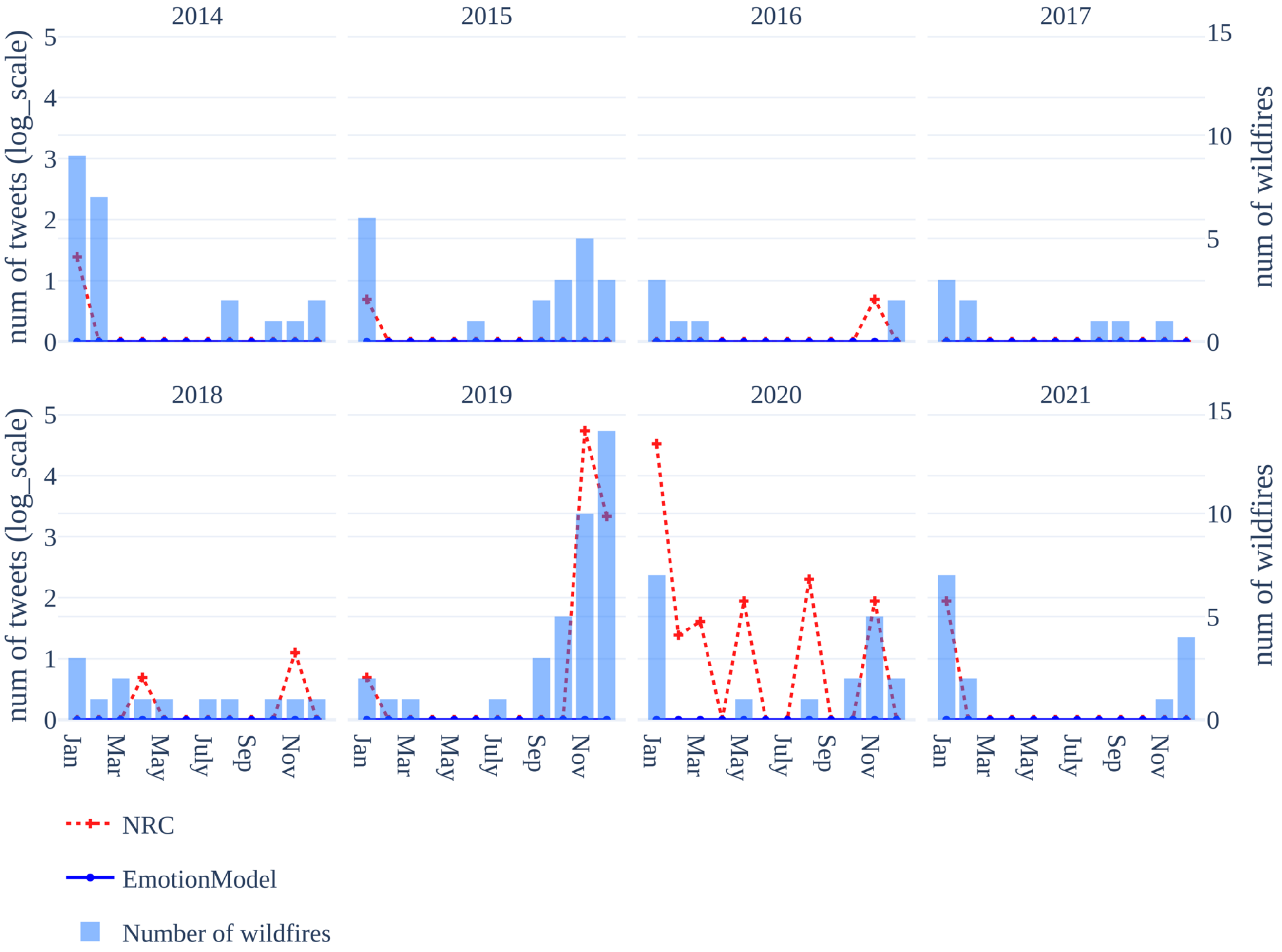


Fig. 23 Surprise tweets predicted by the NRC lexicon and the Emotion model

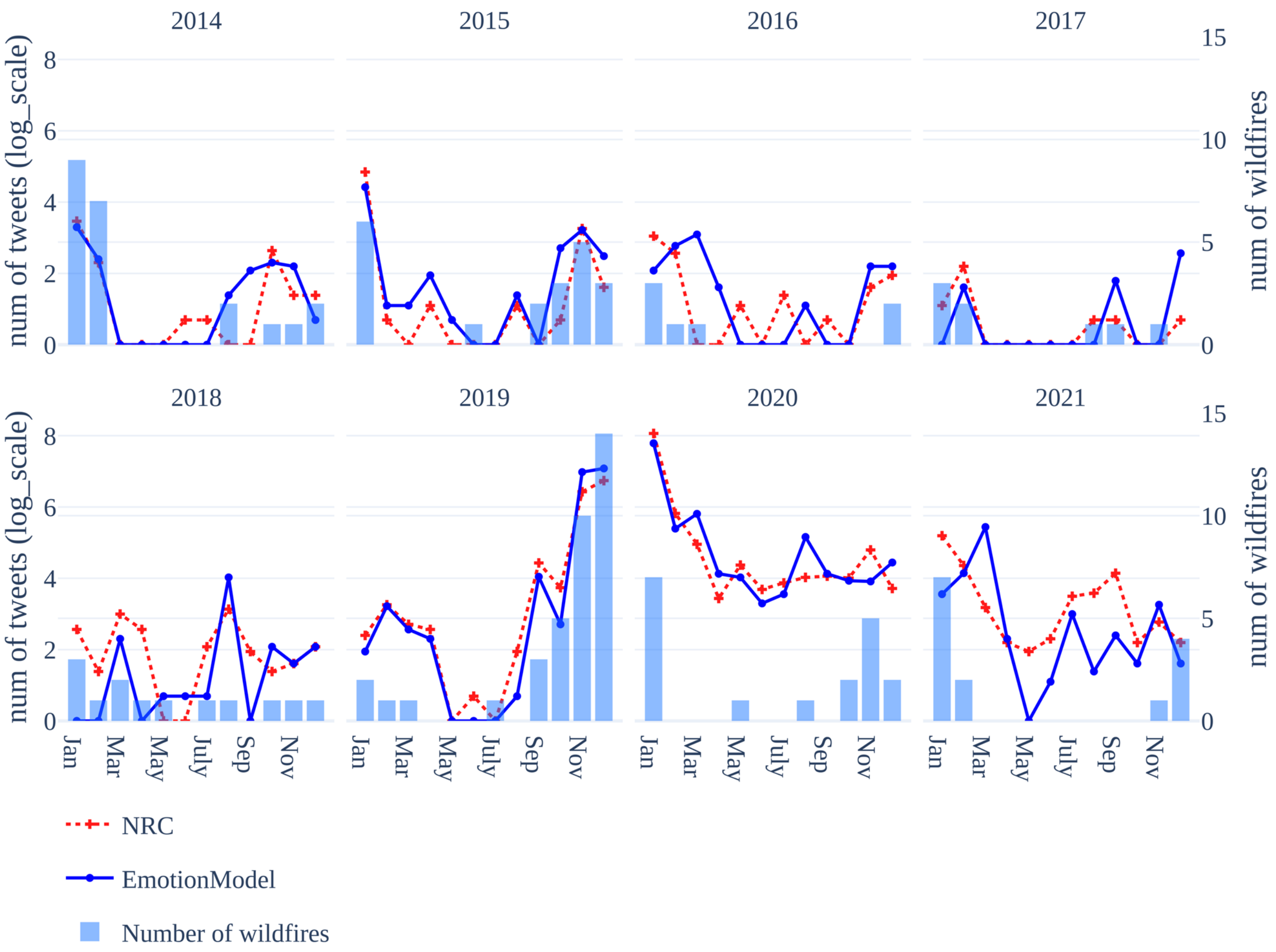


Fig. 24 Positive tweets predicted by the NRC lexicon and the Emotion model

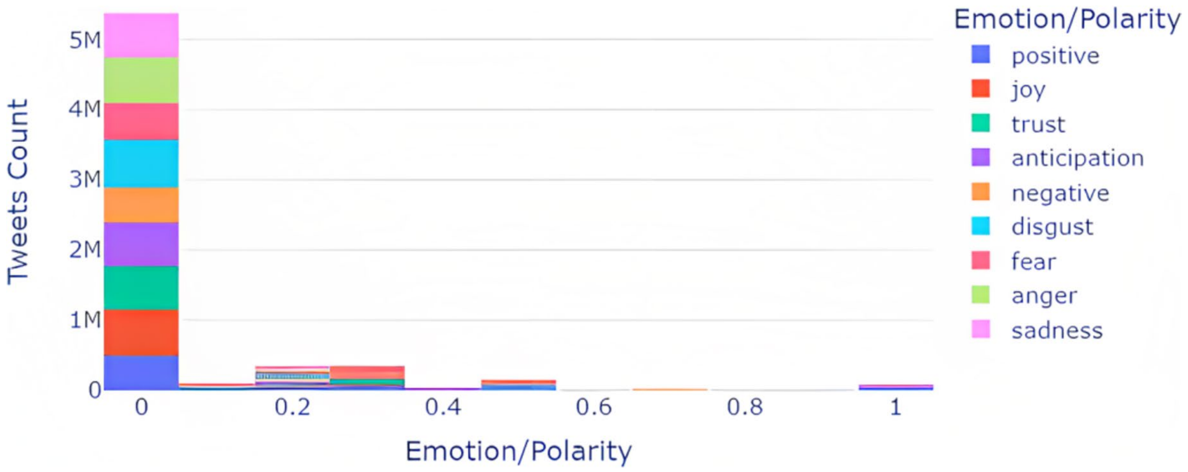


Fig. 25 Distribution of NRC scores across emotions before filtering

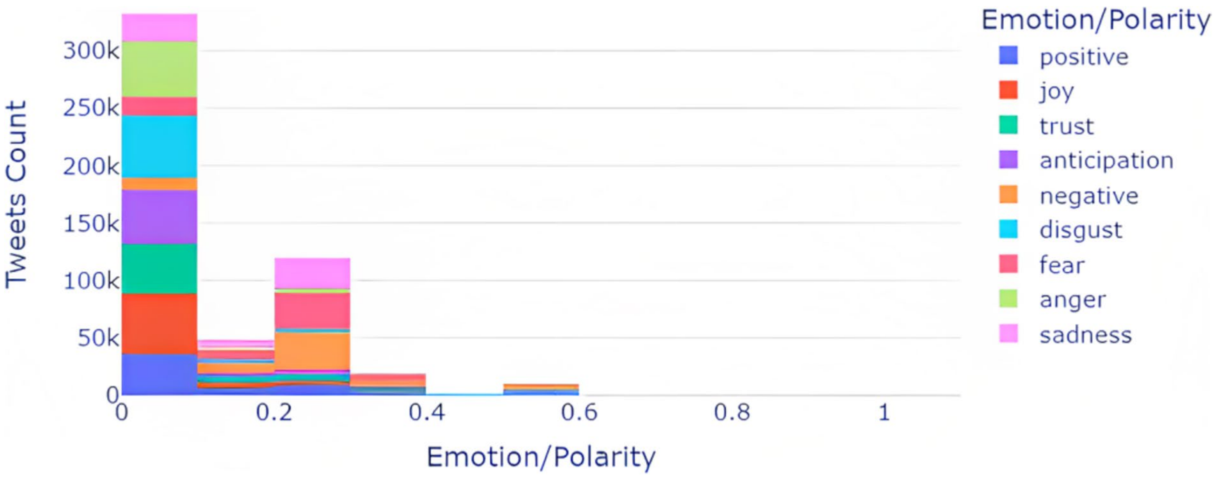


Fig. 26 Distribution of NRC scores across emotions after filtering

Other emotions and sentiments toward Australian wildfires over time, predicted by the Emotion Model and the NRC lexicon.

Acknowledgements This research is supported by the Agency for Science, Technology and Research (A*STAR) under its AME Programmatic Funding Scheme (Project #A18A2b0046). We also appreciate the support from MOE Tier 2 Grant 22-5191-A0001-0: 'Prediction-to-Mitigation with Digital Twins of the Earth's Weather'. In addition, Cuc Duong is supported by Nanyang Research Scholarship.

Data Availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent There is no human participant involved in this research.

Conflict of Interest The authors declare they have no conflict of interest.

References

1. Allan RP, Hawkins E, Bellouin N, Collins B. IPCC, 2021: Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Cambridge University Press; 2021.
2. Blanchi R, Leonard J, Haynes K, Opie K, James M, de Oliveira FD. Environmental circumstances surrounding bushfire fatalities in Australia 1901–2011. Environ Sci Policy. 2014;37:192–203.
3. Richards L, Brew N, Smith L. 20 Australian bushfires—frequently asked questions: a quick guide (Parliament of Australia, 2020). 2019.
4. Cowlishaw S, Metcalf O, Varker T, Stone C, Molyneaux R, Gibbs L, Block K, Harms L, MacDougall C, Gallagher HC, et al. Anger dimensions and mental health following a disaster: Distribution and implications after a major bushfire. J Trauma Stress. 2021;34(1):46–55.
5. Li M, Shen F, Sun X. 2019–2020 Australian bushfire air particulate pollution and impact on the South Pacific Ocean. Sci Rep. 2021;11(1):1–13.
6. van Valkengoed AM, Steg L. Meta-analyses of factors motivating climate change adaptation behaviour. Nat Clim Change. 2019;9(2):158–63.
7. Goldenberg A, Gross JJ. Digital emotion contagion. Trends Cogn Sci. 2020;24(4):316–28.
8. Luo T, Cao Z, Zeng D, Zhang Q. A dissemination model based on psychological theories in complex social networks. IEEE Trans Cogn Develop Syst. 2021;14(2):519–31.
9. Cambria E, Schuller B, Liu B, Wang H, Havasi C. Statistical approaches to concept-level sentiment analysis. IEEE Intell Syst. 2013;28(3):6–9.
10. Amin M, Cambria E, Schuller B. Will affective computing emerge from foundation models and General AI? A first evaluation on ChatGPT. IEEE Intell Syst. 2023;38(2):15–23.
11. Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. J Mach Learn Res. 2003;3:993–1022.
12. Duong C, Liu Q, Mao R, Cambria E. Saving Earth one tweet at a time through the lens of artificial intelligence. In: 2022 International Joint Conference on Neural Networks (IJCNN), p. 1–9, 2022.
13. Mao R, Li X. Bridging towers of multitask learning with a gating mechanism for aspect-based sentiment analysis and sequential metaphor identification. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence, p. 13534–42, 2021.
14. Kirilenko AP, Stepchenkova SO. Public microblogging on climate change: One year of Twitter worldwide. Glob Environ Change. 2014;26:171–82.
15. Kirilenko AP, Molodtsova T, Stepchenkova SO. People as sensors: Mass media and local temperature influence climate change discussion on Twitter. Glob Environ Change. 2015;30:92–100.
16. Dahal B, Kumar SAP, Li Z. Topic modeling and sentiment analysis of global climate change tweets. Soc Netw Anal Min. 2019;9(1):1–20.
17. Willson G, Wilk V, Sibson R, Morgan A. Twitter content analysis of the Australian bushfires disaster 2019–2020: Futures implications. J Tour Futures. 2021.
18. Mao R, Lin C, Guerin F. Word embedding and WordNet based metaphor identification and interpretation. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, vol. 1, p. 1222–31, 2018.

19. Mao R, Li X, Ge M, Cambria E. Metapro: A computational metaphor processing model for text pre-processing. Inf Fusion. 2022;86–87:30–43.
20. Mao R, Liu Q, He K, Li W, Cambria E. The biases of pre-trained language models: An empirical study on prompt-based sentiment analysis and emotion detection. IEEE Trans Affect Comput. 2023.
21. Strapparava C, Valitutti A. WordNet affect: An affective extension of WordNet. In: Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04), Lisbon, Portugal. European Language Resources Association (ELRA); 2004.
22. Esuli A, Sebastiani F. SENTIWORDNET: A publicly available lexical resource for opinion mining. In: Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06), Genoa, Italy. European Language Resources Association (ELRA); 2006.
23. Cambria E, Liu Q, Decherchi S, Xing F, Kwok K. SenticNet 7: A Commonsense-based Neurosymbolic AI Framework for Explainable Sentiment Analysis. In: LREC, p. 3829–39, 2022.
24. Mohammad SM, Turney PD. Crowdsourcing a word-emotion association lexicon. Comput Intell. 2013;29(3):436–65.
25. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. Adv Neural Inf Process Syst. 2013;26.
26. PenningtonJ, Socher R, Manning CD. Glove: Global vectors for word representation. In: Proceedings of The 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), p. 1532–43, 2014.
27. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, p. 4171–86. Association for Computational Linguistics; 2019.
28. Mao R, Lin C, Guerin F. End-to-end sequential metaphor identification inspired by linguistic theories. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, p. 3888–98, 2019.
29. Ge M, Mao R, Cambria E. Explainable metaphor identification inspired by conceptual metaphor theory. In: Proceedings of AAAI, p. 10681–9, 2022.
30. He K, Mao R, Gong T, Li C, Cambria E. Meta-based self-training and re-weighting for aspect-based sentiment analysis. IEEE Trans Affect Comput. 2022.
31. Li W, Zhu L, Mao R, Cambria E. SKIER: A symbolic knowledge integrated model for conversational emotion recognition. Proc AAAI Conf Artif Intell. 2023;37(11):13121–9.
32. Torregrosa J, D'Antonio-Maceiras S, Villar-Rodríguez G, Hussain A, Cambria E, Camacho D. A mixed approach for aggressive political discourse analysis on Twitter. Cognit Comput. 2023;15(2):440–65.
33. Han S, Mao R, Cambria E. Hierarchical attention network for explainable depression detection on twitter aided by metaphor concept mappings. In: Proceedings of the 29th International Conference on Computational Linguistics, p. 94–104, 2022.
34. Yue T, Mao R, Wang H, Hu Z, Cambria E. KnowleNet: Knowledge fusion network for multimodal sarcasm detection. Inf Fusion. 2023;100:101921.
35. Moritz MA, Morais ME, Summerell LA, Carlson JM, Doyle J. Wildfires, complexity, and highly optimized tolerance. Proc Natl Acad Sci. 2005;102(50):17912–7.
36. Penman TD, Bradstock RA, Price O. Modelling the determinants of ignition in the Sydney basin, Australia: Implications for future management. Int J Wildland Fire. 2012;22(4):469–78.
37. Price C, Rind D. Possible implications of global climate change on global lightning distributions and frequencies. J Geophys Res Atmos. 1994;99(D5):10823–31.
38. Goldammer JG, Price C. Potential impacts of climate change on fire regimes in the tropics based on magicc and a giss gcm-derived lightning model. Clim Change. 1998;39(2):273–96.
39. Linnenluecke M, Marrone M. Air pollution, human health and climate change: Newspaper coverage of Australian bushfires. Environ Res Lett. 2021.
40. Wikipedia Contributors. 2013–2014 Australian bushfire season. https://en.wikipedia.org/wiki/2013-14_Australian_bushfire_season. Accessed 22 Aug 2022.
41. Wikipedia Contributors. 2014–2015 Australian bushfire season. https://en.wikipedia.org/wiki/2014-15_Australian_bushfire_season. Accessed 22 Aug 2022.
42. Wikipedia Contributors. 2015–2016 Australian bushfire season. https://en.wikipedia.org/wiki/2015-16_Australian_bushfire_season. Accessed 22 Aug 2022.
43. Wikipedia Contributors. 2016–2017 Australian bushfire season. https://en.wikipedia.org/wiki/2016-17_Australian_bushfire_season. Accessed 22 Aug 2022.
44. Wikipedia Contributors. 2017–2018 Australian bushfire season. https://en.wikipedia.org/wiki/2017-18_Australian_bushfire_season. Accessed 22 Aug 2022.
45. Wikipedia Contributors. 2018–2019 Australian bushfire season. https://en.wikipedia.org/wiki/2018-19_Australian_bushfire_season. Accessed 22 Aug 2022.
46. Wikipedia Contributors. 2019–2020 Australian bushfire season. https://en.wikipedia.org/wiki/2019-20_Australian_bushfire_season. Accessed 22 Aug 2022.
47. Wikipedia Contributors. 2020–2021 Australian bushfire season. https://en.wikipedia.org/wiki/2020-21_Australian_bushfire_season. Accessed 22 Aug 2022.
48. Manning CD, Surdeanu M, Bauer J, Finkel JR, Bethard S, McClosky D. The Stanford CoreNLP natural language processing toolkit. In: Proceedings of 52nd Annual Meeting of The Association for Computational Linguistics: System Demonstrations, p. 55–60, 2014.
49. Jin H, Song Q, Hu X. Auto-Keras: An efficient neural architecture search system. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, p. 1946–56, 2019.
50. Turbé H, Bjelogrlic M, Lovis C, Mengaldo G. Evaluation of post-hoc interpretability methods in time-series classification. Nat Mach Intell. 2023;5(3):250–60.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.