Journal of Organizational and End User Computing
Volume 34 • Issue 3

Topic Modelling and Sentiment Analysis of Global Warming Tweets:

Evidence From Big Data Analysis

Fang Qiao, Xi'an International Studies University, China*
https://orcid.org/0000-0002-1859-4308

Jago Williams, Bangor University, UK

ABSTRACT

With the increasing extreme weather events and various disasters, people are paying more attention to environmental issues than ever, particularly global warming. Public debate on it has grown on various platforms, including newspapers and social media. This paper examines the topics and sentiments of the discussion of global warming on Twitter over a span of 18 months using two big data analytics techniques: topic modelling and sentiment analysis. There are seven main topics concerning global warming frequently debated on Twitter: factors causing global warming, consequences of global warming, actions necessary to stop global warming, relations between global warming and COVID-19, global warming's relation with politics, global warming as a hoax, and global warming as a reality. The sentiment analysis shows that most people express positive emotions about global warming, though the most evoked emotion found across the data is fear, followed by trust. The study provides a general and critical view of the public's principal concerns and their feelings about global warming on Twitter.

KEYWORDS

Big Data, Global Warming, Sentiment Analysis, Topic Modelling, Twitter

INTRODUCTION

Recent events such as the hottest recorded temperature at the Antarctic pole, Australian wildfires, and even the Covid 19 pandemic have often drawn people's attention to environmental issues, especially global warming. Voluminous research on global warming shows that extreme weather events and disasters, be they natural or anthropogenic, tend to spark discussion of global warming (Cody et al., 2015; Kirilenko et al., 2014; Molodtsova et al., 2013; Yeo et al., 2017). Acknowledging the magnitude of global warming, many studies have examined factors contributing to it (Gunnemyr, 2019), its impacts (Brown et al., 2011), its connection with other environmental problems (Le Duff et al., 2020), and attention to it in newspapers (Schmidt et al., 2013). With the arrival of digital era, social media such as Facebook, Instagram and Twitter have become essential platforms for information dissemination and public debate (Segerberg & Bennett, 2011). Twitter, in particular, is seen as the most popular platform to share breaking news, individual experience and personal opinions about current events (Hermida, 2013; Mustaqim et al., 2020; Zhao et al., 2011). Academic attention has

DOI: 10.4018/JOEUC.294901
*Corresponding Author

turned to Twitter data, including through analysing sentiment differences between the UK and Spain concerning global warming (Loureiro & Alló, 2020), media frames of global warming (Jang, 2013), comparing climate change and nuclear weapons (Allen & McAleer, 2018), and the assessment of disaster damage (Kryvasheyeu et al., 2016).

However, the present literature shows a lack of a comprehensive research on people's specific concerns, how they (emotionally) perceive global warming, and how their concerns and emotions are linked. To fill this gap, the current study aims to explore people's main loci of attention and emotions regarding global warming. This study analysed tweets containing the keywords "global warming" from January 2020 to July 2021 using topic modelling, specifically latent Dirichlet allocation (LDA) (Blei et al., 2003) and sentiment analysis based on Plutchik's eight basic human emotions. LDA offers statistical access to the latent topics across the unistructural data, which provides the information about what people are mainly debating. In order to know the attitude of the public towards global warming, sentiment analysis was performed to gain the polarity of people's general opinions and the specific emotions people convey via language.

The paper is structured into two parts: the first part focuses on the topics most popularly discussed during the period under study through LDA; the second part concerns the polarity of emotions that people expressed toward global warming.

METHODOLOGY

The current study examines the global warming discussion on Twitter from the perspective of topics and emotions by employing latent Dirichlet allocation (LDA), a topic modelling technique, and sentiment analysis.

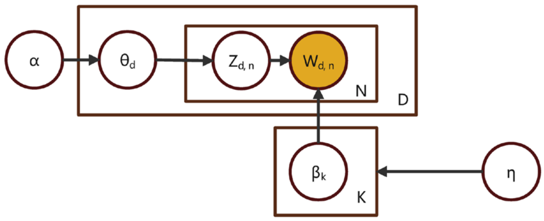
Latent Dirichlet Allocation (LDA) Model

Topic model, also referred as probabilistic topic model, is a statistical model for unearthing the latent semantic structures in a corpus, providing observable topics hidden in the corpus. In the age of information explosion, the information from all sources such as newspapers, web pages, books, and social media is beyond human processing capacity, and people find it prohibitively difficult to find the intended messages. Computational techniques such as topic modelling extract thematic information (topics) to help people find, organize and understand the substantial quantity of unstructured texts (Blei, 2012). LDA is the most popular topic modelling algorithm in the application of topic extraction from a collection of text bodies (Albalawi et al., 2020; Gerlach et al., 2018).

The basic and main assumptions of the LDA model are that each document contains diverse topics, and each topic has a probability distribution over words (Blei et al., 2003). One of the most prominent advantages of the LDA model is that topics can be extracted from collections of documents without any prior knowledge input (Albalawi et al., 2020). The model operates on the basis of following assumptions: there are k topics in the corpus D consisting of M documents, and each document is a sequence of N words w. Hyperparameters α and η denote Dirichlet priors for the document topic distribution θ and word distribution β, respectively. The specific generation process is as follows: (a) β_j is selected in each topic j; (b) θ_m is selected in each document m; (c) a topic z is selected from the distribution represented by θ_m in each word position n in document m; (d) a word is selected from distribution represented by β_z. The plate graph (Figure 1) illustrates how the LDA works specifically. The circles represent variables, and the rectangles represent iteration processes among documents, words and topics. The coloured circle represents the results of words denotating topics, which is the only visible variable in the corpus, the other variables are latent in this model.

The number of topics is s the most crucial parameter for the final results as k is pre-decided before the operation of LDA model. Usually, the number of topics depends on research questions and research aims (Boussalis & Coan, 2016; Quinn et al., 2010). If the k value is too high, the results could be incomplete in terms of information; while if the k value is too low, over-clustering could

Figure 1. The workflow of the LDA model.



happen (Greene et al., 2014). In order to gain a proper k value, LDA model was performed with 7 topics, 10 topics, and 15 topics to compare the quality of the results. Finally, it was decided that the results are satisfactory and suitable for the present study when $k = 7$. LDA was conducted using Python (Rehurek & Sojka, 2010) via the Gensim package.

Sentiment Analysis

Sentiment analysis, also known as opinion mining, is a natural language processing (NLP) technique that explores people's opinions, attitudes, and feelings towards a specific topic, and whether they are positive, neutral or negative (Asif et al., 2020; Stine, 2019). Beyond polarity, it also examines emotional states such as joy, anger, or fear. Sentiment analysis can be traced back to the 1950s (Puschmann & Powell, 2018) when written paper documents were the main research source. With the development of the Internet, information now has a variety of different sources, such as web pages, online news, blogs, comments, reviews and social media in particular. Recent decades have seen a rapid growth in the use of sentiment analysis. It has been applied in different areas to detect people's opinions, such as product reviews (Bhuskute, 2020), newspaper article reviews (Pandiaraj et al., 2021), opinions linked to tourism (Yan et al., 2020), wine reviews (Matheson et al., 2019) and gender studies (Thelwall, 2018). Social media, particularly Twitter, offers an observable and accessible sight into people's views and feelings towards current issues. Many studies using Twitter data to mine opinions to events such as the government's response to wildfires (Mustaqim et al., 2020), happiness level relative to geography (Mitchell et al., 2013), multimodal information (Kumar & Garg, 2019), and extremist tendencies (Asif et al., 2020).

The two most popular approaches to the process of sentiment analysis are sentiment analysis based on wordlists, which are weighted in the form of scores, and sentiment analysis based on machine learning (Stine, 2019). The present study employs the wordlist-based approach, specifically the NRC Word-Emotion Association Lexicon, to track the emotions of people and the proportion of the emotions distributed on Twitter in the discussion of global warming. The NRC Word-Emotion Association Lexicon (Mohammad & Turney, 2013, 2010), which is often used to extract emotions from texts, which is a list of 10,170 English words coded for Plutchik's eight basic human emotions (anger, fear, anticipation, surprise, trust, joy, sadness, disgust) and two polarities (positive and negative). It assigns scores to each word ranging from 0 to 1, representing the lowest and highest amount of emotion of a specific word, respectively. The percentage (p) of a specific emotion e is computed by following equation:

$$p_e = \frac{F_w}{F_p + F_n} \tag{1}$$

where F_w represents the frequency of words with a specific emotion such as joy in a tweet, F_p means the frequency of positive words in a tweet, and F_n is the frequency of negative words in a tweet.

Data Processing

(1) *Data collection:* The corpus consists of Twitter data collected by Python via the Twitter gardenhose Application Program Interface (API). All tweets retrieved contain the words "global warming" during a span from January 1, 2020 to July 30, 2021. After carefully excluding duplicated messages via a duplicate detection algorithm (Rajaraman & Ullman, 2011), the final data returned 538,478 tweets with nearly 22 million words.

(2) *Data preprocessing*: In order to improve validity and reliability of the data, this study followed suggestions from Maier et al. (2018), and performed a sequence of careful and strict preprocessing steps in the following order. The first step is tokenization, dividing documents into smaller units, usually word units. After tokenization, all capital letters are transformed into lowercase for the convenience of term unification and all punctuation marks were removed including period (.), comma (,), question mark (?), exclamation point (!) and special characters such as ampersand (&), slash (/), backslash (\), and the tilde (~), which are uninformative in text-mining based on bag-of-words (Kirelli & Arslankaya, 2020; Maier et al., 2018). Following that, the removal of stop words (e.g., articles, prepositions) is necessary since stop words bear no specific meaning thus have little contribution to the document content (Mustaqim et al., 2020). The next step is unification including lemmatization and stemming. Lemmatization is performed in preference to stemming because lemmatization is much more informative than stemming while stemming is believed to be less precise and more difficult to interpret (Schütze et al., 2008). Lemmatization is a process of transforming inflected forms of words to the lemma, such that "studies" and "studying" become "study". It examines the surrounding context of a word to identify the part of speech of a given word. Stemming is the process of generating word stem, base or root form by cutting derivational and inflectional suffixes, such as "studies" become "studi", and "studying" become "study". Thus, the stem may not be an actual word can be looked up in dictionary. Unlike lemmatization, stemming operates on a single word without consideration of the context of a word. The last procedure is relative pruning, deleting extremely infrequent and extremely frequent words in a corpus to improve the algorithm's performance.

RESULTS AND DISCUSSION

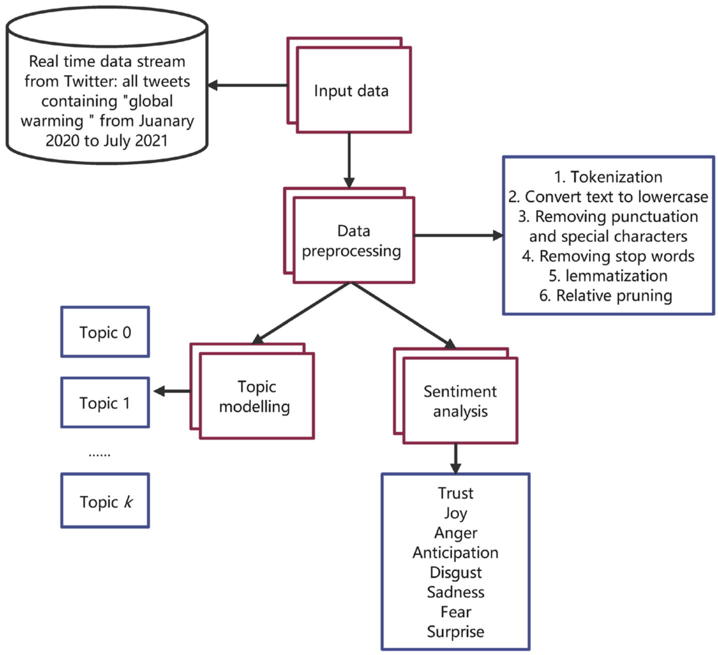
The following part begins with the results of topic modelling by LDA, specifically, the categories of topics, overall changes of topics over time and how the topics are represented in language. Then the results of sentiment analysis are presented in terms of eight different emotions distributed across Twitter data concerning global warming.

Topic Analysis

The LDA model computed the seven most manageable and significant clusters of topics. They are listed in descending order from the highest proportion to the lowest: Factors causing global warming; Impact of global warming; Actions to stop global warming; Relation between global warming and Covid-19; Close relation between Global warming and politics; Global warming as a hoax and Global warming as a reality. Figure 3 shows the proportions of each topic and their developments over time and Figure 4 shows the key topics with highly weighted keywords surrounding.

(1) Factors causing global warming

Figure 2. Research procedure



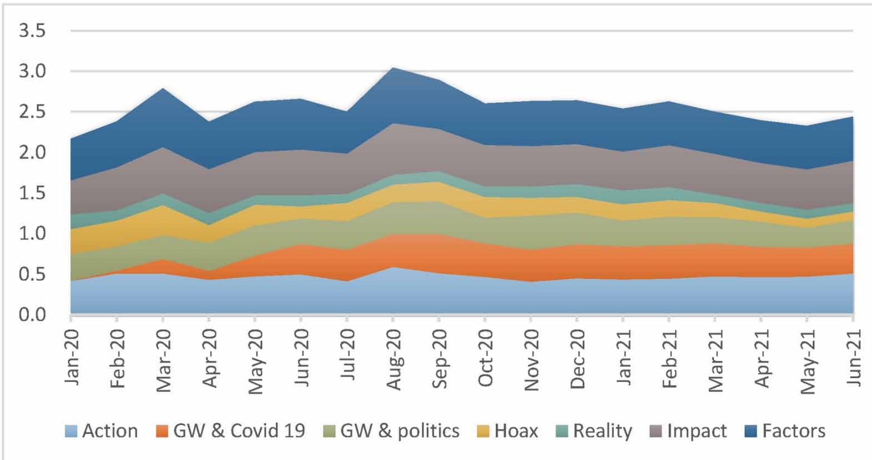
The most frequently debated topic with the highest proportion is the factors causing global warming, as its name shows, focusing on some specific factors contributing to it. The highly frequent words around this topic are “emission”, “carbon”, “pollution”, “human”, “anthropogenic”, “energy”, “environment”, “CO2”, “gas”, “greenhouse”, “fossil”, and “plastic”. The data show that the topic reaches the peaks when disasters or abnormal weather events happen, such as Hurricane Laura and California wildfires. As previous studies have pointed out, abnormal weather and disasters usually intensify the discussion over global warming on Twitter (Hamilton & Stampone, 2013; Molodtsova et al., 2013; Zaval et al., 2014). An example is listed here: “it’s a feedback loop: as peatlands release more carbon, global warming increases, which thaws more peat and causes more wildfires.”.

Generally speaking, the data shows that most people believe it is anthropogenic factors that cause global warming, which is in line with the results of the research by Leiserowitz et al. (2020). The high occurrence of the word “emission” shows that carbon emissions and the greenhouse effect are considered by Twitter users as the most crucial factor leading to global warming. Here are some examples showing this topic: “For 40 years scientists understood human-caused increase concentrations of greenhouse gases are driving global warming but efforts to curtail emissions have not been successful. It’s now 2020, and our inaction is leading many to consider geoengineering.”; “Ocean 2020 Definition: A large body of water filled with oil, trash, micro plastics, acidification, dying ecosystems, rising temperatures and human ignorance. Global warming.”.

(2) Impact of global warming

Just following the topic of Factors, the topic of Impact of global warming comes second in the proportional size of topics. It centres on the consequences, specifically disasters, of global warming.

Figure 3. Proportions of topics



Global warming is considered as a threat, a dangerous phenomenon. The keywords around this topic are "impact", "consequence", "polar", "disaster", "effect", "rise", "threat", "extreme", "weather", "alarm", "sea", and "glacier".

The topic often involves natural disasters and extreme weather events, such as wildfires, hurricane, snow storms, and floods. In other words, most people think those disasters and extreme weather are caused by global warming which results from human activity. Here are some typical examples: "Terrifying consequence of global warming. My heart is with Australia and Australians and their beautiful country and wildlife."; "Global warming pushes temps to near-record levels in 2020, in effect trying 2016 as the hottest year on record, according to data released by US science agencies."; "We all say 'That's 2020' about stuff like multiple consecutive hurricanes and wildfires. It just hit me. It's not '2020'. This is what life is from now on until we address global warming.".

(3) Actions to stop global warming

With so many adverse consequences brought by global warming, a large number of Twitter users advocate for actions to be undertaken immediately, contributing to the third topic—Actions to stop global warming. The most frequently co-occurring words are "action", "act", "fight", "stop", "policy", "scientist", "agreement", "solve", "sustain", and "renew". It is not surprising to find advocation of actions to stop global warming knowing its damages and the factors causing it. The actions are grouped into the following: reduction of toxic emissions, policy-making related to the environment, using products from sustainable sources, zero pollution, etc.

Some examples of this topic: "Peeps are saying '2020 is horrible' as though they expect next year will be better. But will it? Rolling disasters have been predicted by science as the consequence of global warming. Governments must ACT NOW. Elect governments that will take action."; "In a bid to promote eco-friendly industries in the state, the Gujarat Industrial Policy 2020 incentivizes setting-up of green ventures and adoption of clean and green technologies. This new policy will help fight global warming and will help making Gujarat cleaner and greener."; "Set in summer 2020, Valio's emission reduction targets to stop global warming at 1.5 degrees have been certified by the Science-based Target initiative."

(4) Global warming and Covid-19

The fourth topic extracted from data is the connection between global warming and Covid-19. The two seemingly unconnected issues are unexpectedly connected with each other. The analysis begins with the keywords of the topic, which are as follows: "Covid-19", "virus", "hoax", "people", "emission", "coronavirus", "Trump", "kill", "worry", "pandemic", "carbon", "news", "lockdown", "life", "emergency", and "science".

Examining the tweets more closely, two general trends were found. The first focuses on carbon emission reduction influenced by Covid-19 during the lockdown. As one Twitter user put it: "Although the Covid-19 pandemic will cause a dip in 2020 emission, this will not bring the words closer to the Paris Agreement goal of limiting global warming this century to well below 2°C and pursuing 1.5°C.". This phenomenon also attracted academic attention to the relations between toxic emissions and the pandemic, and the findings showed a significant decrease in emissions in most countries in the world during the lockdown period (Evangeliou et al., 2021; Sarfraz et al., 2021).

The second draws parallels between global warming and Covid-19, highlighting the importance and severity of problems humans are confronting, such as "2021 will not be a new year because global warming, the extinction crisis, the Covid-19 global pandemic, along with unemployment, hunger and poverty, will be part of that year as they are in 2020.". Interestingly, it also usually involves blame or sarcasm about Trump, for example, "Funny how Trump dismisses scientist's warning regarding global warming and Covid 19, but he is desperately hoping they will produce a vaccine before Nov 3."; "Trump doesn't believe in science. He said Covid-19 was a 'Democratic hoax'. He also claims global warming is a Democrat hoax.". From the data, it is inferred that a sizeable number of Twitter users were not satisfied with statements by Trump that Covid-19 and global warming are both unimportant.

(5) Close relations between global warming and politics

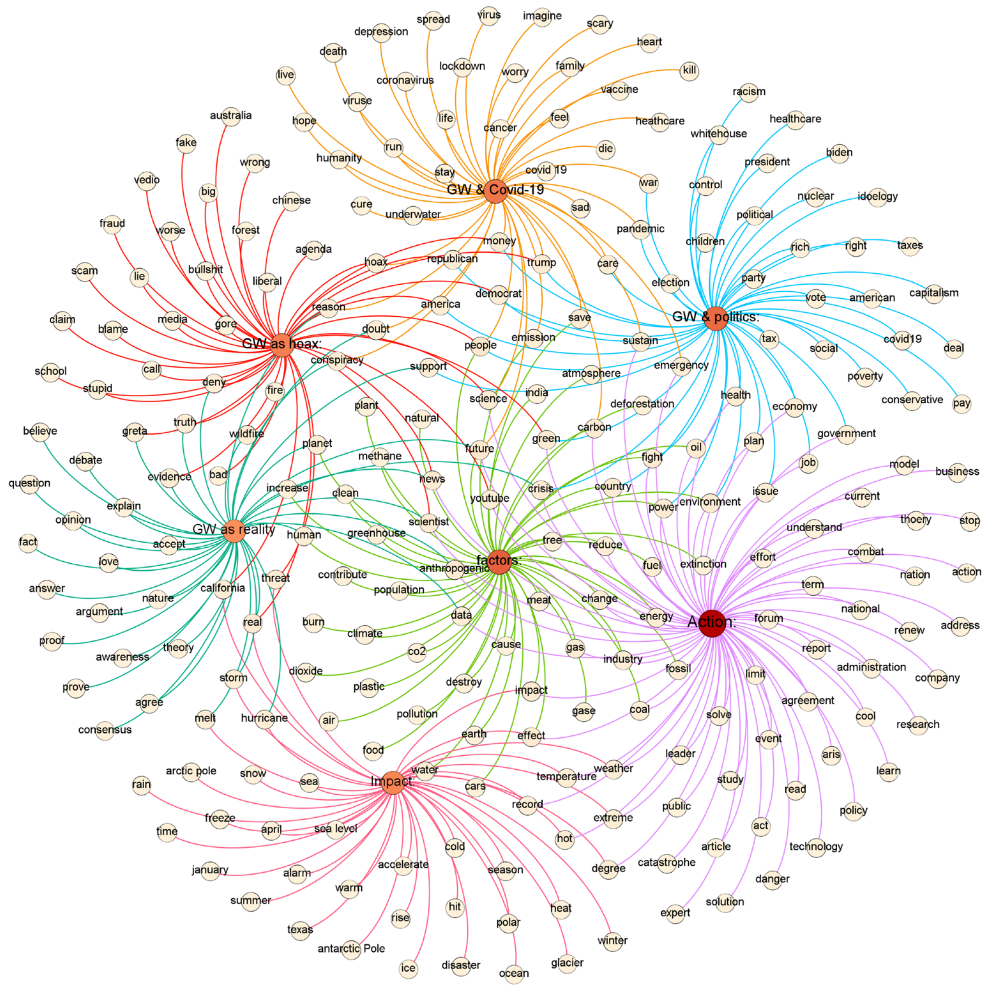
The results show that global warming is closely connected with politics, especially with American parties. The keywords surrounding the topic are "politic", "Trump", "election", "ideology", "people", "government", "Biden", "vote", "power", "party", and "Democrat". It is discovered that global warming was connected in the corpus with the 2020 American presidential election. Here are some illustrative examples: "Breaking: Global warming is the top most important issue among liberal Democrats deciding who they'll vote for in the 2020 Presidential Election, followed by healthcare, income gap and environment protection."; "2020 ELECTION: Global warming surges as a voting issue! It is now the top 1 voting issue (out of 29) among liberal Democrats and top 5 among moderate conservative Democrats.".

Global warming has been politicalized since the late 1980s when it was listed in the US national agenda. McCright and Dunlap (2011) investigated the American public's opinions towards global warming over ten years and discovered that the global warming issue was significantly dominant in political ideologies and partisan polarization. Our data also show the same results, as this tweet demonstrates: "So farewell global warming, you have served your political purpose. Hallo extreme weather."

(6) Global warming as a hoax

Contrary to the "common sense" that climate scientists believe, a number of people remain sceptical about the truth of global warming. A study found that almost 1/3 people in America were sceptical about global warming and denying that global warming is caused by human activity (Leiserowitz et al., 2015). Similarly, our results show that many Twitter users believe that global warming is just a hoax, denying global warming. The high frequency weighted words surrounding the

Figure 4. Key topics of global warming discussion on Twitter



topic in decreasing order are “hoax”, “news”, “fake”, “Trump”, “blame”, “lie”, “people”, “media”, “scam”, “deny”.

These highly weighted words show following important information: First, many Twitter users are sceptical about the gradual increase in temperature of the Earth, considering it “fake news”, “scam”, and “lie”, such as in: “Man-made global warming is the biggest and most expensive lie in the history of the world.”; “Climate change is a natural phenomenon. Human cause global warming is a Hoax.” Second, the news about global warming is “framed” by media, disseminating among people, for example, “global warming is a fake news.”; “Ordinary people are denied the right to be heard as the global warming scam is embraced once more by gullible Western leaders. The world media is a dishonest supporter of the hoax by failing to be objective and reporting only one side.”; Third, the discourse about the “hoax” theory is closely related to the former American president Donald Trump, such as in: “I remember when Trump claimed that global warming was a hoax created by China.”; “I thought Trump said global warming is a hoax.”

Donald Trump, who has long been a denier of global warming, tweeted some messages in 2012 that global warming is a lie. The tweet "The concept of global warming was created by and for Chinese in order to make US manufacturing non-competitive," was retweeted, according to reports by The New York Times, over 104,000 times, and "liked" nearly 66,000 times (Wong, 2016). It is suggested that his tweets about global warming had a great impact on people's view of it, since social media, especially Twitter, has been an important place for political communication, influencing a large number of people's opinions (Buccoliero et al., 2020; Hong & Nadler, 2011).

(7) Global warming as a reality

The last topic with the least proportion across the whole data set is that global warming is a reality, with highly frequent co-occurrence of words "real", "true", "evidence", "believe", "anthropogenic", "fact", "news", "Greta", "crisis", "wildfire", and "melt" (Figure 2). As overwhelming scientific evidence shows, the average temperature of the Earth is getting higher, and this is primarily caused by human activity. Consistent with the findings of Leiserowitz et al. (2020) that almost half Americans strongly believe that global warming is happening, our results show the topic of believing the truth of global warming.

Looking more closely the tweets, it is evident that this topic has two prominent features, i.e., persuading people into believing the fact of global warming (related to topic 6) and listing natural disasters and extreme weather events such as hurricanes, wildfires, melting of glaciers, and high temperatures at the poles.

Regarding the first feature, many Twitter users talk about the truth of global warming, yet at the same time mention those who don't believe in it. This is probably due to the prevalence of the "hoax" frame, i.e., some believe global warming is a hoax. Some examples are listed here: "The fact that people still don't believe in global warming even though 2020 have been a year that have shown us a lot of consequences is crazy. Keep this up and this is our future, our every year. Nature disasters, new disease… Global warming is real."; "2-4 inches of rain expected in the next two days. Just in case anyone doesn't believe global warming is real. January 8th 2020 that should be snow."; "It's a disgrace that 'views on global warming or climate change' is even a thing adults will talk about in 2020. Nobody has 'views on whether dogs are real' or 'views on days being longer in summer than winter'. These are facts. Climate change is a fact.".

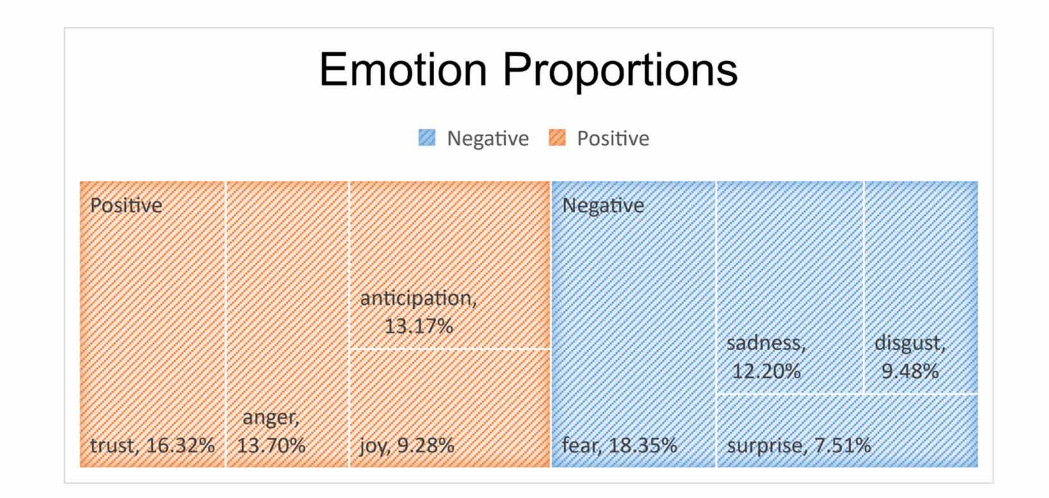
As for tweeting a list of natural disasters and extreme weather events, the data shows that this seemingly serves as evidence proving the fact of global warming. Some examples of this: "Global warming is real because January 2020 was the hottest in 141 years."; "On August 8, 2020, the Canadian ice shelf larger than Manhattan collapses into the sea. Global warming is real and it's a crisis."; "I'm saying this for a reason, flooding, tornadoes, snow storms, hail, wind, etc. Resources need to be available no matter what. We are in 2021 and earth has already proved, with no avail, that global warming is real and upon us. Mother nature is mad and we're the issue behind it.".

According to the results, what people care most in the collective sense is the factors, then the impacts and actions need to be done when it comes to global warming. This is typical blame behaviour, blaming someone or something else first when people are in uncomfortable situations and defending themselves from cognitive dissonance (Hein, 1998). The seven topics reveal the main attentions as well as behaviours of Twitter users in the debate of global warming.

Sentiment Analysis

The sentiment analysis was performed using Python based on the NRC Word-Emotion Association Lexicon, which revealed the distribution and proportion of the eight emotions by Plutchik across the data. Plutchik proposed the wheel of emotions and divided emotions into eight categories, viz., joy, trust, fear, surprise, sadness, disgust, anger and anticipation (Plutchik, 1980, 1982). This is one of the most influential emotion theories. He suggested a bipolarity in the eight emotions: joy is the opposite of sadness, fear is the opposite of anger, anticipation is the opposite of surprise and disgust is the opposite of trust. Joy, trust, anticipation and anger are considered as positive in valence, while fear, sadness, disgust and surprise denote polarity of negativeness. In opposition to common understanding, the negative valence of anger is classified as a positive emotion and surprise as negative. This is because anger is considered a sign of strength, motivating action (Hess, 2014) and indicates a path of "moving toward" a goal (TenHouten, 2014) . As for surprise, it involves a violation of people's psychological territory (TenHouten, 2006), and usually brings about unpleasantness (Noordewier & Breugelmans, 2013).

Figure 5. Emotion proportions in relation to global warming



Very differently to the common understanding of the public perception of global warming as a negative issue, the sentiment analysis results show that the proportion of positive views of global warming is, surprisingly, higher than the negative one (Figure 5). The results confirmed the Pollyanna hypothesis (Boucher & Osgood, 1969), that people tend to think about the positive side of events. Similarly, the study conducted by Loureiro and Alló (2020) also showed similar results when they compared the Twitter messages containing to keywords "climate change" in the UK and Spain, showing that the overall polarity was positive in the UK, while the opposite was true in Spain. Perhaps it is attributable to the high proportion of trust (16.32%) and anger (13.7%) which is usually viewed as a negative feeling.

The emotion most evoked by global warming is fear, with a proportion of 18.35%, contributing most to the negative valence, which is followed by the emotion of trust with 16.32%, contributing most to the positivity in valence. Anger and anticipation occupy comparatively big portions with 13.7% and 13.17% respectively, which is followed by sadness with 12.20%. The feeling that global warming evokes least is surprise (7.51%). Disgust (9.48%) and joy (9.28%) account for about 18% in positive and negative valence, respectively.

Wordcloud is used for the display of most frequent words assigned to different emotions (Figure 6). The bigger size of the word is, the more frequent the word appears in terms of a specific emotion, and the more contributions to the emotion classification it makes.

The results show that fear is the most evoked emotion, just as Hulme (2008) points out, "we are living in a climate of fear about our future climate". The most frequent words serving the emotion are “change”, “pandemic”, “government”, “kill”, “fight”, “threat”, and “bad”. It is inferable that the changes in climate and the pandemic are both considered “threats”, causing the fear emotion. According to Plutchik and Kellerman (2013), the stimulus “threat” causes a cognition of danger, resulting a feeling of fear. The corresponding behaviour is to run away from it. The results indicate that humans are still not prepared for the fact or consequences of global warming, being greatly confused or even frightened by all those extreme weather events. The highly frequent words such as “scientist”, “level”, “economy”, “money”, “tree”, “president”, “policy”, “deal”, “expert” lead to the high proportion of the emotion of trust. The emotion of trust means a mental acceptance, a willingness to have a close relationship with (something). It suggests that people such as scientists, presidents, and experts are considered “friends” as the Plutchik’s emotion theory argues and the “policy” and “deal” made by them are also convincing. In this regard, 16.3% people showed confidence in the global warming issue, still holding a very positive view of it. The stimulus of obstacles is considered as an enemy cognitively, from which the anger emotion arises. The words “hoax”, “threat”, “bad”, “hot”, “fight”, “blame”, “disaster”, and “storm” contribute most to the feeling of anger. The results show that the belief that global warming is a hoax causes a strong anger across Twitter, as do all disasters. As for anticipation, it involves the stimulus of new territory, resulting in a corresponding behaviour of examination. The words “scientist”, “time”, “tree”, “money”, “plan”, “deal”, and “hope”. are crucial to the emotion of anticipation. The sad feeling deriving from the loss of valued people or things focuses on the consequences of different disasters, with the keywords “hoax”, “pandemic”, “bad”, “die”, “kill”, “wildfire”, and “disease”. The belief that global warming is a hoax, disasters, and anthropogenic pollution generate the feeling of disgust with keywords “hoax”, “blame”, “tree”, “bad”, “lie”, “shit”, “death” and so forth. A small proportion of joy elicited by global warming is affiliated with the words “love”, “tree”, “money”, “save”, “green”, “hope”, “humanity” along with others. The least roused feeling is surprise, which usually produces interrelated behaviour of alertness and halting. The keywords are “Trump”, “hoax”, “money”, “disaster”, “death”, “free”, and “hope”. It is interesting to note that the word “hoax” appeared very often in the wordcloud with a high weight in the emotion of sadness, disgust, anger and surprise. In other words, the belief that global warming is a hoax tends to provoke a negative feeling.

In order to further examine the attentions and emotions of the public in relevance to global warming, the connections between emotions and topics were analysed through Groupby sum.

The results (Figure 7) show that the fear emotion takes a leading position in all topics, followed by trust. It is noted that people show more trust than other emotions when talking about the reality of global warming and actions needed to be done, and show the least trust for the topic of Hoax. Anticipation appears least in the topic Hoax, and most in the discussion of impact of global warming. Sadness is displayed most in topic of the relations between global warming and Covid-19. The topic Reality shows more joy emotion than other topics while the topic Hoax exhibits least joy. It is interesting to note that the topic Hoax show a very prominent negativity in emotion with high scale of fear, least trust, joy and anticipation, as well as most disgust and surprise.

The topics exhibits the general attention of the public in relevance to global warming on Twitter and the sentiment analysis show the mental state of people towards it. The comprehensive research of global warming from the perspective topic and emotion shed light on people’s attitudes towards it.

Figure 6. Wordcloud of each emotion

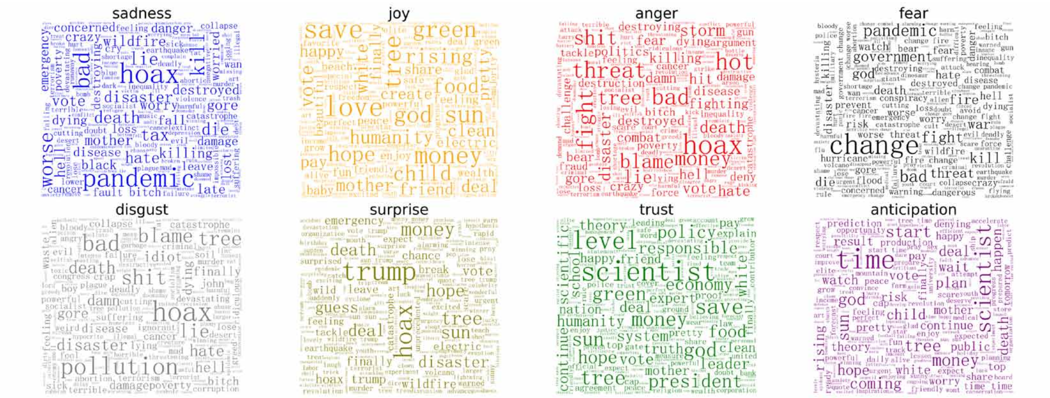
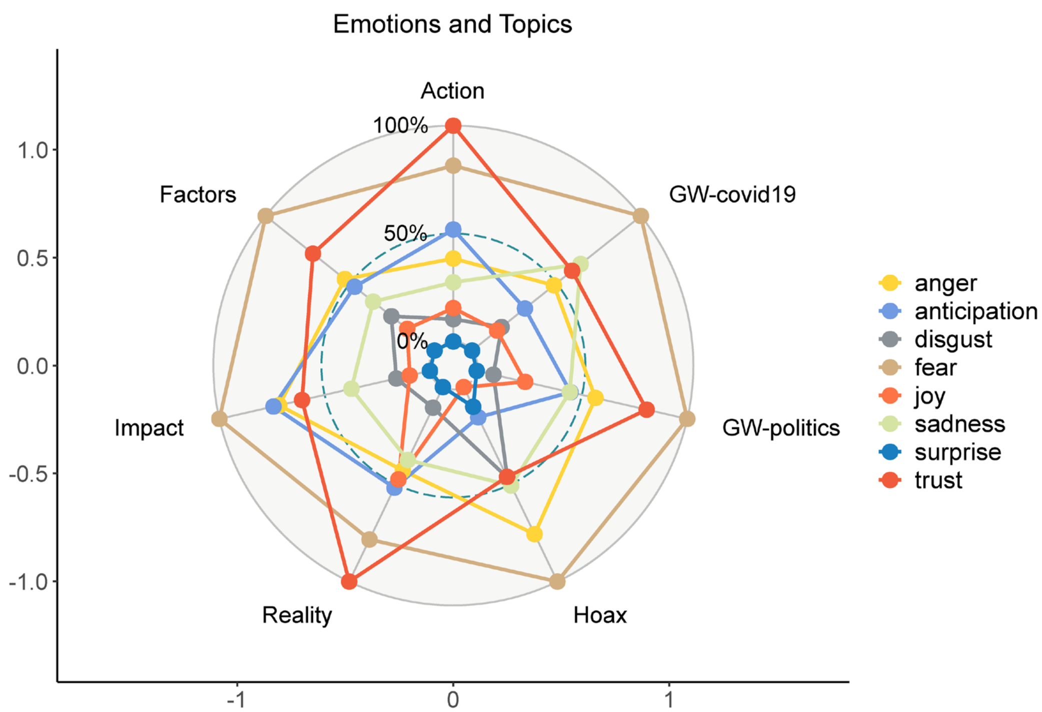


Figure 7. Connections between emotions and topics



CONCLUSION

The present study, which is based on big data analytics, extracted 538,478 tweets about global warming, spanning 18 months, analysed people's attention and emotion regarding global warming through LDA, a topic modelling method, and sentiment analysis on the basis of Plutchik's emotion theory.

Seven significant clusters of topics emerged from the data using the LDA. They were as follows: Factors causing global warming, the most frequently occurring set in the data, in which anthropogenic causes of global warming, especially greenhouse emissions, are very frequently mentioned; Impact of global warming, which contains references to natural disasters caused by global warming; Actions to stop global warming, which covers (largely very urgent) calls for actions to counteract global warming; Relation between global warming and Covid-19, which contains two main trends – the reduction in carbon emissions owing to the lockdown and highlighting the parallels between Covid-19 and global warming as crises; Close relation between Global warming and politics, in which there were a number of references to American political parties; Global warming as a hoax, or expressions of scepticism and denials of the truth of global warming, in which anger was a prominently expressed emotion and Global warming as a reality, in which lists of natural disasters and other extreme environmental events are frequently mentioned, along with attempts to persuade others of the truth.

The sentiment analysis showed the positive discussion of global warming is more prevalent than negative discussion, which provides empirical evidence for the Pollyanna phenomenon whereby people use more positive words to describe an event if is a disaster, showing a bright side of things (Boucher & Osgood, 1969). The most evoked emotion in the discussion over global warming on Twitter is fear, followed by trust. And the least roused emotion by the issue is surprise, followed by joy. It is interpreted that most Twitter users considered global warming as a threat, which gives rise to a sense of danger. The fear produced in them probably is closely related with what had already happened, such as the severe consequences (disasters) caused by it or unpredictable situations. Still, a number of people showed trust in regard to global warming, accepting the current situation. Regarding the connections between topics and emotions, fear dominates most of topics while the topic of action and reality show more trust than others.

Global warming, as an issue affecting all human beings, has been extensively debated for years through different platforms such as newspapers, Facebook, Twitter. Understanding the public's perception of global warming is of importance to economic development, policy making, lifestyle decisions, etc. The present study provides an academic information for the research of emotions evoked by the discourse on global warming.

ACKNOWLEDGMENT

This research was supported by China Scholarship Council [grant number 201808610242], Xi'an International Studies University PhD Scholarship [grant number syjsb201704], and Key Research Project from Xi'an International Studies University [grant number 17XWZD04].

REFERENCES

Albalawi, R., Yeap, T. H., & Benyoucef, M. (2020). Using topic modeling methods for short-text data: A comparative analysis. *Frontiers in Artificial Intelligence*, *3*(42), 1–14. doi:10.3389/frai.2020.00042 PMID:33733159

Allen, D. E., & McAleer, M. (2018). President Trump tweets supreme leader Kim Jong-Un on nuclear weapons: A comparison with climate change. *Sustainability*, *10*(7), 1–6. doi:10.3390/su10072310

Asif, M., Ishtiaq, A., Ahmad, H., Aljuaid, H., & Shah, J. (2020). Sentiment analysis of extremism in social media from textual information. *Telematics and Informatics*, *48*, 1–20. doi:10.1016/j.tele.2020.101345

Bhuskute, L. (2020). Sentiment analysis of product reviews. *International Journal of Advanced Science and Technology*, *29*(4), 7671–7677.

Blei, D. M. (2012). Introduction to probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. doi:10.1145/2133806.2133826

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Boucher, J., & Osgood, C. E. (1969). The Pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behavior*, *8*(1), 1–8. doi:10.1016/S0022-5371(69)80002-2

Boussalis, C., & Coan, T. G. (2016). Text-mining the signals of climate change doubt. *Global Environmental Change*, *36*, 89–100. doi:10.1016/j.gloenvcha.2015.12.001

Brown, T., Budd, L., Bell, M., & Rendell, H. (2011). The local impact of global climate change: Reporting on landscape transformation and threatened identity in the English regional newspaper press. *Public Understanding of Science (Bristol, England)*, *20*(5), 658–673. doi:10.1177/0963662510361416 PMID:22164705

Buccoliero, L., Bellio, E., Crestini, G., & Arkoudas, A. (2020). Twitter and politics: Evidence from the US presidential elections 2016. *Journal of Marketing Communications*, *26*(1), 88–114. doi:10.1080/13527266.2018.1504228

Cody, E. M., Reagan, A. J., Mitchell, L., Dodds, P. S., & Danforth, C. M. (2015). Climate change sentiment on Twitter: An unsolicited public opinion poll. *PLoS One*, *10*(8), 1–18. doi:10.1371/journal.pone.0136092 PMID:26291877

Evangeliou, N., Platt, S. M., Eckhardt, S., Lund Myhre, C., Laj, P., Alados-Arboledas, L., Backman, J., Brem, B. T., Fiebig, M., Flentje, H., Marinoni, A., Pandolfi, M., Yus-Dìez, J., Prats, N., Putaud, J. P., Sellegri, K., Sorribas, M., Eleftheriadis, K., Vratolis, S., & Stohl, A. et al. (2021). Changes in black carbon emissions over Europe due to COVID-19 lockdowns. *Atmospheric Chemistry and Physics*, *21*(4), 2675–2692. doi:10.5194/acp-21-2675-2021

Gerlach, M., Peixoto, T. P., & Altmann, E. G. (2018). A network approach to topic models. *Science Advances*, *4*(7), 1–11. doi:10.1126/sciadv.aaq1360 PMID:30035215

Greene, D., O'Callaghan, D., & Cunningham, P. (2014). *How many topics? Stability analysis for topic models*. 10.1007/978-3-662-44848-9_32

Gunnemyr, M. (2019). Causing global warming. *Ethical Theory and Moral Practice*, *22*(2), 399–424. doi:10.1007/s10677-019-09990-w

Hamilton, L. C., & Stampone, M. D. (2013). Blowin' in the wind: Short-term weather and belief in anthropogenic climate change. *Weather, Climate, and Society*, *5*(2), 112–119. doi:10.1175/WCAS-D-12-00048.1

Hein, F. (1998). The blame game. *IEEE Software*, *15*(6), 89–91. doi:10.1109/52.730852

Hermida, A. (2013). Journalism: Reconfiguring journalism research about twitter, one tweet at a time. *Digital Journalism*, *1*(3), 295–313. doi:10.1080/21670811.2013.808456

Hess, U. (2014). Anger is a positive emotion. In *The positive side of negative emotions* (pp. 55–75). The Guilford Press.

Hong, S., & Nadler, D. (2011). Does the early bird move the polls? The use of the social media tool ‘Twitter’ by U.S. politicians and its impact on public opinion. *The Proceedings of the 12th Annual International Conference on Digital Government Research*, 182–186. doi:10.1145/2037556.2037583

Hulme, M. (2008). The conquering of climate: Discourses of fear and their dissolution. *The Geographical Journal*, *174*(1), 5–16. doi:10.1111/j.1475-4959.2008.00266.x

Jang, S. M. (2013). Framing responsibility in climate change discourse: Ethnocentric attribution bias, perceived causes, and policy attitudes. *Journal of Environmental Psychology*, *36*, 27–36. doi:10.1016/j.jenvp.2013.07.003

Kirelli, Y., & Arslankaya, S. (2020). Sentiment analysis of shared tweets on global warming on Twitter with data mining methods: A case study on Turkish language. *Computational Intelligence and Neuroscience*, *2020*, 1–9. doi:10.1155/2020/1904172 PMID:32963511

Kirilenko, A. P., Molodtsova, T., & Stepchenkova, S. O. (2014). People as sensors: Mass media and local temperature influence climate change discussion on Twitter. *Global Environmental Change*, *30*, 92–100. doi:10.1016/j.gloenvcha.2014.11.003

Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Hentenryck, P. V., Fowler, J., & Cebrian, M. (2016). Rapid assessment of disaster damage using social media activity. *Science Advances*, *2*(3), 1–11. doi:10.1126/sciadv.1500779 PMID:27034978

Kumar, A., & Garg, G. (2019). Sentiment analysis of multimodal twitter data. *Multimedia Tools and Applications*, *78*(17), 24103–24119. doi:10.1007/s11042-019-7390-1

Le Duff, M., Dumas, P., Allenbach, M., & Cohen, O. (2020). An orientation for coastal disaster risks management and prevention policy in a global warming context: Case study in Ouvea (New Caledonia). *Marine Policy*, *117*, 1–13. doi:10.1016/j.marpol.2018.12.012

Leiserowitz, A., Maibach, E., Rosenthal, S., Kotcher, J., Bergquist, P., Ballew, M. T., Goldberg, M., & Gustafson, A. (2020). Climate change in the American mind: November 2019. PsyArXiv. 10.31234/osf.io/z3wtx

Leiserowitz, A., Maibach, E., Roser-Renouf, C., Feinberg, G., & Rosenthal, S. (2015). Climate change in the American mind. Yale University and George Mason University.

Loureiro, M. L., & Alló, M. (2020). Sensing climate change and energy issues: Sentiment and emotion analysis with social media in the U.K. and Spain. *Energy Policy*, *143*, 1–11. doi:10.1016/j.enpol.2020.111490

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, *12*(2–3), 93–118. doi:10.1080/19312458.2018.1430754

Matheson, K., Plangger, K., Kietzmann, J., Vella, J., & Grant, P. (2019). The serious side to funny cartoons: Understanding public perception of wine through cartoon content analysis. *Journal of Wine Research*, *30*(2), 95–106. doi:10.1080/09571264.2019.1587396

McCright, A. M., & Dunlap, R. E. (2011). The politicization of climate change and polarization in the American public’s views of global warming, 2001–2010. *The Sociological Quarterly*, *52*(2), 155–194. doi:10.1111/j.1533-8525.2011.01198.x

Mitchell, L., Frank, M. R., Harris, K. D., Dodds, P. S., & Danforth, C. M. (2013). The geography of happiness: Connecting Twitter sentiment and expression, demographics, and objective characteristics of place. *PLoS One*, *8*(5), 1–15. doi:10.1371/journal.pone.0064417 PMID:23734200

Mohammad, S. M., & Turney, P. D. (2010). Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. *Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, 26–34.

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, *29*(3), 436–465. doi:10.1111/j.1467-8640.2012.00460.x

Molodtsova, T., Kirilenko, A., & Stepchenkova, S. (2013). Utilizing the social media data to validate’climate change’indices. *AGU Fall Meeting Abstracts,* GC11D-1039.

Mustaqim, T., Umam, K., & Muslim, M. A. (2020). Twitter text mining for sentiment analysis on government's response to forest fires with vader lexicon polarity detection and k-nearest neighbor algorithm. *Journal of Physics: Conference Series*, *1567*(3), 1–8. doi:10.1088/1742-6596/1567/3/032024

Noordewier, M. K., & Breugelmans, S. M. (2013). On the valence of surprise. *Cognition and Emotion*, *27*(7), 1326–1334. doi:10.1080/02699931.2013.777660 PMID:23560688

Pandiaraj, A., Sundar, C., & Pavalarajan, S. (2021). Sentiment analysis on newspaper article reviews: Contribution towards improved rider optimization-based hybrid classifier. *Kybernetes*, 1–38. doi:10.1108/K-08-2020-0512

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In *Theories of emotion* (pp. 3–33). Elsevier. doi:10.1016/B978-0-12-558701-3.50007-7

Plutchik, R. (1982). A psychoevolutionary theory of emotions. *Social Sciences Information. Information Sur les Sciences Sociales*, *21*(4–5), 529–553. doi:10.1177/053901882021004003

Plutchik, R., & Kellerman, H. (2013). *Theories of emotion* (Vol. 1). Academic Press.

Puschmann, C., & Powell, A. (2018). Turning words into consumer preferences: How sentiment Analysis is framed in research and the news media. *Social Media + Society*, *4*(3), 1–13. doi:10.1177/2056305118797724

Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., Radev, D. R., Purpura, S., Schrodt, P., Sin, G., Sinclair, B., Ward, M., Wilkerson, J., Wood, D., & Zorn, C. (2010). How to analyze political attention with minimal assumptions and costs. *American Journal 21 Political Science*, *54*(1), 209–228. doi:10.1111/j.1540-5907.2009.00427.x

Rajaraman, A., & Ullman, J. D. (2011). *Mining of massive datasets*. Cambridge University Press. doi:10.1017/CBO9781139058452

Rehurek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. doi:<ALIGNMENT.qj></ALIGNMENT>10.1.1.695.4595

Sarfraz, M., Mohsin, M., Naseem, S., & Kumar, A. (2021). Modeling the relationship between carbon emissions and environmental sustainability during COVID-19: A new evidence from asymmetric ARDL cointegration approach. *Environment, Development and Sustainability*, *23*(11), 16208–16226. doi:10.1007/s10668-021-01324-0 PMID:33782633

Schmidt, A., Ivanova, A., & Schäfer, M. S. (2013). Media attention for climate change around the world: A comparative analysis of newspaper coverage in 27 countries. *Global Environmental Change*, *23*(5), 1233–1248. doi:10.1016/j.gloenvcha.2013.07.020

Schütze, H., Manning, C. D., & Raghavan, P. (2008). *Introduction to information retrieval* (Vol. 39). Cambridge University Press Cambridge.

Segerberg, A., & Bennett, W. L. (2011). Social media and the organization of collective action: Using Twitter to explore the ecologies of two climate change protests. *Communication Review*, *14*(3), 197–215. doi:10.1080/10714421.2011.597250

Stine, R. A. (2019). Sentiment analysis. *Annual Review of Statistics and Its Application*, *6*(1), 287–308. doi:10.1146/annurev-statistics-030718-105242

TenHouten, W. D. (2006). *A general theory of emotions and social life*. Routledge. doi:10.4324/9780203013441

TenHouten, W. D. (2014). *Emotion and reason: Mind, brain, and the social domains of work and love*. Routledge. doi:10.4324/9780203093634

Thelwall, M. (2018). Gender bias in sentiment analysis. *Online Information Review*, *42*(1), 45–57. doi:10.1108/OIR-05-2017-0139

Wong, E. (2016, November 18). Trump has called climate change a Chinese Hoax. Beijing says it is anything but. *The New York Times*. https://www.nytimes.com/2016/11/19/world/asia/china-trump-climate-change.html

Yan, Y., Chen, J., & Wang, Z. (2020). Mining public sentiments and perspectives from geotagged social media data for appraising the post-earthquake recovery of tourism destinations. *Applied Geography (Sevenoaks, England)*, *123*, 1–13. doi:10.1016/j.apgeog.2020.102306

Yeo, S. K., Handlos, Z., Karambelas, A., Su, L. Y.-F., Rose, K. M., Brossard, D., & Griffin, K. (2017). The influence of temperature on #ClimateChange and #GlobalWarming discourses on Twitter. *Journal of Science Communication*, *16*(05), 1–25. doi:10.22323/2.16050201

Zaval, L., Keenan, E. A., Johnson, E. J., & Weber, E. U. (2014). How warm days increase belief in global warming. *Nature Climate Change*, *4*(2), 143–147. doi:10.1038/nclimate2093

Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E.-P., Yan, H., & Li, X. (2011). Comparing twitter and traditional media using topic models. *European Conference on Information Retrieval*, 338–349. doi:10.1007/978-3-642-20161-5_34

ENDNOTE

1 Parameter α means Dirichlet prior for the document topic distribution; parameter β is the Dirichlet for the word distribution; θ represents the vector for topic distribution across a document d, z represents a topic extracted from document, w is the specific words in N, rectangle D is the corpus, and rectangle N refers to the number of words in the document.

Fang Qiao is a PhD candidate at Xi'an International Studies University.

Jago Williams is a postgraduate researcher in Linguistics at Bangor University.