Saving Earth One Tweet at a Time through the Lens of Artificial Intelligence

Cuc Duong[1,2], Qian Liu[2], Rui Mao[2], and Erik Cambria[2]
1. *Interdisciplinary Graduate Program, Nanyang Environment & Water Research Institute, Singapore*
2. *School of Computer Science and Engineering, Nanyang Technological University, Singapore*
duongthi001@e.ntu.edu.sg, {liu.qian, rui.mao, cambria}@ntu.edu.sg

***Abstract*—The impacts of climate change and global warming have become more visible globally because of the increasing frequency of extreme weather events, abnormal heatwaves, and other climate crises. Besides the traditional survey method, it is beneficial to automatically distillate climate change opinions from social platforms to measure public reactions quickly. We investigate how to organize climate change opinions on Twitter into meaningful categories to support perspective summarizing tasks. We find that merely using the available taxonomy for this task is ineffective; hence we must consider the entire text content. We recommend five high-level categories (Root cause, Impact, Mitigation, Politics or Policy, Others) and assemble ClimateTweets, a dataset with category and polarity labels. In addition, we construct category classification and polarity detection tasks with a range of opinion mining baselines. The experimental results show that both tasks are challenging for existing models. We release the ClimateTweets dataset to facilitate investigation in public opinion mining using text content and artificial intelligent methods. We hope this study could pave the way for future studies in the climate change domain.**

***Index Terms*—climate change, opinion mining, topic modeling**

I. Introduction

The impacts of climate change and global warming have become more visible to global citizens because of the increasing frequency of extreme events, abnormal waves of temperatures, and wildfires [1], [2]. These events have caused considerable losses to people's lives and the economy. We argue that monitoring these hazards through people is needed to timely evaluate the cost of climate change and recognize the solutions on which people commonly agree. Researchers regularly survey public opinions surrounding the effectiveness of the deployed solutions [3], people's readiness for new policies [4], motivating factors to increase individual climate adaptation actions [5]–[7], and other aspects. These popular research topics highlight the information needs about climate change opinion mining in downstream analysis.

To complement these traditional surveys (i.e., questionnaires), mining people's opinions on a social network could provide much more data insights (million posts compared with thousand survey respondents) over a long period (via social network archived data) at a lower cost. For example, Twitter[1] offers a free full-archive search service (data from 2006) for academic research. In addition, analyzing data from Twitter could reach impactful discoveries [8]–[10].

[1]https://twitter.com/

Twitter is also a popular tool researchers use to investigate climate change opinions [11]. Studies have shown that data from Twitter are relevant to climate events and valuable for measuring opinion polarization [9], [12], [13]. Due to the absence of large annotated datasets in different climate change-related topics, most previous works took an unsupervised learning fashion [14], [15] to classify tweets into topic categories. These methods use topic models to obtain extractive topic words, then map the topic words to major topic categories by the semantics of the topic words. However, we find that the extractive topic words cannot correctly represent the actual topics of the tweets in many cases (see Table IX). Other studies used taxonomies and pre-defined rules to classify tweets into different label classes [16]–[18]. Table I shows that the extractive topic words in the taxonomy may lead to incorrect categories. These gaps highlight the significance of contextualized classification in the climate change domain.

With the rapid development of artificial intelligence (AI) and machine learning, the performance of opinion mining has been significantly improved [19]–[23]. Recent works in neurosymbolic AI [23], [24] are also making results increasingly interpretable. Supervised learning models, however, are usually trained with substantial labeled data to achieve robust performance. As a result, they are hungry for data in task-specific domains. A realistic annotated dataset can also serve as a reliable testing set before deploying the model to the market and a resource for corpus study. Unlike other research domains, e.g., movie review, product review, and news with rich label resources for opinion mining [25]–[28], the data annotated in the climate change domain are insufficient.

Table II summarizes recent datasets in the climate change domain. The Global Warming Stance Dataset (GWSD) [29] and SemEval-2016 Task 6 [26] labeled the text with people's stance toward climate change (i.e., whether or not climate change is a serious concern). Similarly, Pearce, Holmberg, Hellsten, and Nerlich [30] labeled users' stance toward the 2013 IPCC Working Group 1 report, an event related to climate change. An et al. [31] labeled the text as subjectivity or objectivity and the sentiment polarities. Climate Fever [32] labeled the data for fact-checking purposes (i.e., a statement is supported, refuted, or disputed by facts). These annotation paradigms consider climate change as *one subject* in opinion mining, where climate change sub-topics were not differentiated.

TABLE I: Correct and incorrect examples for topic modeling using Global Pulse [16] taxonomy. The first column contains the actual tweets about climate change. The second column contains the predicted categories using the taxonomy, and the last column is the general opinion about these categories.

Text	Taxonomy	Human
"A new report from the medical journal {Name} finds that human-caused #climatechange is worsening human #health in just about every measurable way."	**Risk/Disaster**, deducted from matched string *health*	Generally agree
"The system can track bleaching events in near-real-time and provide an overall view of trends and changes in coral reef health."	**Risk/Disaster**, deducted from matched string *health*	Generally disagree
"#Flooding in {Place} worsens off-season amid #climatechange"	**Risk/Disaster**, deducted from matched string *flooding*	Generally agree
"#BeFloodPrepared and establish a family communication plan for emergencies."	**Risk/Diaster**, deducted from matched string *flood*	Generally disagree
"Huge oysters that grew too big to eat are now going to be used to restore acres of oyster beds around {Name}, {City} where the restored reefs will protect against storm surges from higher sea levels #ClimateChange #Oyster"	**Agriculture/Forestry**, deducted from matched string *acres*	Generally disagree

On the other hand, another type of climate change-related dataset was developed to distinguish the impact of climate change in different domains, e.g., energy, agriculture, ocean, economy, etc. [16]–[18] (see Table III). However, these datasets do not have manually annotated opinion labels, limiting their research scope in analyzing public opinions on climate change. Besides, from these datasets, one cannot gain insight into climate change causes and mitigation as in the aforementioned downstream research [5]–[7]. We believe that annotations in both opinion and multi-dimensional climate change research topics are essential. The labeled data and its resulting models help other experts develop targeted solutions and policies to save climate change from various aspects. However, there has not been such a dataset to the best of our knowledge.

To facilitate climate change opinion mining, gaining insight into possible solutions from different aspects, we propose a novel annotated Twitter dataset, termed ClimateTweets. The dataset includes two types of labels, general categories and sentiment polarities. The general categories include labels such as ***Root cause, Impact, Mitigation, Politics or Policy***, and ***Others***. These labels distinguish tweets related to climate change from the dimensions of causality and activity, which is motivated by the information needs of downstream studies [5]–[7]. Specifically, these novel label classes help answer research questions related to the root causes, the severity of the impacts, the approval rating of the solutions, the political influences, with and without other opinions. The sentiment polarities are classified as ***positive, negative***, and ***neutral***. This label set helps us understand public views toward the above general categories. The dataset contains 2300 text samples over a year, from December 2020 to November 2021. We also conduct experiments to show the preliminary results of classifiers on the dataset. We observe that the classical pre-trained language model, e.g., BERT [33], can only achieve 0.67 micro F1 on the general category classification task and 0.66 micro F1 on the sentiment analysis task. These results demonstrate the challenges of our proposed tasks and the necessity of developing task-specific models.

TABLE II: Types of labels from recent works. The second column contains the source of the data. The third column lists the type of labels. The fourth and last column are the sample size and their availability to the public. * SemEval-2016 Task 6 [26] has 564 climate change tweets over 4870 total samples.

Dataset	Source	Labels	Count	Available
[32]	Wikipedia	Fact-checking	1535	**Yes**
[29]	News articles	Stance	2000	**Yes**
[30]	Twitter	Stance	239	No
[31]	Twitter	Subjectivity, polarity	2550	No
[26]	Twitter	Stance and polarity	564*	**Yes**
Ours	Twitter	Category, polarity	2300	**Yes**

Finally, we systematically analyze public opinions on the labeled aspects and present valuable findings and suggestions based on ClimateTweets. These findings inspire future downstream applications and corpus studies based on the models trained with our proposed dataset and the model inferences from non-annotated big data in the open domain.

The contribution of this work can be summarized as threefold: (1) We propose a new dataset[2] with novel annotation paradigms for climate change study; (2) We conduct experiments to show the preliminary results of classical classifiers on the dataset; (3) We conduct preliminary corpus studies to show the insight findings of public opinions toward climate change by using our dataset.

II. Related Works

Current climate change opinion mining can be divided into two main streams: stance detection and topic modeling.

A. Stance detection

Studies on stance detection aim to sort opinions into groups such as supportive and non-supportive to a climate policy [30] or to climate change in general [26], [29], [31], [34]. Williams, McMurray, Kurz, and Lambert [34] grouped Twitter accounts

[2] https://github.com/CucDuong/ClimateTweets

into ‘sceptic’ and ‘activist’ on climate change topics by manually labeling several ‘seed’ accounts and using a similarity reasoning method to assign labels to the remaining users. Similarly, Pearce, Holmberg, Hellsten, and Nerlich [30] labeled an account as ‘supportive’ or ‘non-supportive’ toward the 2013 IPCC Working Group 1 report, a narrower scope than [34]. The strategy of propagating labels from seed accounts to other accounts is sensible. Yet, it is hard to evaluate its accuracy because of the missing direct evidence from the users.

On the other hand, GWSD [29] and SemEval-2016 Task 6 [26] labeled people’s stance toward climate change based on sentences. Their supervised approaches are more reliable because they utilized the entire text content to determine the viewpoint. However, these datasets cover only the topic: whether climate change is a serious concern. Since these datasets consider climate change as *one subject*, they do not apply to other research questions that potentially help save climate change-related issues. Hence, we propose to group the population into five categories from the perspectives of causality and activity: ***Root cause, Impact, Mitigation, Politics or Policy, and Others***, which helps us determine the most urgent and wanted solutions by understanding the causes, the impacts, and the desired mitigating and political supports.

B. Topic modeling

Latent Dirichlet Allocation (LDA) is one of the most popular topic modeling methods on text data [35]. With the advantage of no labeled data required, LDA is suitable for quickly investigating the top prevalent topics [14], [15]. The extractive topic word list, given by LDA-liked topic models, is based on the symbolic string frequency, where the intention and contextualized meanings of the text are weakly represented by the topic words. Besides, extracting topic words from short text is particularly challenging for topic models [36].

Futhermore, some natural language processing (NLP)-based climate change studies used taxonomy to classify tweets into different label classes [16]–[18]. The taxonomy contains topic words that were pre-defined for specific research questions. These taxonomies mainly focus on different types of related impacts of climate change, e.g., weather, economy, energy, air quality, etc. (see Table III). However, the benefits of studying these taxonomies are inadequate for exploring the causes of climate change and making effective solutions. The absence of opinion annotations also limits researchers to rank different climate change factors from the perspective of public opinions.

To bridge these gaps, we annotate both the five general categories and sentiment polarities for climate change.

III. Data Collection

In this study, we develop a climate change dataset, termed ClimateTweets. The data are gathered from Twitter. The dataset incorporates two types of labels: (1) ***Impact, Mitigation, Root cause, Politics or Policy, and Others*** of climate change; (2) ***negative, positive, neutral*** sentiment polarities of public opinions. We mean to improve the research of addressing climate change by understanding public opinions.

TABLE III: Category comparison to existing works. [17] and [18] modified Global Pulse taxonomy to fit their applications.

Source	**Category**
Global Pulse [16]	General, Politics/Opinion, Weather, Economy, Risk/Disaster, Energy, Agriculture/Forestry, Arctic, Ocean/Water
[17]	Global Pulse + Negotiation/Summit,Campaigns, Air quality, Sandstorm
[18]	Energy, Weather, Economy, Agriculture/Forestry, Water, Security, Climate Denial, Air Issues, Animals
ClimateTweets (Ours)	Impact, Root Cause, Politics or Policy, Mitigation, Others

A. Querying Tweets

Using the *full archive search* function of Twitter Application Programming Interface (API) version 2 and Twitter Academic Research access [37], tweets with two hashtags `#ClimateChange` and `#climatechange` were collected. Climate change is a broad topic, and many hashtags (e.g., `#globalwarming`, `#ClimateCrisis`, `#Sustainability`) are used. However, these two hashtags are the most general ones to ensure no sampling bias toward any topic. Several users tend to re-post the same content over time, in multiple accounts, or re-post slightly modified tweets. During the annotating process, annotators may annotate and add similar tweets to the database. In order to tackle this problem, analogous tweets were detected and filtered out using the *fuzzywuzzy*[3] Python3 package.

B. Investigating the Population

We focus on the text component of a tweet. The text is labeled with one of the five categories if it composes a complete meaning. For example, this sentence, “This is fantastic! #ClimateActionNow #climatechange ReadingCAN ReadingCouncil” is out of the scope of this work because it does not deliver a complete message. In this example, one cannot figure out what the opinion target of the message owner is (e.g., the information of the referring target *This* is unknown). If the other media (e.g., images, videos, or attached news) present the central message while the text is too obscure to guess, it will not be included in this dataset. Additionally, we make sure that ClimateTweets contains no hate speech samples through manual validation.

The data investigation shows that a climate change tweet usually contains more than one topic. For example, in this tweet, “Warming temperatures have reduced the size of many #birds over the last four decades; this is emblematic of the scale of #ClimateChange impacts the world biological diversity. There is an urgent need for action,” the first sentence is about the impacts on birds, while the last sentence is a general statement to call for actions. In addition, users who advertise their sustainable products on Twitter typically start their tweets with an impact sentence, followed by the solution.

[3]https://pypi.org/project/fuzzywuzzy/

For these cases, the text is split into multiple pieces before labeling. The purpose is to maintain the consistency and accuracy of the dataset. Opinions about climate change have a broad range of sub-topics. Examples of low level and closely related topics are: “ice melting, sea-level rise, coast erosion, ocean biodiversity, sea turtles,” or “transportation, public transport, electric vehicle, vehicle battery, battery disposal.” Breaking them into meaningful categories can help identify each group’s generic features and is convenient for sentiment summarizing tasks.

C. Label Definition

We propose to categorize climate change issues into five classes. These categories are defined as follows:

- **Impact**: sentences cover the impacts, costs, and consequences of climate change to human health and security, the economy, extreme weather events, and the environment.
- **Mitigation**: sentences cover mitigation and adaptation strategies, solutions, or potential solutions to tackle climate change and climate change caused issues.
- **Root cause**: sentences cover direct and indirect factors that cause climate change or worsen the situation.
- **Politics or policy (P&P)**: sentences discuss actions from politicians, governments, inter-governmental political forums, political organizations, countries and territories, policies, other political events that affect the climate and climate solutions.
- **Others**: other sentences such as general opinions, climate advocate, call for climate action, climate denial, media, product and event advertisement, or other general climate change-related topics.

We define the sentiment polarities labels as follows:

- **Positive**: the opinion holder is supportive toward the aspect mentioned in the sentence or has a positive long-term point of view about the aspect.
- **Negative**: the opinion holder is unsupportive toward the mentioned aspect or has a negative long-term view about the aspect.
- **Neutral**: the opinion holder is neither supportive nor unsupportive toward the mentioned aspect or has neither a positive or negative point of view about the aspect.

D. Task Description

ClimateTweets defines two tasks: category detection (T1) and sentiment classification (T2). T1 is to identify which of the five categories the textual sequence is about and T2 is to detect the opinion holder’s sentiment polarity toward the textual sequence. T1 and T2 facilitate sentiment aggregation within each group so that people with climate interests can monitor public reactions at higher granularity than before.

E. Annotation Process

Step 1: Annotators were asked if a sentence(s) contained a clear message. If the answer is ‘Yes’, annotators continue to the next step, discard the sentence(s) otherwise.

```
{'tweet_id': '1402050489974505474',
 'start_char': 186,
 'end_char': 280,
 'category': 'mitigation',
 'category_polarity': 'neutral'}
```

Fig. 1: ClimateTweets format.

Step 2: Annotators located essential keywords or targets of the sentence(s). The keyword can be an explicit aspect or imply an aspect. For example, in the sentence, “#ClimateChange creeping in, #Country is one of the worlds most water-deficient countries and is now in the grip of one of the most severe #droughts in history.”, the keyword is “droughts”.
Step 3: Annotators identified the category of the sentence(s) by the hints from the located keywords, the overall meaning of the sentence(s), and the intention of the statement(s). In the previous example, the keyword here is *droughts* and the sentence is about the ***Impact*** category.
Step 4: Annotators determined the sentiment polarity toward the category. In the previous example, the opinion holder has a *negative* viewpoint toward the ***Impact*** category.

The annotators are one graduate student working on sentiment analysis for climate change and two undergraduate students. English is their official language at university. All annotators were volunteers of this project; therefore, no monetary compensation was given. The annotation process was conducted offline via an in-house Python3 interactive program. Each sample was annotated by two annotators. After that, the data manager performed a cross-checked to resolve conflicting cases (i.e., two annotators label the same sample differently). Comparing the outcomes shows that the most conflicts occurred in the polarity task, between neural-positive and neutral-negative labels. In some cases, the polarization from neutrality is subtle. Therefore, different labels were given by different annotators. For example, in this sentence, “They want to open a #SustainabilityHub in the center of #City to support the community to take action on #ClimateChange, share ideas, and learn new skills.”, the opinion holder talks about a community solution in a neutral-to-slightly-supportive view. Though the tone is neutral, the final assigned polarity is ***positive*** as the outcome of the hub is positive.

Format and Availability. ClimateTweets is made available with columns: *tweet_id, start_char, end_char, category, category_polarity*, where *start_char* and *end_char* are the start token index and end token index of the sentence from the original tweet. Users can recall the sentence using *tweet_id* and the character indices in 0-based format. We keep the start and end token index information of a target sentence so that future aspect level classification can be also conducted based on ClimateTweets.

F. ClimateTweets Properties

Figure 2 depicts the distribution of each category and its polarity. The ***Impact*** tweets are strongly correlated to the ***negative*** sentiment, which is logical because the consequences of climate change are generally harmful. Similarly, the ***Root cause*** tweets are primarily ***negative*** as a sense of urgency of this alarming phenomenon. On the other hand, the ***Mitigation*** tweets are more inclined to the ***positive*** side though the correlation is not as strong. Some negative ***Mitigation*** tweets represent the disapproval opinions toward the target solution. The potent pattern of these three categories can help segregate them from the population. This advantage demonstrates the usefulness of our proposed category list that is missing from the other works. Apart from these three skewed bars, the sentiment polarity for ***P&P*** and ***Others*** tweets is more evenly distributed. Regarding the number of samples, except for the ***Root cause***, they are relatively balanced among classes. However, the small percentage of the ***Root cause*** tweets signals a challenging mission to identify this category. Table IV gives typical examples of sentences for each category and their associated polarity.

ClimateTweets contains samples from 1739 Twitter users. Table VI shows that more than 80% of the users contribute just one sample. Hence, ClimateTweets maintains diversity in opinions toward different aspects of climate change. Using Twitter users information API[4], the locations of 1436 users have been retrieved. We only deduce the countries of 1242 users, because the others provide non-geographic information (e.g., internet, global). Table VII shows that nearly half of the users come from North America, followed by Europe and Asia. The top three countries with the most users are the United States, the United Kingdom, and India. Samples in ClimateTweets have an average length of around 20 tokens and a maximum length of 50 tokens. The distribution of the sample length is shown in Figure 3. While the ***P&P*** category curve has a bell shape, the others are skewed toward the mean and analogous to the total distribution. Table V shows that the dataset covers a wide range of sub-topics. For example, in the ***Impact*** category, these sub-topics spread from extreme weather events to agriculture and security, which are comparable to the list of the previous works (Table III).

Table V shows that some topics belong to more than one category, such as *transportation* and *fossil fuel*. If existing taxonomies are used, text that has the term, e.g., *transportation* will be classified into the ***Economy*** group[5]. However, our contextual dependent annotations show that these topic words may distribute in different classes. For example, "Emissions from the transportation sector are worsening #airpollution and accelerating #climatechange." is not relevant to the economy but rather to climate change. Hence, our proposed categories fit the content of this message better than the existing works. Moreover, people discuss transportation policy with a more positive sentiment than when they refer it as the source of emissions. Therefore, allocating transportation tweets to either ***Root cause*** or ***P&P*** can help aggregate the sentiment polarity appropriately, whereas it does not fit the existing taxonomy and rules (Table III).

[4]https://api.twitter.com/1.1/users/show.json
[5]http://unglobalpulse.net/climate/taxonomy/

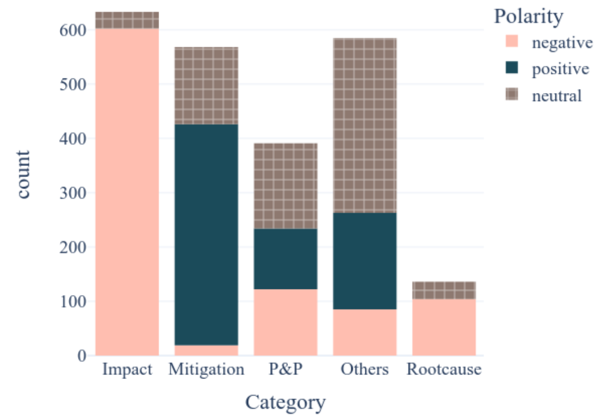

Fig. 2: Number of samples per category and polarity.

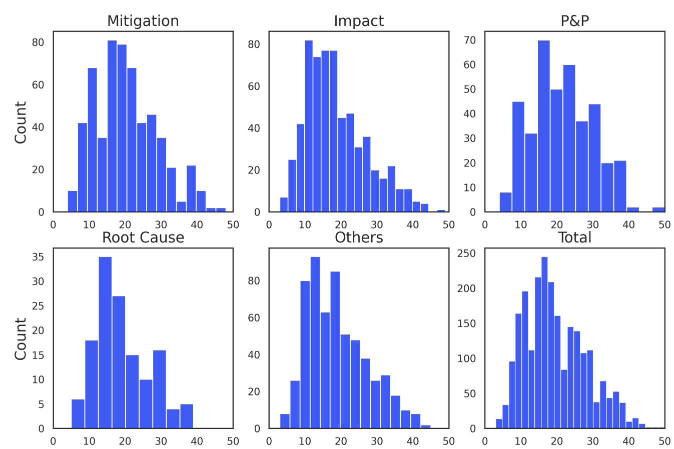

Fig. 3: Distribution of number of tokens per sample.

This characteristic is the most crucial advantage of our proposed category list compared to the previous works. Totally, we gather 2300 samples in ClimateTweets. The size of ClimateTweets is comparable to previous climate change opinion mining datasets (see Table II). It is also comparable to other sentiment analysis datasets from different domains. For example, the SemEval-2016 Task 5 dataset [25] subtask-2 has 425 English text with 1839 pairs of labels (i.e., {category, polarity}) for restaurant reviews and 475 English text with 2627 pairs for laptop reviews.

IV. Experiments

A. Baselines

Table VIII lists three classical sequence labeling classifiers employed to test on our dataset. The 100-dimension GloVe [38] pre-trained encoder is used as the input embedding layer for the GloVe-LSTM and GloVe-GRU models. While the GloVe-LSTM has two latent LSTM layers [39], the GloVe-GRU has two GRU [40] layers. On the other hand, BERT-FC utilizes the 768-D BERT encoder [33], which encodes

TABLE IV: Example sentences from each category and their associated polarity.

Category	Sentence	Polarity
Mitigation	"This #AI technique could use a #digital version of #Earth to help fight #ClimateChange #WhatsYourTechStory."	Positive
Mitigation	"With regards to #climatechange, recycling isn't the answer."	Negative
Impact	"Due to prolonged dry period within {Name} county and the entire lake {Name} basin region our tree seedlings have began to dry up."	Negative
Impact	"Glacier blood could be key to understanding impacts of climate change."	Neutral
P&P	"{Forum} confirmed pledges to increase #climatefinance contributions as part of efforts to reduce #emissions that contribute to #ClimateChange and help a move toward #CleanEnergy."	Positive
P&P	"World leaders are not showing #leadership when it comes to #climatechange."	Negative
P&P	"Putting {Country}'s #carbon tax axing into perspective"	Neutral
Root cause	"Greenhouse Gas Emissions From Food Production are Far Greater Than Previous Estimates Suggest #ClimateChange #agriculture #food."	Negative
Root cause	"Heat accounts for 37% of {Country} emissions."	Neutral
Others	"I am excited to moderate a brilliant panel of climate leaders as they explore the way forward to an equitable, climate resilient and sustainable future."	Positive
Others	"I am excited to moderate a brilliant panel of climate leaders as they explore the way forward to an equitable, climate resilient and sustainable future."	Positive

TABLE V: Most popular topics in each category. From left to right is in decreasing frequency order.

Category	Top frequent topics
Impact	extreme weather events, humanity (including health), environment, water shortage, drought, high temperatures, economy, wildlife, agriculture, wildfire
Mitigation	sustainability, new technology, alternative energy, planting trees, reduce reuse recycle, sustainable agriculture, electric vehicle, natural solutions, net zero movement, water
P&P	politician or key figure, country or region, summit, government, climate policy, political organization, fossil fuel policy, funding and financing policy, transportation policy, political party
Root cause	greenhouse gases (including carbon dioxide), deforestation, fossil fuel, agriculture, consumption, transportation, pollution (including plastic), urbanization, growing population, energy inefficiency
Others	advertisement, educational activity (e.g., conference, webinar, podcast), call for actions, denial, advocate, awareness, news media, climate change in general, private sector, protest

TABLE VI: Distribution of samples per account. Column name denotes the number of samples per account.

Num_samples	1	2	3	4	5	>5
Percentage	81.3%	13.7%	2.82%	0.63%	0.63%	0.92%

TABLE VII: Distribution of locations of the accounts. *NA: North America, EU: Europe, AS: Asia, OC: Oceania, AF: Africa, SA: South America.*

Continent	NA	EU	AS	OC	AF	SA
Percentage	48.6%	26.5%	19.2%	3.0%	2.3%	0.4%

syntactic and semantic information of a text. All models end with a fully connected (FC) layer activated by the *softmax* function to compute the probability of each class. The models were optimized by the Adam optimizer [41] on the cross-entropy losses. Additionally, we implement an LDA model with five components (i.e., five groups of topics) as the baseline for unsupervised learning methods.

TABLE VIII: Experimental models. *D* denotes dimension. *L1, L2, L3 denote Layer 1, 2, 3.*

Model	Embedding	L1	L2	L3
GloVe-LSTM	GloVe 100D	LSTM	LSTM	FC
GloVe-GRU	GloVe 100D	GRU	GRU	FC
BERT-FC	BERT 768D	FC	-	-

Later, we will compare it with the other supervised methods in the category detection task. Topic words are generated by the LDA model, then manually mapped to our defined categories according to their literal meanings. In the polarity detection task, Stanford Corenlp API [42] is used as the pre-trained cross-domain baseline.

B. Setups

In the preprocessing step, we removed URLs, HTML characters such as '&amp,' the new line character, the hex character '#,' and the mention character '@'. We set decoding rules for popular acronyms (e.g., EU: Europe, COP: climate change conference, EV: electric vehicle). However, some acronyms

TABLE IX: Category detection performance measured in F1 score. *Miti. denotes Mitigation, and R.C. denotes Root cause.*

Model	Impact	Miti.	P&P	R.C.	Others	Micro F1
LDA	0.13	0.32	0.26	0.02	0.17	0.22
GloVe-LSTM	0.50	0.38	0.27	0.05	0.30	0.38
GloVe-GRU	0.37	0.45	0.28	0.07	0.34	0.36
BERT-FC	**0.79**	**0.64**	**0.68**	**0.45**	**0.61**	**0.67**

TABLE X: Sentiment classification performance measured in F1 score.

Model	Positive	Negative	Neutral	Micro F1
GloVe-LSTM	0.22	0.60	0.22	0.43
GloVe-GRU	0.46	0.60	0.04	0.46
Stanford Corenlp	0.52	0.66	0.45	0.56
BERT-FC	**0.60**	**0.81**	**0.46**	**0.66**

were remained the same as they were related to a specific project or self-defined. The Python3 package *wordsegment*[6] was utilized to segment words that are written together without a space (e.g., #ClimateCrisis is decoded to climate crisis, #RenewableEnergy to renewable energy). A stratified sampling strategy was applied to divide ClimateTweets into train, validation, and test set with the ratio of 60% : 20% : 20%. The three models (Table VIII) were implemented in Python3 with Keras API [43]. In addition, the early stopping scheme and model checkpoint were enabled during the training. Each experiment was repeated three times with randomization, and the reported scores are the average F1 scores.

C. Results

a) Category detection: Table IX shows that the LDA model performs worse than the others. There is a 45% micro F1 score gap between LDA and BERT-FC. Its poor performance is because words generally from different categories are mixed into one topic, such as *crisis*, *carbon*, and *solar*. Hence, the topic ineffectively reflects a clear and meaningful category. On the other hand, ClimateTweets plays a crucial role in the topic modeling task to guide the other models to group tweets based on their meanings. GloVe-LSTM and GloVe-GRU achieve better results than the topic model-based method, whereas the performance of the two models is still much worse than BERT-FC. This shows that contextualized classification, e.g., RNNs vs. LDA is important for the category detection task in climate change and Twitter domains. A stronger contextualizing ability e.g., RNNs vs. BERT, yields better performance. Although BERT-FC achieves the best performance, there is still space for improvement. The most accurate class, ***Impact***, just reaches a 79% F1 score. The least one, ***Root cause***, has a 45% F1 score, likely because its sample size is only 30% of the average size. These results show that our proposed category detection task is challenging, demanding tailored task-specific models to achieve more accurate results.

b) Sentiment classification: Table X summarizes the performance of different methods on the sentiment classification task. BERT-FC method outperforms the other three baselines, achieving the highest micro F1 score (66%). However, the performance is still not as exciting as sentiment analysis on other domains, e.g., movie review [33]. The gap comes from two aspects: domain-specific knowledge and pragmatics (we will detail these issues in the later error analysis). These issues exacerbate the difficulty of sentiment classification on our proposed dataset.

[6]https://pypi.org/project/wordsegment/

Besides, we also find that all the examined models likely yield weak results on the neutral class. Many tweets are slightly polarized from neutrality; hence, their sentiment toward the target is hard to recognize. The miss-classified tweets often hold a neutral tone but a supportive or unsupportive view toward the target (see § III-C for label definitions). For example, the sentence, "The role of AI in sustainable manufacturing: AI and robotics required to tackle climate change", sounds neutral about AI, and the same polarity is assigned by the BERT-FC model. However, the participle adjective *required* presents strong support of the opinion holder toward the AI solution, so the accurate label is positive.

D. Error analysis

In this section, we summarize common errors of the BERT-FC classifier in both category detection and sentiment classification tasks.

Commonsense knowledge. Errors come from the missing commonsense knowledge about the target of the opinion. This sentence, "3 months of rain IN 5 DAYS!!", is predicted to belong to the ***Others*** category with positive sentiment. The model recognizes the exclamation marks closely related to non-neutral feelings and chooses a positive label, probably because no negative term appears in the text. However, the sentence describes an abnormal rain to imply the negative impact of irregular weather patterns. Without knowing about rain, months versus days, the model cannot correctly identify the sentiment polarity and category. Fusing commonsense knowledge [44]–[46] into the model can help to add extra information to reinforce its performance.

Dominant terms. Many sentences contain key terms or phrases that determine the mean of the whole sentence. Consider this text, "The big increase in carbon emissions started with the industrial revolution and has continued right into the 21st C. Facts Matter Climate Change Green New Deal". If the phrase *carbon emission* is removed, the attention would go to the phrase *industrial revolution*. As a result, the category would flip from ***Root*** cause to ***Mitigation***. Inside the mode, given their adjacent positions, a slight difference in their weighting factors might greatly affect the outcome.

Metaphors. Metaphoric expressions are also problematic for the sentiment classification task. The BERT-FC classifies the tweet, e.g., "we learned that carbon dioxide levels in the atmosphere just hit an all-time high" as negative, due to the metaphoric verb "hit" likely conveys negative sentiment. However, its contextual meaning is "reach", which is neither positive nor negative. One may employ metaphor processing methods [47], [48] to pre-process the text by paraphrasing the metaphors with their literal counterparts to achieve more accurate predictions.

Climate change-specific languages. Domain terminology serves a vital role in determining the polarity label. For example, token "floods" frequently appears in the ***Impact*** category and signals a negative sentiment. It probably influences BERT-FC to predict a negative polarity for this sentence, "{Name} floods highlight the need to factor in the environment while planning development.". However, the target of the opinion is about environment feasibility assessment which the opinion holder supports. Hence, the appropriate label is a positive polarity to the ***Mitigation*** category. In this case, the model seemingly does not have enough attention to the phrase "factor in the environment".

V. ClimateTweets Corpus Study Findings

Root cause. People commonly identify greenhouse gases and pollution are the direct causes. Other secondary causes are deforestation, burning fossil fuels, and agriculture. Since most of these are related to humans, it can be inferred that people view climate change as an anthropogenic phenomenon. Besides, the growing population is mentioned by a few people. We must take this cause cautiously as it helps promote innovations in food production (e.g., synthesized meat), but it can also lead to low birth rates in climate supporters.

Impact. People generally agree on the negative impacts of climate change on any living thing. Tagging climate change while mentioning extreme weather events, changing temperatures, droughts, and wildfires shows that people regard these phenomena as climate change consequences. These hazards are discussed more than other intangible impacts such as the economy or human health, suggesting that people worry about climate change more when they see or experience these events. This finding is also in line with other studies using questionnaires [49], [50].

Mitigation. Existing works on quantifying people's emotions about climate change mainly focus on negative feelings [51]. This study shows that positive sentiment is not uncommon. For example, people express their gratitude when completing a planting trees project or their hopes for a sustainable future. Future works should include positive feelings to evaluate better climate change conditions and the progress of mitigation and adaptation activities.

Politics or policy. This category has the highest polarized sentiment. Given the geographical distribution of the collected tweets (i.e., most tweets come from multi-party countries), this pattern is consistent with other studies on debating socio-economic issues [52]. Besides criticizing individuals or groups of politicians, people express desires for meaningful climate policies in fossil fuels, transportation, and funding for climate change solutions. In contrast, several concerns that policymakers might exploit climate policies for personal benefits.

Others. Many people advertise their eco-friendly products or services on Twitter with climate change hashtags. Educational events such as webinars, podcasts, conferences, and art shows are also promoted here. These tweets commonly sound positive or neutral; hence, these two labels contribute more than 80% of all samples. On the other hand, climate change denial tweets exist at low frequency.

VI. Ethical Considerations

We strictly follow the Twitter Developer Agreement and Policy[7] in querying, processing, and labeling the data. With the current format (see Figure 1), ClimateTweets could be released to the public while upholding Twitter terms and conditions. Through manual selection, we ensure that ClimateTweets contains no hate speech and entirely focuses on climate change aspects.

VII. Conclusion

Although climate change is widely recognized as a serious global issue, sentiment analysis datasets for climate change are much fewer than those in other domains, such as movie and product reviews. Previous works had to rely on unsupervised topic models due to missing a categorical-labeled dataset. At the same time, the extractive topic words are likely misclassified with incorrect labels due to the absence of contextual information. This work proposes a novel dataset, named ClimateTweets, containing category and sentiment labels to enhance opinion mining in the climate change domain. The two types of labels allow researchers to conduct future works in both aspects. The proposed categories, ***Root cause, Impact, Mitigation, Politics or Policy, and Others***, help bridge the gaps of analyzing public opinions toward climate change by the dimensions of causality and activity. Meaningful findings during the ClimateTweets corpus study show the values of the proposed labels.

We conduct preliminary experiments to show the results of classical classifiers on ClimateTweets. The moderate performance demonstrates that the classification tasks on ClimateTweets are challenging. The errors are mainly due to the inefficiency of the classifiers in managing commonsense knowledge, dominant terms, metaphoric expressions, and climate change-specific languages. These errors imply that one may need a task-specific model to achieve better results on the proposed dataset. However, these supervised baseline models exceed the unsupervised topic model-based method, yielding a large margin in the category detection task.

Acknowledgment

The authors would like to thank the volunteering annotators of this project. Cuc Duong is supported by Nanyang Research Scholarship.

References

[1] Intergovernmental Panel on Climate Change, *Climate Change 2014 – Impacts, Adaptation and Vulnerability: Part A: Global and Sectoral Aspects: Working Group II Contribution to the IPCC Fifth Assessment Report*. Cambridge University Press, 2014, vol. 1.

[2] ——, *Climate Change 2014 – Impacts, Adaptation and Vulnerability: Part B: Regional Aspects: Working Group II Contribution to the IPCC Fifth Assessment Report*. Cambridge University Press, 2014, vol. 2.

[7] https://developer.twitter.com/en/developer-terms/agreement-and-policy

[3] M. Povitkina, S. Carlsson Jagers, S. Matti, and J. Martinsson, "Why are carbon taxes unfair? disentangling public perceptions of fairness," *Global Environmental Change*, vol. 70, p. 102356, 2021.

[4] B. K.Sovacool, C. Demski, and L. Noel, "Beyond climate, culture and comfort in european preferences for low-carbon heat," *Global Environmental Change*, vol. 66, p. 102200, 2021.

[5] A. M. van Valkengoed and L. Steg, "Meta-analyses of factors motivating climate change adaptation behaviour," *Nature Climate Change*, vol. 9, no. 2, pp. 158–163, Feb 2019.

[6] T. Bouman, M. Verschoor, C. J. Albers, G. Böhm, S. D. Fisher, W. Poortinga, L. Whitmarsh, and L. Steg, "When worry about climate change leads to climate action: How values, worry and personal responsibility relate to various climate actions," *Global Environmental Change*, vol. 62, p. 102061, 2020.

[7] G. L. Bradley, Z. Babutsidze, A. Chai, and J. P. Reser, "The role of climate change risk perception, response efficacy, and psychological adaptation in pro-environmental behavior: A two nation study," *Journal of Environmental Psychology*, vol. 68, p. 101410, 2020.

[8] A. P. Kirilenko, T. Molodtsova, and S. O. Stepchenkova, "People as sensors: Mass media and local temperature influence climate change discussion on twitter," *Global Environmental Change*, vol. 30, 2015.

[9] A. P. Kirilenko and S. O. Stepchenkova, "Public microblogging on climate change: One year of twitter worldwide," *Global Environmental Change*, vol. 26, pp. 171–182, 2014.

[10] C. Mumenthaler, O. Renaud, R. Gava, and T. Brosch, "The impact of local temperature volatility on attention to climate change: Evidence from spanish tweets," *Global Environmental Change*, vol. 69, 2021.

[11] W. Pearce, S. Niederer, S. Ozkula, and N. Sanchez-Querubin, "The social media life of climate change: Platforms, publics, and future imaginaries," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 10, 2018.

[12] S. M. Jang and P. S. Hart, "Polarized frames on "climate change" and "global warming" across countries and states: Evidence from twitter big data," *Global Environmental Change*, vol. 32, pp. 11–17, 2015.

[13] T. H. Y. Chen, A. Salloum, A. Gronow, T. Ylä-Anttila, and M. Kivelä, "Polarization of climate politics results from partisan sorting: Evidence from finnish twittersphere," *Global Environmental Change*, vol. 71, 2021.

[14] Y. Jiang, X. Song, J. Harrison, S. Quegan, and D. Maynard, "Comparing attitudes to climate change in the media using sentiment analysis based on latent dirichlet allocation," in *Natural Language Processing meets Journalism Workshop*, 2017, pp. 25–30.

[15] B. Dahal, S. A. Kumar, and Z. Li, "Topic modeling and sentiment analysis of global climate change tweets," *Social Network Analysis and Mining*, vol. 9, 2019.

[16] United Nations, "United nations global pulse," http://unglobalpulse.net/climate/, 2021, accessed: 2021-09-01.

[17] S. Abbar, T. Zanouda, L. Berti-Equille, and J. Borge-Holthoefer, "Using twitter to understand public interest in climate change: The case of qatar," in *AAAI Conference on Web and Social Media*, 2016.

[18] N. Pathak, M. Henry, and S. Volkova, "Understanding social media's take on climate change through large-scale analysis of targeted opinions and emotions," in *AAAI Spring Symposium Series*, 2017.

[19] B. Liang, H. Su, L. Gui, E. Cambria, and R. Xu, "Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks," *Knowledge-Based Systems*, vol. 235, no. 107643, 2022.

[20] W. Xue and T. Li, "Aspect based sentiment analysis with gated convolutional networks," in *ACL*, 2018.

[21] Q. Liu, H. Huang, Y. Gao, X. Wei, Y. Tian, and L. Liu, "Task-oriented word embedding for text classification," in *COLING*, 2018.

[22] R. Mao and X. Li, "Bridging towers of multi-task learning with a gating mechanism for aspect-based sentiment analysis and sequential metaphor identification," in *Proceedings of AAAI*, vol. 35, no. 15, 2021.

[23] E. Cambria, Q. Liu, S. Decherchi, F. Xing, , and K. Kwok, "SenticNet 7: A commonsense-based neurosymbolic ai framework for explainable sentiment analysis," in *LREC*, 2022.

[24] E. Cambria, A. Kumar, M. Al-Ayyoub, and N. Howard, "Guest editorial: Explainable artificial intelligence for sentiment analysis," *Knowledge-Based Systems*, vol. 238, 107920, 2022.

[25] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq *et al.*, "SemEval-2016 task 5: Aspect based sentiment analysis," in *SemEval*, 2016.

[26] S. Mohammad, S. Kiritchenko, P. Sobhani, X. Zhu, and C. Cherry, "SemEval-2016 task 6: Detecting stance in tweets," in *SemEval*, 2016.

[27] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *ACL*, 2011.

[28] Q. Liu, J. Lu, G. Zhang, T. Shen, Z. Zhang, and H. Huang, "Domain-specific meta-embedding with latent semantic structures," *Information Sciences*, vol. 555, pp. 410–423, 2021.

[29] Y. Luo, D. Card, and D. Jurafsky, "Detecting stance in media on global warming," in *Findings of EMNLP*, 2020.

[30] W. Pearce, K. Holmberg, I. Hellsten, and B. Nerlich, "Climate change on twitter: Topics, communities and conversations about the 2013 IPCC Working Group 1 report," *PloS one*, vol. 9, no. 4, 2014.

[31] X. An, A. Ganguly, Y. Fang, S. Scyphers, A. Hunter, and J. G. Dy, "Tracking climate change opinions from twitter data," in *Workshop on Data Science for Social Good*, 2014.

[32] T. Diggelmann, J. Boyd-Graber, J. Bulian, M. Ciaramita, and M. Leippold, "CLIMATE-FEVER: A dataset for verification of real-world climate claims," *Tackling Climate Change with Machine Learning Workshop*, 2020.

[33] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *proceedings of NAACL*, 2019.

[34] H. T. Williams, J. R. McMurray, T. Kurz, and F. H. Lambert, "Network analysis reveals open forums and echo chambers in social media discussions of climate change," *Global environmental change*, vol. 32, pp. 126–138, 2015.

[35] D. Blei, A. Ng, and M. Jordan, "Latent dirichlet allocation," *Advances in Neural Information Processing Systems*, vol. 14, 2001.

[36] I. Vayansky and S. A. Kumar, "A review of topic modeling methods," *Information Systems*, vol. 94, p. 101582, 2020.

[37] Twitter, "Twitter api," https://developer.twitter.com/en/docs/twitter-api, 2021, accessed: 2021-09-01.

[38] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of EMNLP*, 2014.

[39] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[40] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proceedings of EMNLP*, 2014.

[41] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of ICLR*, 2015.

[42] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky, "The Stanford CoreNLP natural language processing toolkit," in *Proceedings of ACL*, 2014, pp. 55–60.

[43] "Keras API," https://keras.io/api/, 2021, accessed: 2021-09-01.

[44] E. Cambria, A. Hussain, C. Havasi, and C. Eckl, "Common sense computing: From the society of mind to digital intuition and beyond," in *Biometric ID Management and Multimodal Communication*, ser. Lecture Notes in Computer Science. Berlin Heidelberg: Springer, 2009, vol. 5707, pp. 252–259.

[45] D. Rajagopal, E. Cambria, D. Olsher, and K. Kwok, "A graph-based approach to commonsense concept extraction and semantic similarity detection," in *WWW*, 2013, pp. 565–570.

[46] E. Cambria, T. Mazzocco, A. Hussain, and C. Eckl, "Sentic medoids: Organizing affective common sense knowledge in a multi-dimensional vector space," ser. Lecture Notes in Computer Science. Berlin Heidelberg: Springer-Verlag, 2011, vol. 6677, pp. 601–610.

[47] R. Mao, C. Lin, and F. Guerin, "Word embedding and WordNet based metaphor identification and interpretation," in *Proceedings of ACL*, 2018.

[48] M. Ge, R. Mao, and E. Cambria, "Explainable metaphor identification inspired by conceptual metaphor theory," in *Proceedings of AAAI*, 2022.

[49] C. Demski, S. Capstick, N. Pidgeon, R. G. Sposato, and A. Spence, "Experience of extreme weather affects climate change mitigation and adaptation responses," *Climatic Change*, vol. 140, 2017.

[50] E. A. Albright and D. Crow, "Beliefs about climate change in the aftermath of extreme flooding," *Climatic Change*, vol. 155, 2019.

[51] J. K. Swim, R. Aviste, M. L. Lengieza, and C. J. Fasano, "Ok boomer: A decade of generational differences in feelings about climate change," *Global Environmental Change*, vol. 73, 2022.

[52] G. Pennycook, J. McPhetres, B. Bago, and D. G. Rand, "Beliefs about covid-19 in canada, the united kingdom, and the united states: A novel test of political polarization and motivated reasoning," *Personality and Social Psychology Bulletin*, 2021.