Taylor & Francis
Taylor & Francis Group

Journal of Location Based Services

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tlbs20

Exploring the potential of social media to study environmental topics and natural disasters

Kenzo Milleville, Samuel Van Ackere, Jana Verdoodt, Steven Verstockt, Philippe De Maeyer & Nico Van de Weghe

JOURNAL OF LOCATION BASED SERVICES
https://doi.org/10.1080/17489725.2023.2238663

Exploring the potential of social media to study environmental topics and natural disasters

Kenzo Milleville [a,b], Samuel Van Ackere [a], Jana Verdoodt [a], Steven Verstockt [b], Philippe De Maeyer [a] and Nico Van de Weghe [a]

[a]CartoGis, Ghent University, Ghent, Belgium; [b]IDLab - Imec, Ghent University, Ghent, Belgium

ABSTRACT
Social media has become an important means of communication and new insights can be gained from processing this data on a large scale. Our goal is to develop and implement a pipeline to automatically extract and analyse Twitter data on natural disasters and environmental topics. We aim to provide an additional layer of spatiotemporal data that can be used to study the immediate and lasting impacts of natural disasters, climate change, and environmental topics on the global population. An initial analysis of forest fires was conducted in four different languages confirming the need for multilingual support for global analysis. We found a positive correlation between wildfire occurrence and tweeting behaviour, as well as the geographic spread of fires. We found that simple sentiment predictions add little value when aggregating data on a large scale. A subsequent test using a fine-tuned stance detection model proved promising in determining the stance of tweets towards nuclear energy. We intend to expand our dataset and develop customised models in the future that can be used to analyse the global impact of natural disasters and environmental topics.

ARTICLE HISTORY
Received 15 February 2023
Accepted 15 July 2023

KEYWORDS
Social media; Geospatial data analysis; Natural language processing; Sentiment analysis

1. Introduction

In recent years, governments at regional, national, and supranational levels have invested heavily in the standardisation (e.g. Inspire), collection (e.g. Copernicus), and use of remote sensing data to monitor environmental parameters. The qualitative development of geospatial information technologies and services over the past two decades has led to a dramatic increase in the amount of data that can be used to assess the state of the environment. Although the amount of data has increased substantially, the quality of decisions made based on this data has not improved much (Kuemmerle et al. 2015). Remote sensing data collection is typically performed using satellite imagery. This method requires significant investment in hardware and software to process the data into indicators. Various characteristics describing the ecological condition of areas

CONTACT Kenzo Milleville kenzo.milleville@ugent.be

have been derived from satellite images taken at different moments in time, resulting in spatiotemporal indicators (Tikunov et al. 2017). Atlas information systems and their qualitative advancements, such as Digital Earth, are becoming increasingly important, especially in the field of environmental management. New sources of information and new approaches to aggregate and analyse information greatly expand the ability to monitor the ecological condition of areas and support environmental decision-making. In addition to remote sensing data, new types of data are being incorporated into environmental studies. Research in citizen science, crowdsourcing, and social media has shown the potential to gather both knowledge and insights to address environmental emergencies (Horita et al. 2013; Simon, Goldberg, and Adini 2015). We propose to develop an information and analysis system that uses social media data to analyse the global impact of environmental topics and natural disasters. Much of this research will focus on the collection and integration of geolocated social media data. Although the use of social media data is useful for disaster management, further research is needed to sift through the (potentially) interesting information and provide valuable insights for environmental indicators (de Albuquerque et al. 2016).

The biggest challenges lie in collecting and aggregating the myriad of data across different technologies. As each social media post in itself provides little additional information, they must be aggregated both semantically (via natural language processing) and spatiotemporally (via clustering techniques). Once the data is aggregated, the next step is to explore the transformation of this information into valuable indicators at local and global levels.

Using social media data, we aim to provide an additional dimension to environmental data by analysing public opinion about different environmental topics and studying the impact of natural disasters. We expect to find regional differences on these topics and want to examine how public opinion has changed over time. Do people think environmental topics, such as global warming and renewable energy, are important? If so, when did this change occur, and can we hypothesise about what caused this change in mentality? Concerning natural disasters, we will study their immediate impact online and determine whether they had a long-term impact on people's views. For instance, did the forest fires around the world reignite the discussion about global warming? Did the tragedy in Fukushima lead to a negative change in mentality about nuclear energy? Especially when it comes to polarising topics, this data can provide new insights. We hope to answer most of these questions and provide a framework that will allow other researchers and legislators to freely access the data and gain insights to answer similar questions.

The paper is organised as follows: Section 2 discusses related work regarding natural language processing, sentiment analysis, and studies using social media data for disaster management. Section 3 details the tweet collection, processing, and geocoding. Section 4 presents the obtained results and visualisations.

Section 5 presents a discussion of our approach and results. The paper finishes with a conclusion in Section 6.

2. Related work

To automatically analyse large text datasets, Natural Language Processing (NLP) techniques can be used. NLP is the field of computer science that deals with the automatic extraction of information from (unstructured) text (Cambria and White 2014). With recent advances in neural network architectures, faster computing, and larger datasets to train on, the field has made tremendous progress. Many NLP models can be used immediately on new, unseen datasets and still perform well. Sentiment analysis is a popular NLP technique to predict an author's sentiment from their text. Usually, sentiment is denoted as a class (positive, negative, or neutral) or as a numerical value (−1 to 1). For most people, it is often straightforward to determine the sentiment of a given text. However, due to the complexity of natural language, the small amount of text per tweet, sarcasm, and the unique vocabulary used in certain subcultures, it can be difficult to determine sentiment in an automated way.

A related problem is stance detection, which involves determining whether the author is in favour, against, or neutral towards a given statement or topic. The author's stance can either be explicitly mentioned or implied in the text. In Mohammad et al. (2016), this problem was posed as a supervised learning task by annotating a dataset of tweets on five different topics: atheism, climate change, feminism, Hillary Clinton, and abortion. The annotated data was used to create the SemEval-2016 Task 6 challenge. The highest F_{avg} score of the participating teams was 67.8, indicating that this problem is not easy to solve. In recent work, these scores have been improved by using large pre-trained language models. One popular model, BERT, is a state-of-the-art language model that uses a bidirectional transformer architecture (Devlin et al. 2019). This model was pre-trained on huge corpora of unlabelled text, allowing rapid fine-tuning on a wide range of NLP tasks. Even though these models can outperform traditional techniques, they typically require labelled data for each specific topic or statement to fine-tune. Automatically determining the specific topic or statement being talked about positively or negatively is even more challenging.

In the event of a (natural) disaster, social media users produce many posts with disaster-related information that can be useful for analysis (Acar and Muraki 2011; Muralidharan et al. 2011; Ukkusuri et al. 2014). For instance, Neppalli et al. (2017) found that extracting sentiments during Hurricane Sandy could help emergency responders develop better situational awareness of the disaster area. In another study, a BERT model was applied to a set of tweets related to the Jakarta floods in early 2020 to identify relevant tweets that could provide information on disaster response (Maharani 2020). Besides natural disasters, previous research has shown correlations between mode of travel and tweet sentiment (Greg et al. 2018)), between temperature anomalies and tweeting behaviour (Kirilenko, Molodtsova, and Stepchenkova 2015), and between people's concern about climate change and the severity of weather anomalies (Sisco, Bosetti, and Weber 2017). By analysing tweets made with the hashtag #WorldEnvironmentDay, Reyes-Menendez, Ramón Saura, and Alvarez-Alonso (2018) found that certain environmental topics (climate change, clean water, and pollution) carried a negative sentiment. While other topics such as public health and clean energy, were rather positive. These results can potentially be used by NGOs or policymakers to focus on the most concerning topics. In Zotova, Agerri, and Rigau (2021), a semi-automatic method was developed to label over 20,000 tweets regarding their stance on the independence of Catalonia. Their method greatly speeded up the labelling process by leveraging user-based relations. Their best models were based on the BERT architecture and achieved an F_{avg} score of 0.7468 and 0.7472 on Catalan and Spanish tweets, respectively.

More recently, large language models such as ChatGPT (GPT-3.5), are being used for sentiment analysis or stance detection. These models can be applied in a zero-shot fashion (using no labelled data or examples) and achieve close to state-of-the-art performance on certain benchmarks (Wei et al. 2022; Zhang, Ding, and Jing 2023). One study even found that ChatGPT outperformed crowd workers to label tweets, for a fraction of the cost Gilardi, Alizadeh, and Kubli (2023). While these large language models are computationally intensive, they are becoming increasingly popular to tackle a wide range of NLP tasks.

3. Methods

The proposed procedure for investigating public opinion and spatiotemporal differences related to natural disasters consists of four phases: tweet collection, tweet processing, tweet georeferencing, and analysis.

3.1. Collection and processing of tweets

Over the past decade, social media has become an integral part of global communication and is therefore frequently used as a data source for large-scale analyses. Twitter, a social network where users can send tweets (short messages of up to 140 characters, increased to 280 characters in 2017), is often used to collect such data. It has an open API for research purposes with detailed query functionalities. It is estimated that over 500 million tweets are sent daily, allowing for extensive data collection on virtually any topic. However, many people do not use Twitter, preferring alternatives such as Facebook, Reddit, and Sina Weibo (popular in China). We are aware that these alternatives exist and that their exclusion could lead to geospatial bias, but we will focus only on

Twitter data to limit the scope of the project. However, the proposed methods can be applied to virtually any social media platform.

First, a query relating to forest fires was performed in four languages: English, Spanish, Chinese, and Russian. Each query consisted of multiple writing variations on the topic of forest fires. The tweets were collected from January 2012 until August 2021. The last two years of tweets were processed and analysed in detail. Table 1 shows the number of retrieved tweets in those last two years, the percentage of geotagged tweets, the number of user locations found, and the percentage that was successfully geolocated using our algorithm (see 3.2). Most tweets are written in English or contain some English keywords. Although Chinese is the second most popular language in the world, we found almost 30 times more tweets searching with English keywords. This is likely because Twitter is less popular in Chinese-speaking countries and due to translation errors. Additionally, many viral keywords, hashtags, and trends are often written in English. This causes many non-native English speakers to tweet in English or use such keywords to increase their reach. Using the estimated language provided by Twitter, we find that 94% of the tweets found with English keywords were written in English.

Because tweets can be sent in any language, this complicates both the retrieval of tweets and their analysis. A popular approach is to translate all collected tweets into a common language (usually English) and then process them using language-specific models. To query topics in different languages, keywords or hashtags have to be translated using available translation services and models. However, sometimes these automatic translations do not reflect the correct translation, or there are several common spellings for the specified topic. For instance, if you aim to collect tweets regarding the coronavirus epidemic, you should use multiple related keywords, such as”covid””,coronavirus”, and”corona epidemic”. Even after searching for all possible spelling variations of the topic, many tweets related to the topic may not be found. The collected tweets consisted of multiple fields that contained additional information about the tweet itself. Table 2 lists some of the most important fields with their explanations.

To perform a large-scale spatiotemporal analysis, we require coordinate information. Twitter allows users to tag their tweets with the exact coordinates of their location (geotag). Our findings show that less than 2.5% of all collected tweets were geotagged. Therefore, we need to rely on the location provided in

Table 1. The number of tweets and user locations found for each query language.

Language	Total tweets (thousands)	Geotagged tweets (%)	User locations (thousands)	Geolocated user locations (%)
English	2,465	2.32	1,890 (76.7%)	69.7
Spanish	433	2.62	349 (80.6%)	73.8
Chinese	84	0.45	46 (54.4%)	12.4
Russian	17	1.03	12 (69.0%)	28.0

Table 2. Description of the tweet fields used in this work.

Field name	Explanation
text	The content of the tweet
lang	Language of the tweet, detected by Twitter
created_at	Creation time of the tweet
id	Unique identifier of this tweet
author_id	Unique identifier of this user
user_location	(Optional) User-submitted profile location in plain text
geo	(Optional) Details & coordinates of the geotagged location

the user's Twitter profile to determine an approximate location. Between 60–80% of users provided their location in plain text. Users can enter anything as their location, so many locations will not provide useful information. For instance, one query found that over 10% of users listed their location as 'earth' or 'planet earth'. To determine the coordinates that relate to these location names, geocoders can be used.

3.2. Geocoding

Geocoding is the process of transforming a location description (address or place name) into a coordinate on the earth's surface. There are a variety of algorithms for geocoding, but they all follow roughly the same process. First, the address to be geocoded is entered in plain text. Then, the address is normalised into an acceptable format (usually street name, house number, city name, and postal code). Finally, an iterative comparison of this address with a reference dataset (e.g. a street and city database) is performed, from which the geographic coordinates of the address can be calculated (Yang et al. 2004).

Geocoders are usually accessed via a REST API. Popular examples include Google Geocoding[1] and the open-source solutions Geonames[2] and Nominatim.[3] These APIs have tight limitations and can become expensive to geocode millions of user locations. Therefore, we developed a simple algorithm to geocode popular locations and place names. Our method used a reference dataset containing all countries, their main cities, and provinces.[4] In total, this dataset contained about 43,000 place names. An iterative algorithm was developed that considered exact string matches of the found place names with this reference dataset.

First, the user location was queried for a country name. If it contained a country name, we checked if it contained the name of a city or province/state of that country. If it did, the coordinates were extracted, favouring cities over provinces, as they provide more localised information. If no country was found, we checked for a matching province or state and city name (e.g". Nashville, TN" was matched to"Nashville, Texas"). If no province or state was found, we queried for just a city name. If multiple matches were found with the same city name, the most populous option was chosen (e.g".Paris" was matched to"Paris, France" rather than"Paris, Texas"). This disambiguation could be further improved by considering the language of the tweet and the native language of the matching countries. The algorithm worked quite well with place names written in the Latin alphabet and was able to geocode 70.3% of the user locations (see Table 1). However, it performed poorly in the initial tests with other alphabets (Russian and Chinese). This is because the reference dataset often does not contain place names in the local alphabet. After geocoding all unique user locations with this algorithm, the locations that did not result in a match can be geocoded with a public API, greatly reducing the number of requests.

3.3. Sentiment analysis and stance detection

After processing and geocoding the collected tweets, sentiment analysis was performed using Textblob. Textblob is a popular sentiment analysis model, available as an open-source Python library (Loria 2018). In addition to sentiment analysis, the library provides a consistent API for common NLP tasks such as part-of-speech tagging, noun phrase extraction, and more. Textblob determines the sentiment with a predefined dictionary that classifies negative and positive words. All words in the analysed sentence receive an individual score depending on whether they are positive or negative. A pooling operation, such as the average of all sentiments, is then used to calculate the final sentiment. TextBlob provides two types of information about the input sentiment: polarity and subjectivity. Polarity ranges from [−1,1], where −1 represents a negative sentiment and 1 represents a positive sentiment. Subjectivity ranges from [0,1] and tries to distinguish facts from opinions. Higher subjectivity means that the text contains personal opinions rather than factual information. The tweets were first preprocessed by removing all mentions (@username), URLs, and hashtag symbols. Then, both polarity and subjectivity were calculated using Textblob. Using the predicted polarity, the goal was to visualise how public opinion varies spatiotemporally. We found that tweets related to natural disasters were not very polarising and were difficult to analyse after aggregation on a large scale (see Section 4).

Therefore, an additional test was conducted on tweets about alternative energy sources (nuclear, solar, wind), which represented a more polarising topic. Instead of using a generic sentiment analysis model, we fine-tuned a language model for stance detection on a small subset of the collected tweets related to nuclear energy. The tweets were manually labelled as either in favour, against, or neutral (neither) towards nuclear energy as an alternative energy source. During the labelling process, we found many irrelevant tweets. Some discussed nuclear weapons, some were job ads, and some had nothing to do with nuclear energy but contained one or more keywords. Due to the large number of irrelevant tweets, we added irrelevance as an additional label. A total of 500 tweets were labelled using the open-source tool Label Studio (Tkachenko et al. 2020). These labelled tweets were then used to fine-tune a BERTweet model (Nguyen, Thanh, and Tuan Nguyen 2020)), which is a pre-trained language model that uses a similar architecture as BERT. BERTweet was pre-trained on large corpora of English tweets and outperformed other pre-trained models on NLP tasks on tweets. In addition, the model and code are released under an open-source licence.

4. Results

4.1. Sentiment analysis on forest fires

To compare our results related to wildfires, the international disaster database EM-DAT[5] of the Centre for Research on the Epidemiology of Disasters was used. All Spanish tweets related to wildfires (from January 2012 to July 2021) were collected, geocoded, and filtered for tweets posted from Spain. Figure 1 shows these tweets along with some of the major wildfires in Spain reported in the EM-DAT database. There is clearly a recurring pattern in posts about forest fires during the summer. There is a clear overlap between the Twitter data and the wildfire occurrences, whether at the local (in Spain) or global level. For instance, four peaks in the Twitter data correspond to reported wildfires in Spain: July 20 June 201217 October 2017, and July 2021.

Remarkably, the large spike in Spanish tweets in August 2019 did not coincide with a reported wildfire in Spain. The only major wildfires reported in the EM-DAT database were in Australia (New South Wales, Queensland). Looking at the English tweets posted in Europe dealing with wildfires, this peak is also noticeable. We can conclude that it is possible to detect the occurrence of wildfires using Twitter data. However, user locations do not indicate where these wildfires are occurring, as many people across the globe tweet about major wildfires. Further analysis of the text content is needed to determine the wildfire location.

Figure 2 shows the complementary spatial distributions of Spanish and English tweets related to wildfires. There is a clear correlation between

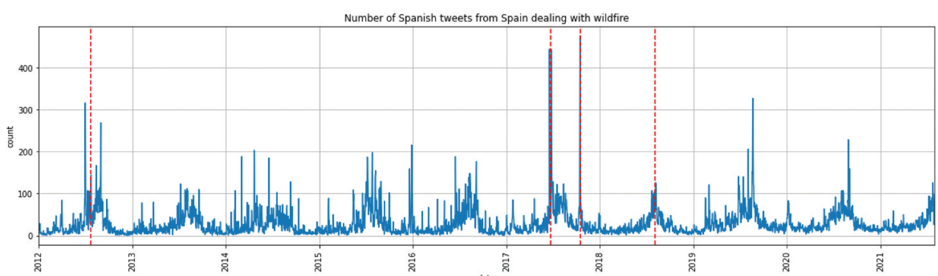

Figure 1. Number of Spanish tweets from Spain dealing related to wildfires, red dashed lines represent some major reported wildfires in Spain from EM-DAT.

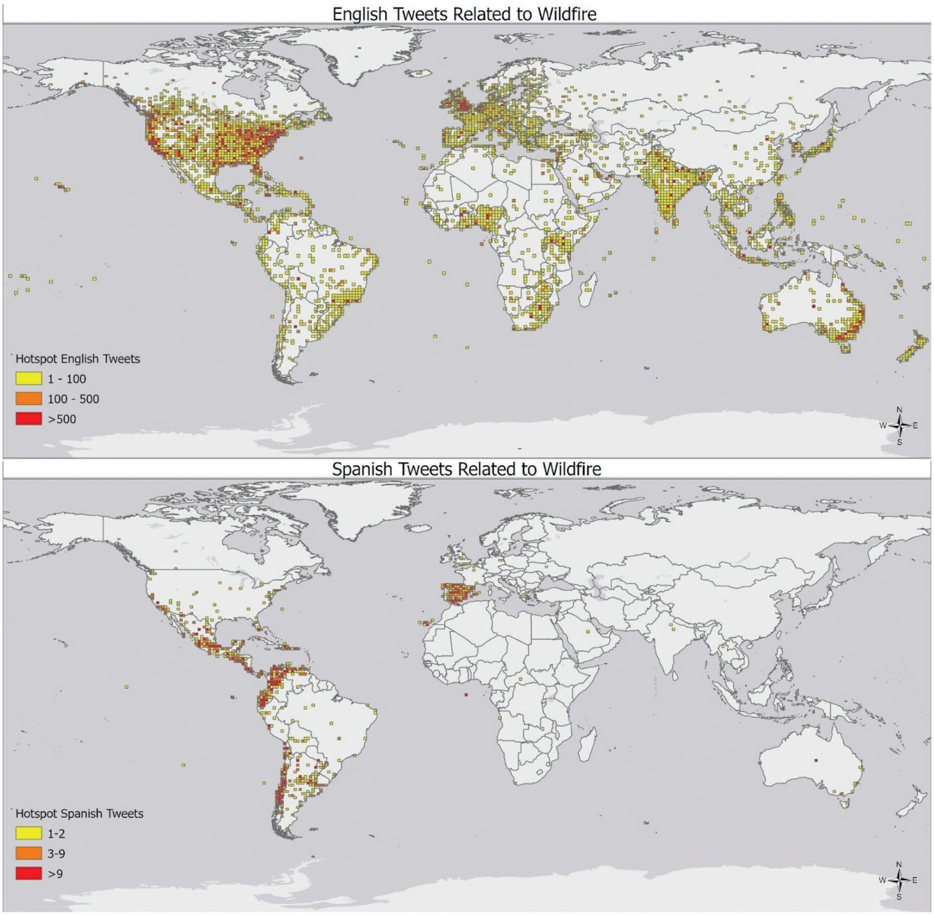

Figure 2. The distribution of English (top) and Spanish (bottom) tweets related to wildfires.

Spanish user locations and countries where Spanish is the official native language. For the English query, we see better global coverage, showing that these English keywords and hashtags are used by many non-native speakers. Furthermore, we found that over 99% of the tweets found with the Spanish query were unique and not included in the English query.

After taking a closer look at many tweets and their predicted sentiment, we concluded that the Textblob model cannot accurately assess sentiment. Many tweets contain relevant keywords or hashtags (e.g. forest fire, wildfire), but are irrelevant to the topic. We suspect that many of the viral hashtags related to wildfires are used to gain more reach for an individual's tweets, even though the tweet is unrelated to wildfires. These irrelevant tweets heavily influence the results, therefore, they either need to be filtered out beforehand or the model should ignore them. Some positive and negative predicted tweets are shown in Table 3. Interestingly, the model predicted many tweets with a positive sentiment, most of which were irrelevant with

Table 3. Some sample tweets related to wildfires, grouped by their predicted sentiment. Many of the collected tweets contained viral hashtags related to wildfires but were irrelevant.

Tweets with negative predicted sentiment	Irrelevant
Sam Wood and Snezana Markoski raise $20,000 in donations for bushfire relief in just 24 hours The Bachelor's Sam Wood and Snezana Markoski are doing their part in helping Australians affected by the devastating bushfire crisis	
There will be a day That all the diabolical and evil deeds of these politicians will be met by a raging wildfire that will engulf them and riches they have robbed this nation of.	X
Neguse Curtis Launch Bipartisan Wildfire Caucus Introduce Legislation to Help Communities Recover From Devastating 2020 Wildfire Season. TY Sen Neguse!	
My brain cannot wrap itself around a fire crossing the continental divide How can a wildfire reach 11 −12,000 feet Absolutely insane.	
Tweets with positive predicted sentiment	Irrelevant
You can find them best the year after a forest fire.	X
Best of luck to all the nominees #Wolfwalkers #DatingAmber #Wildfire #Vivarium #HereAreTheYoungMen #SeaFever	X
Beautiful sunrise underway in Missoula courtesy of the wildfire smoke! You can expect hazy skies again today but lessening going into tomorrow MTwx	
Man this bird is awesome #lyrebirds #leonardthelyrebird #bluemountains #AustralianBushfires	X
What a brilliant idea watch it catch on like wildfire!	X

regards to wildfires. Furthermore, when the sentiments are aggregated over large areas, they tend to average out to neutral and provide little insight into the public opinion.

4.2. Stance detection on nuclear energy

Out of the 500 manually labelled tweets for stance detection regarding nuclear energy, 98 were labelled as irrelevant. Of the 402 others, 169 were labelled as"in favour", 86 as"neither", and 147 as"against". Two tests were performed: one to predict the tweet's relevance and one to predict the author's stance with respect to nuclear energy. For both tests, 20% of the data was used for validation (100 tweets). The relevance prediction performed surprisingly well, with an F_1 score of 0.92. The fine-tuned model was clearly able to distinguish tweets related to nuclear energy from unrelated tweets.

Because we are mainly interested in favourable or negative opinions, the irrelevant tweets were considered as"neither" for stance detection. To evaluate the overall performance, we used the macro-average of the F_1 scores (denoted as F_{avg}) for the"in favour" and"against" classes. This is the same metric that was used in (Mohammad et al. 2016). The stance detection was less accurate than the relevance prediction with an F_{avg} of 0.67. The model was also much better at predicting the favourable class. For completeness, Table 4 lists the precision, recall, and F_1 scores for each class in the validation set.

Taking a closer look at some of the incorrect predictions on the validation set, we saw that the model sometimes made confident mistakes. Other times, none of the predictions had a high probability, so these could be ignored by using a threshold. For instance, if we only consider predictions with a minimum threshold of 0.75, the F_{avg} score rises to 0.765, but at the cost of discarding

Table 4. Validation scores for each class of the stance detection.

Stance	Precision	Recall	F_1 score
Against	.63	0.63	0.63
Neither	.81	0.69	0.75
In favor	.67	0.76	0.71

Table 5. Sample tweets from the validation set with incorrect predictions and associated scores and labels.

Tweet text	Prediction	Label
I've been reading a book about the Chernobyl accident and it's had me thinking. Considering how the Russian government botched the building and managing of those reactors, imaging the disaster if the trump admin were to attempt something like nuclear energy.	Neither (.848)	Against
@CKscullycat Not to mention, nuclear power plants	Against (.696)	Neither
Observing the #printergate debacle, I think it was wise we eschewed nuclear energy.	Favor (.449)	Against
@GavinNewsom How about spending money on infrastructure, nuclear power, etc. . . to accommodate the CA population's need for energy? Just like H2O, with proper planning these "emergencies" can be avoided	Against (.575)	Favor
went down a nuclear energy rabbit hole tonight like how did we not ditch the whole "atomic age" thing after chernobyl? fukushima? we're really still out here burying radioactive waste in concrete sarcophaguses in 2020? wild	Neither (.393)	Against

52% of the tweets in the validation set. Some example tweets with incorrect predictions are presented in Table 5.

5. Discussion

In this paper, we presented a generic pipeline for spatiotemporal analysis of tweets on environmental topics and our preliminary results. We showed that simple sentiment analysis models often underperform on tweets. Furthermore, the predicted sentiment does not provide sufficient information to perform an in-depth analysis of public opinion when aggregated over larger regions.

The relatively simple geocoding algorithm was able to geocode 70.3% of the collected tweets in the Latin alphabet by using the locations of the users in their Twitter bio. The locations that did not yield a match can be geocoded using a public API, greatly reducing the number of queries. These user locations are critical because less than 2.5% of all tweets were geotagged. However, when using geocoding APIs, certain user locations such as"earth" and"nowhere" can match a real place name, resulting in false positives. Automatically removing these false positive matches will be a challenge.

Upon closer examination of the collected tweets, we found that many of them were irrelevant to the queried topic. Many news reports, job ads, or tweets on similar topics (e.g. nuclear weapons) contained some of the keywords. The inclusion of these irrelevant tweets will lead to an overestimation of the number of tweets and people discussing the topics at hand. However, we showed that it is possible to accurately filter out irrelevant tweets by fine-tuning a language model. While this approach produced good results, it can be time-consuming when applied to the full dataset of millions of tweets. Additionally, this approach was tested for a single topic (nuclear energy). Future research will show whether a single model can be used to filter out most irrelevant tweets across topics, or whether a separate model is needed for each topic. We estimate that news reports, job ads, financial information, and other similarly structured irrelevant tweets can be automatically filtered out.

The retrained BERTweet model for stance detection regarding nuclear energy achieved an F_{avg} score of 0.67. Considering that the model was only trained on 400 tweets and validated on the remaining 100, this is a promising result. When analysing the incorrect model predictions, we saw that many of them were replies to another tweet, were too short, or were written in a convoluted way where it is difficult to determine the stance without additional context. These problems were also mentioned in (Lai et al. 2020; Zotova, Agerri, and Rigau 2021). Although our analysis focused solely on tweets, the discussed methods can be applied with little adjustments to other social media platforms featuring text-based content such as Facebook and Reddit.

Our goal is to label additional data for stance detection in queries about alternative energy sources (nuclear, solar, wind, etc.) to visualise the spatiotemporal evolution of public opinion over the last decade. We will investigate the use of large language models like ChatGPT and GPT-4 to speed up the labelling process, as these offer exceptional zero and few-shot performance (Gilardi, Alizadeh, and Kubli 2023; Wei et al. 2022; Zhang, Ding, and Jing 2023). The resulting dataset will be anonymised and published with a permissive licence to stimulate further research. We also intend to conduct a small study of the model's performance on tweets that were automatically translated into English. This translation is likely to affect the performance of the model, which is important if we are to include multilingual queries.

6. Conclusion

In this work, we have highlighted the need for a generic pipeline to process and analyse social media data related to natural disasters and environmental topics. Such a pipeline should preferably include multilingual support to achieve better global coverage. Our initial tests show that there are spatiotemporal correlations between the occurrence of wildfires and the corresponding tweets. However, our current methods are not detailed enough to perform a thorough analysis of the immediate and long-term effects of these wildfires on global tweet behaviour. Additionally, many of the collected tweets were not relevant to the queried topic but simply contained the same keywords or hashtags. Basic sentiment analysis models often fail to predict the correct sentiment and do not add much value when aggregated over large regions.

Stance detection models may be able to solve this issue, as our initial results showed good performance in determining the stance with respect to nuclear energy. We plan to expand our dataset, label a larger number of tweets, and fine-tune state-of-the-art NLP models to gain further insights into the impact of environmental topics on Twitter.

Notes

1. https://developers.google.com/maps/documentation/geocoding/overview
2. https://www.geonames.org/
3. https://nominatim.org/
4. https://simplemaps.com/data/world-cities
5. http://www.emdat.be/database

Acknowledgments

The research activities as described in this paper were funded by Ghent University and The Research Foundation - Flanders (FWO) (Grant number: G0F2820N).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the Fonds Wetenschappelijk Onderzoek [G0F2820N].

ORCID

Kenzo Milleville http://orcid.org/0000-0002-9765-6000
Samuel Van Ackere http://orcid.org/0000-0002-2462-7858
Jana Verdoodt http://orcid.org/0000-0003-2645-9904
Steven Verstockt http://orcid.org/0000-0003-1094-2184
Philippe De Maeyer http://orcid.org/0000-0001-8902-3855
Nico Van de Weghe http://orcid.org/0000-0002-5327-4000

References

Acar, Adam, and Yuya Muraki. 2011. "Twitter for Crisis Communication: Lessons Learned from Japan's Tsunami Disaster." *International Journal of Web Based Communities* 7 (3): 392–402. https://doi.org/10.1504/IJWBC.2011.041206.

Cambria, Erik, and Bebo White. 2014. "Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]." *IEEE Computational Intelligence Magazine* 9 (2): 48–57. https://doi.org/10.1109/MCI.2014.2307227.

de Albuquerque, João Porto, Melanie Eckle, Benjamin Herfort, and Alexander Zipf. 2016. "Crowdsourcing Geographic Information for Disaster Management and Improving Urban Resilience: An Overview of Recent Developments and Lessons Learned." In *European Handbook of Crowdsourced Geographic Information*, 309–321. https://doi.org/10.5334/bax.w.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. "Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding." *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Minneapolis, Minnesota, 1: 4171–4186. http://dx.doi.org/10.18653/v1/N19-1423.

Gilardi, Fabrizio, Meysam Alizadeh, and Maël Kubli. 2023. "ChatGpt Outperforms Crowd-Workers for Text-Annotation Tasks." *arXiv preprint arXiv:2303.15056*. https://doi.org/10.48550/arXiv.2303.15056.

Greg, Rybarczyk, Syagnik Banerjee, Melissa D. Starking-Szymanski, and Richard R. Shaker. 2018. "Travel and Us: The Impact of Mode Share on Sentiment Using Geo-Social Media and GIS." *Journal of Location Based Services* 12 (1): 40–62. https://doi.org/10.1080/17489725.2018.1468039.

Horita, Flávio, Livia Degrossi, Luiz Fernando Assis, Alexander Zipf, and Joao De Albuquerque. 2013. "The Use of Volunteered Geographic Information and Crowdsourcing in Disaster Management: A Systematic Literature Review." *In Proceedings of the Nineteenth Americas Conference on Information Systems*, August 15-17, Chicago, Illinois. 5: 06.

Kirilenko, Andrei P, Tatiana Molodtsova, and Svetlana O Stepchenkova. 2015. "People as Sensors: Mass Media and Local Temperature Influence Climate Change Discussion on Twitter." *Global Environmental Change* 30:92–100. https://doi.org/10.1016/j.gloenvcha.2014.11.003.

Kuemmerle, Tobias, Jed O Kaplan, Alexander V Prishchepov, Ilya Rylsky, Oleh Chaskovskyy, Vladimir S Tikunov, and Daniel Müller. 2015. "Forest Transitions in Eastern Europe and Their Effects on Carbon Budgets." *Global Change Biology* 21 (8): 3049–3061. https://doi.org/10.1111/gcb.12897.

Lai, Mirko, Alessandra Teresa Cignarella, Delia Irazú Hernández Farías, Cristina Bosco, Viviana Patti, and Paolo Rosso. 2020. "Multilingual Stance Detection in Social Media Political Debates." *Computer Speech & Language* 63:101075. https://doi.org/10.1016/j.csl.2020.101075.

Loria, Steven. 2018. "Textblob Documentation." *Release 0.15*.

Maharani, Warih. 2020. "Sentiment Analysis During Jakarta Flood for Emergency Responses and Situational Awareness in Disaster Management Using BERT." In *2020 8th International Conference on Information and Communication Technology (ICoICT)*, Yogyakarta, Indonesia, 1–5. IEEE.

Mohammad, Saif, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. "Semeval-2016 Task 6: Detecting Stance in Tweets." In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, June 16-17, San Diego, California, 31–41.

Muralidharan, Sidharth, Leslie Rasmussen, Daniel Patterson, and Jae-Hwa Shin. 2011. "Hope for Haiti: An Analysis of Facebook and Twitter Usage During the Earthquake Relief Efforts." *Public Relations Review* 37 (2): 175–177. https://doi.org/10.1016/j.pubrev.2011.01.010.

Neppalli, Venkata K, Cornelia Caragea, Anna Squicciarini, Andrea Tapia, and Sam Stehle. 2017. "Sentiment Analysis During Hurricane Sandy in Emergency Response." *International Journal of Disaster Risk Reduction* 21:213–222. https://doi.org/10.1016/j.ijdrr.2016.12.011.

Nguyen, Dat Quoc, Vu Thanh, and Anh Tuan Nguyen. 2020. "BERTweet: A Pre-Trained Language Model for English Tweets." *In Proceedings of the 2020 Conference on Empirical*

Methods in Natural Language Processing: System Demonstrations, Online, 9–14. Association for Computational Linguistics.

Reyes-Menendez, Ana, José Ramón Saura, and Cesar Alvarez-Alonso. 2018. "Understanding #worldenvironmentday User Opinions in Twitter: A Topic-Based Sentiment Analysis Approach." *International Journal of Environmental Research and Public Health* 15 (11): 2537. https://doi.org/10.3390/ijerph15112537.

Simon, Tomer, Avishay Goldberg, and Bruria Adini. 2015. "Socializing in Emergencies—A Review of the Use of Social Media in Emergency Situations." *International Journal of Information Management* 35 (5): 609–619. https://doi.org/10.1016/j.ijinfomgt.2015.07.001.

Sisco, Matthew R, Valentina Bosetti, and Elke U Weber. 2017. "When Do Extreme Weather Events Generate Attention to Climate Change?" *Climatic Change* 143 (1): 227–241. https://doi.org/10.1007/s10584-017-1984-2.

Tikunov, Vladimir, Yu Chereshnya, Marina Gribok, and Vasily Yablokov. 2017. "Assessment of Russian Regions in Terms of the Air Pollution Level." *Vestnik Moskovskogo Universiteta, Seriya 5: Geografiya* 5 (1): 43–48.

Tkachenko, Maxim, Mikhail Malyuk, Nikita Shevchenko, Andrey Holmanyuk, and Nikolai Liubimov. 2020. "Label Studio: Data Labeling Software." *Open source software* https://github.com/heartexlabs/label-studio.

Ukkusuri, Satish V, Xianyuan Zhan, Arif Mohaimin Sadri, and Ye. Qing. 2014. "Use of Social Media Data to Explore Crisis Informatics: Study of 2013 Oklahoma Tornado." *Transportation Research Record* 2459 (1): 110–118. https://doi.org/10.3141/2459-13.

Wei, Jason, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, et al. 2022. "Emergent Abilities of Large Language Models." *Transactions on Machine Learning Research*.

Yang, Duck-Hye, Lucy Mackey Bilaver, Oscar Hayes, and Robert Goerge. 2004. "Improving Geocoding Practices: Evaluation of Geocoding Tools." *Journal of Medical Systems* 28 (4): 361–370. https://doi.org/10.1023/B:JOMS.0000032851.76239.e3 .

Zhang, Bowen, Daijun Ding, and Liwen Jing. 2023. "How Would Stance Detection Techniques Evolve After the Launch of ChatGpt?" *arXiv preprint arXiv:2212.14548*. https://doi.org/10.48550/arXiv.2212.14548.

Zotova, Elena, Rodrigo Agerri, and German Rigau. 2021. "Semi-Automatic Generation of Multilingual Datasets for Stance Detection in Twitter." *Expert Systems with Applications* 170:114547. https://doi.org/10.1016/j.eswa.2020.114547.