Social Media Attributions in the Context of Water Crisis

Rupak Sarkar[1*], **Sayantan Mahinder**[2*] and **Hirak Sarkar**[3] and **Ashiqur R. KhudaBukhsh**[4]
[1]Maulana Abul Kalam Azad Institute of Technology
[2]Independent Researcher
[3]University of Maryland
[4]Carnegie Mellon University
rupaksarkar.cs@gmail.com, hsarkar@cs.umd.edu, sayantan.mahinder@gmail.com, akhudabu@cs.cmu.edu

Abstract

Attribution of natural disasters/collective misfortunes is a widely studied social science problem. At present, most such studies rely on surveys or external signals such as voting outcomes. Typically, these surveys are costly to conduct and often have considerable turnaround time. In contrast, procuring social media data is vastly cheaper and can be obtained at varying spatiotemporal granularity. In this paper, we describe our recent work[1] that looked into the viability of estimating attributions through social media discussions. To this end, (1) we focus on the 2019 Chennai water crisis, a major instance of recent environmental resource crisis; (2) construct a substantial corpus of 72,098 YouTube comments posted by 43,859 users on 623 videos relevant to the crisis; (3) define a novel natural language processing task of *attribution tie detection*; and (4) design a neural classifier that achieves a reasonable performance. We also release the first data set on this novel task and important domain.

1 Introduction

Apportioning attribution for a collective misfortune is a critical research challenge previously studied in the political science literature on retrospective voting (see, e.g., [Peffley, 1984; Ferejohn, 1986]) or psychological literature on attribution (see, e.g., [Shaver, 2012]). Most such studies either relied on traditional surveys [Griffin *et al.*, 2008], or external signals such as voting outcomes [Ferejohn, 1986]. In this paper, we describe our recent work that leverages social media data and explores the viability of estimating attributions through social media discussions.

Specifically, we focus on the 2019 Chennai water crisis and design a neural classifier that takes a short social media text as input and outputs a set (empty if no attribution exists) of attribution factors (see, Table 1 for examples inputs and outputs). Water crisis is a pressing environmental challenge and grim forecasts indicate that nearly two-thirds of the world population could be water stressed by 2025 [Seckler *et al.*, 1999]. India is in the forefront of the water-stressed zones and the most-recent Chennai water crisis received widespread attention from scientists, environment activists and global media. We choose Youtube as our source of data because of it's widespread popularity in the Indian subcontinent [HindustanTimes, 2019] and global reach.

Attribution factor	Social media text
Overpopulation	people need to stop having kids otherwise this lack of good water problem will spread
Climate change	coastline cities like mumbai and chennai is going to sink under water after sea rise due to global warming while we fight for water
Deforestation	plant trees dumb ass trees will hold water as well as soil you have no trees at all that is why you have not water

Table 1: Examples of *attribution ties* in our data set.

Benefits: We expect our research (and unique data set) will open the gates for research in this important domain of nuanced analysis of crises through the lens of social media. Estimating attributions from social media data has the following benefits. Unlike traditional surveys, social media analyses are vastly cheaper, have faster turnaround time, can be conducted at different spatiotemporal granularities, and aggregate a larger number of opinions than traditional surveys can usually afford. For instance, the most-recent PEW survey[2] focused on India was conducted in 2018 on only 2,521 users. In contrast, our data set (described later) consists of comments from 43,859 users.

Challenges: While utilizing social media data to apportion attributions has undeniable benefits, unstructured text data presents a myriad of challenges that requires subtle understanding of language constructs. Consider the following example: '*stop have* ***9 kids family***'. While there is no surface level text match with the term *'population'*, humans can still infer that a growing population has been attributed as the possible cause from the semantic equivalence of *population'* and *'9 kids family'*. As there can be many equivalent ways of expressing attributions, a semantic understanding of the language is necessary for the task. However, that alone cannot guarantee success. Consider another example: '*can't feed my* ***9 kids family***'. In this example, we again see that the same

[*]Rupak Sarkar and Sayantan Mahinder are equal-contribution first authors. Ashiqur R. KhudaBukhsh is the corresponding author.

[1]This work is accepted as a long paper (paper link) in the main track of EMNLP 2020.

[2]Pew research link.

phrase *'9 kids family'* is present, yet the comment is not about the water crisis. Hence, to correctly identify attribution, we also need to understand the context in which an attribution is mentioned.

Beyond the aforementioned challenges of this nuanced task of attribution ties detection, specific to social media data in the Indian subcontinent, we faced an additional challenge in the form of spelling and grammar disfluencies observed in non native English speakers. A detailed study of the problems encountered in working with English as it is used in the Indian subcontinent can be found in [Sarkar *et al.*, 2020a].

2 Related Work

Our research draws inspiration from water research from a broad range of communities that include food policy research [Hanjra and Qureshi, 2010], earth science [Qin *et al.*, 2007], social science [Foltz, 2002], and water research [Schindler and Donahue, 2006; Narula *et al.*, 2011; von Medeazza, 2006] in a sense that we construct our list of potential attribution factors for prominent lines of work in these fields. Our work contrasts with [Oz and Bisgin, 2016], that formed different attribution hypotheses and then accepted or rejected those hypotheses based on randomly sampled data labelled by annotators, in our strong focus on automated methods.

Methodologically, our work is closest to [Liang *et al.*, 2019] that sought to detect *blame ties* between entities and a collective misfortune (the 2008 financial crisis) from well-formed texts in news articles. Our works differs in (1) its focus on unstructured text data from noisy social media with a vast majority of non-native English speakers contributing content; and (2) the lack of crisp entity boundaries.

3 Dataset

We use the publicly available YouTube API and search YouTube using the following two queries: *chennai water crisis*; and *india water crisis*. We sort the results by *relevance* and obtain a set $\mathcal{V}$, consisting of 623 videos. The comments from these videos make up our overall comment data set $\mathcal{D}_{all}$ containing 72,098 comments posted by 43,859 users.

Due to the large linguistic diversity of India (22 languages recognised by the constitution), we found that a significant portion of our data set consists of comments written in other native languages. To filter out the English comments, we use a recently proposed linguistic identification method [Palakodety *et al.*, 2020a] which has been used for document and token level language identification [KhudaBukhsh *et al.*, 2020] in similar multilingual settings [Palakodety *et al.*, 2020b; Palakodety *et al.*, 2020c]. We filter out comments made in other languages to obtain a filtered set $\mathcal{D}$ of 41,791 English comments.

3.1 Data Pruning

We first ground our work to relevant research from the urban planning, environmental science, political science and water research communities [Schindler and Donahue, 2006; Hanjra and Qureshi, 2010; Qin *et al.*, 2007; Foltz, 2002; Marshall, 2011; Rodell *et al.*, 2009; Narula *et al.*, 2011; von Medeazza, 2006]. Considering existing literature covering major water crises across the world including India, we construct a list of several factors scientific experts deem responsible for the water crisis. Next, we divide these factors in a set of 20 broad categories listed in Table 2, denoted by $\mathcal{F}$. An additional category, *religion*, was provided by our annotators upon dataset inspection.

In order to confine our search space to discussions regarding water crisis and eliminate peripheral discussions on unrelated contemporaneous events (e.g., the 2019 India-Pakistan conflict [Palakodety *et al.*, 2020a]), we further refine our dataset using an embedding based [Pennington *et al.*, 2014] similarity measure (details are presented in [Sarkar *et al.*, 2020b]). Filtering in this manner, we obtain our pruned data set $\mathcal{D}_{pruned}$ consisting of 2,282 comments (9,004 sentences). Our final dataset consists of 1,500 comments randomly sampled from $\mathcal{D}_{pruned}$ and 1,000 comments randomly sampled from $\mathcal{D}$. Upon human annotation, we obtain 24.03% comments from the randomly sampled set with at least one attribution. In contrast, 73.87% comments from $\mathcal{D}_{pruned}$ contains at least one attribution indicating the efficacy of our pruning strategy. Our final data set consists of 8,222 sentences (2,385 positives and 5,837 negatives).

Broad Category	Sub-categories
Agriculture	agricultural use, water intensive irrigation, inefficient irrigation, water intensive crops
Climate change	climate change, global warming, weather
Corruption	corruption, mismanagement
Damming	damming, impoundments
Deforestation	deforestation, nutrient loss in soil
Desalination	desalination
Government inaction	government inaction, indifference of policy makers, lack of proper funding
Groundwater exploitation	groundwater exploitation, strain on natural resources
Human activity	human activity, water intensive protein rich diet, consumption by livestock
Industrial development	industrial development, petroleum industry, water intensive industries, oil sands development
Lack of awareness	lack of awareness, lack of study
Lack of infrastructure	lack of infrastructure, inefficient distribution system
Lack of harvesting	lack of rainwater harvesting, lack of water preservation
Loss of water bodies	loss of water bodies, loss of water tables
Natural calamities	drought, flood
Overpopulation	overpopulation, excessive demand, population shift
Pollution	pollution, contamination, industrial waste water, industrial draining
Public water wastage	public water wastage, excessive usage
Religion	religion, Hindu caste system, Islam
Water Withdrawals	water withdrawals, irresponsible water pumping
Urbanization	urbanization, expansion of urban areas, land conversion, urban waste

Table 2: 21 broad categories of attribution factors.

Annotation: We use a two-step process to annotate our data set. In the first step, the annotators ascertain whether a sentence contains an attribution. If it does, in the second step, the attribution factor (factors) is (are) chosen from the factors listed in table 2. Three annotators proficient in Hindi, English and Bengali performed the annotation. A high Fleiss' κ value of 0.86 in the first step indicates a strong inter-rater agreement. We note that overpopulation, climate change, and public water wastage are the dominating attributed factors.

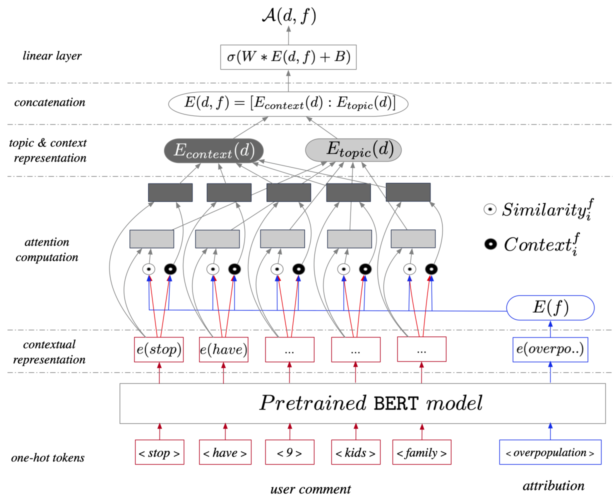

Figure 1: Model architecture.

4 Model Specification

4.1 Task

Our task in this paper boils down to detecting whether a sentence d from the sentence set $\mathcal{D}$ contains an attribution tie, and if it does, what factors f from the factor set $\mathcal{F}$ does it hold responsible for the water crisis. The second part of our task is modeled as a probability density estimation problem, where, given a sentence/attribution pair $\langle d, f \rangle$, we want to find the probability with which f is attributed in the sentence d. This distinguishes our task from standard multi-class classification problems, as our formulation can be evaluated on attribution factors that the model has not previously seen.

Due to paucity of annotated data, here we are dealing with a small data problem. We thus use a pre-trained language model, BERT [Devlin *et al.*, 2018], leaving a few parameters trainable to prevent overfitting. In order to cope with certain quirks present in social media Indian English [Sarkar *et al.*, 2020a], we use $\texttt{BERT}_{Indian}$, a BERT model fine-tuned on Indian English [Palakodety *et al.*, 2020b].

4.2 Model Architecture

The model architecture is illustrated in Figure 1. Consider a sentence-attribution pair $\langle d, f \rangle$. For every word $w_i \in d$ where $i = 1, 2, \ldots, n$ are the indices of each word in the sentence, we define $Similarity_i^f$ as a semantic similarity measure between word w_i and attribute f. Similar to the attention mechanism proposed by [Bahdanau *et al.*, 2015], we use cosine similarity between the representations of the attribution factor f and representations of w_i from the language model (LM).

We then define $Context_i^f$, which is formulated as an inversely correlated function of $Similarity_i^f$ used to capture the context in which f is mentioned in d. Our intuition here is that words that are not used to represent a topic in a positively labeled sentence are used to capture its context. Applying this intuition for a pair $\langle d, f \rangle$, we first obtain the representations of the words in the topic $e(w_j)$ for $w_j \in f$ and the representation of the words in the sentence $e(w_i)$ for $w_i \in d$. The representation $E(f)$ for the attribution factor f is calculated as the mean embedding of its constituent words, $mean_{j \in |f|}\{e(w_j)\}$ (Eq. 1).

We construct the probability function, $Similarity_i^f$ (Eq. 3), for a factor f and a word w_i, by using the cosine similarity (denoted as c_i in Eq. 2) between $E(f)$ and $e(w_i)$. The *topical similarity*, $E_{topic}(d)$ (Eq. 5) for the entire sentence d is represented as a linear combination of the contextual word embedding, $e(w_i)$ weighted by individual $Similarity_i^f$. Finally, we use $1 - c_i$ as a loose measure of inverse cosine similarity for constructing the probability function, $Context_i^f$ (see Eq. 4) and similarly generate the non-topical contextual representation $E_{context}(d)$ (Eq. 6) for the sentence.

$$E(f) = mean_{j \in |f|}\{e(w_j)\} \quad (1)$$

$$c_i = Cosine(e(w_i),\ E(f)) \quad (2)$$

$$Similarity_i^f = \sigma(\alpha * c_i + \beta) \quad (3)$$

$$Context_i^f = \sigma(\alpha * (1 - c_i) + \beta) \quad (4)$$

$$E_{topic}(d) = \sum_{i \in |d|} Similarity_i^f * e(w_i) \quad (5)$$

$$E_{context}(d) = \sum_{i \in |d|} Context_i^f * e(w_i) \quad (6)$$

In Equations 3 and 4, α and β are the hyper-parameters and $\sigma(.)$ is the sigmoid function to scale the cosine similarities to $[0, 1]$ range. The concatenation of the $E_{topic}(d)$ and $E_{context}(d)$ is used as the final representation of the $\langle d, f \rangle$ pair:

$$E(d, f) = [E_{topic}(d) : E_{context}(d)] \quad (7)$$

The final representation, $E(d, f)$, is passed through a linear layer with dropouts to model the attribution function $\mathcal{A}$ and is trained with Binary Cross Entropy loss (BCELoss) using binary labels. The linear layer, with learnable parameters W and B, is defined as follows,

$$\mathcal{A}(d, f) = \sigma(W * E(d, f)\ +\ B) \quad (8)$$

4.3 Baselines

We consider the following baselines and models.

- **Word embedding ($\mathcal{M}_{\texttt{GloVe}}$)**: Our baseline is a similarity measure based on cosine similarity between idf (inverse document frequency) weighed mean GloVe [Pennington *et al.*, 2014] embedding of the comment and the topics.
- **Classification over BERT** ($\mathcal{M}_{\texttt{BERT}}^{simple}$): This is a standard linear layer classifier trained on mean BERT embedding of the constituent words of a sentence and an attribution factor.
- **Final architecture** ($\mathcal{M}_{\texttt{BERT}}^{final}$): This is the model described in Figure 1.
- **Switching to $\texttt{BERT}_{Indian}$** ($\mathcal{M}_{\texttt{BERT}_{Indian}}^{final}$): This model is identical to $\mathcal{M}_{\texttt{BERT}}^{final}$ with the sole difference that instead of BERT, we use $\texttt{BERT}_{Indian}$.

5 Results

We evaluate the performance of our models at two granularities – attribution detection and resolution. A successful detection amounts to correctly evaluating if an input sentence

government inaction	wow that is insane i feel so bad for the people of flint how has the governor kept his job so many people should be punished for this
pollution	the land is poisoned sitting around and wishing for the magical government to fix it is what children do either install water filtering stations like in arizona or move
corruption	rick snyder is a corrupt lying sociopath cutting people off from bottles of clean water is just incredibly cruel they need to vote him out of office

Table 3: Random sample of comments detected as positives from our Flint data set.

government inaction	it would help if governments and presidencies listened less to lobbyists who have destroyed the natural world 20 times over and really invested in producing a cleaner environment
climate change	this will be the norm for many more cities if we don't start serious action on climate change...
overpopulation	cape town has seen a 70% plus rise in population in recent times. there is your answer.

Table 4: Random sample of comments detected as positives from our Cape Town data set.

contains an attribution or not. For an input sentence, a successful resolution amounts to predicting a set of factors that have a non-zero overlap with the set of labelled factor(s).

Both $\mathcal{M}^{final}_{\texttt{BERT}_{Indian}}$ and $\mathcal{M}^{final}_{\texttt{BERT}}$ perform substantially better than the baseline on both the detection and resolution tasks (see, Table 5). We obtain a modest performance improvement at the resolution task when we use $\texttt{BERT}_{Indian}$. Table 6 lists a random sample of example sentences our model correctly resolved. We notice that in the case of multiple attributions, our model could predict *overpopulation* and *deforestation*, both factors successfully. Table 7 lists some of the examples on which our model failed. One particular example points to current language models' well-established limitation to handle negations [Kassner and Schütze, 2019].

6 Performance in the Wild

We now assess our model's broad applicability in (1) detecting unseen attribution factor (2) in the wild detection. and (3)

Model	*Metric*	*Detection*	*Resolution*
$\mathcal{M}^{final}_{\texttt{BERT}_{Indian}}$	Precision	75.88	**70.14**
	Recall	81.99	61.22
	F1	**78.81**	**65.38**
	Accuracy	87.34	**81.37**
$\mathcal{M}^{final}_{\texttt{BERT}}$	Precision	66.92	59.17
	Recall	**92.58**	**66.31**
	F1	77.68	62.54
	Accuracy	86.42	79.42
$\mathcal{M}^{simple}_{\texttt{BERT}}$	Precision	74.52	22.54
	Recall	83.05	8.26
	F1	78.55	12.09
	Accuracy	**88.07**	68.28
$\mathcal{M}_{\texttt{GloVe}}$	Precision	38.30	7.45
	Recall	86.95	11.28
	F1	53.18	8.98
	Accuracy	57.62	36.77

Table 5: Performance comparison of our models and baselines. For a given task and a performance measure, the best model's performance is highlighted in bold.

public water wastage	everyone forgot within 2 or 3 month later again forget to save and waste water.
deforestation	we cut trees to build flat malls multi stored buildings
government inaction	discorperted i know where you are coming from but do not blame the farmers i think it is more of a governmental problem but farmers should not be in the reap where you sow
overpopulation | deforestation | pollution	the basic reason is population for everything cause this planet had a limit to hold people and to add more we are doing deforestation polluting our rivers air pollution and wasting water...

Table 6: Examples that our classifier correctly resolved.

public water wastage | human activity	the best way to save water is to stop consuming animal products so much of our precious water is used for animal agriculture
government inaction | urbanization	urban people are the reason for water shortage
overpopulation | no attribution	it has nothing to do with population control
human activity | government inaction	otherwise all our development is a waste if the people are being eliminated by carcinogens created due our irresponsible administration

Table 7: Misclassified instances. Misclassified attribution factor is marked with red, ground truth is marked with blue.

generalizability across other water crises.

On an input sentence *'this flu caused the water crisis'* and an additional dummy attribution factor *pandemic*, our model is able to predict *pandemic* with the highest probability.

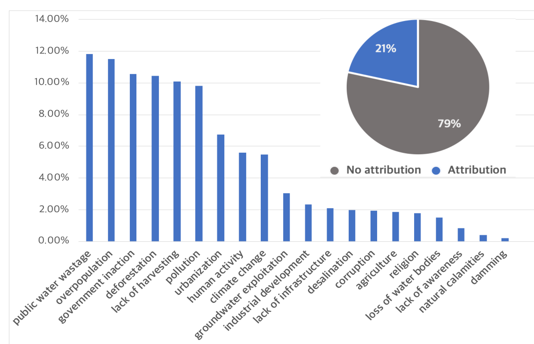

Figure 2: Distribution of number of comments detected by $\mathcal{M}^{final}_{\texttt{BERT}_{Indian}}$ model on 40k comments.

To obtain a big picture, we run our model on all 40K comments and found that *public water wastage*, *overpopulation*, and *government inaction* are the three major factors detected by our model. A human inspection of randomly sampled 200 comments aligns with the classifier predictions.

Finally, on two new data sets of 5,000 comments on the Flint water crisis (obtained from 503 videos relevant to the crisis) and the Cape Town water crisis (obtained from 237 videos relevant to the crisis), our model predicted *government inaction*, *pollution* (subsumes contamination according to Table 2), and *corruption* as the dominant factors for the Flint water crisis; and *government inaction*, *climate change* and *overpopulation* as dominant factors for the Cape Town water crisis. A human inspection of 200 randomly sampled comments aligns with the classifier's predictions.

References

[Bahdanau *et al.*, 2015] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *CoRR*, abs/1409.0473, 2015.

[Devlin *et al.*, 2018] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[Ferejohn, 1986] John Ferejohn. Incumbent performance and electoral control. *Public choice*, 50(1):5–25, 1986.

[Foltz, 2002] Richard C Foltz. Iran's water crisis: cultural, political, and ethical dimensions. *Journal of agricultural and environmental ethics*, 15(4):357–380, 2002.

[Griffin *et al.*, 2008] Robert J Griffin, Zheng Yang, Ellen Ter Huurne, Francesca Boerner, Sherry Ortiz, and Sharon Dunwoody. After the flood: Anger, attribution, and the seeking of information. *Science Communication*, 29(3):285–315, 2008.

[Hanjra and Qureshi, 2010] Munir A Hanjra and M Ejaz Qureshi. Global water crisis and future food security in an era of climate change. *Food policy*, 35(5):365–377, 2010.

[HindustanTimes, 2019] HindustanTimes. Youtube now has 265 million users in india, 2019. Online; accessed 20-April-2020.

[Kassner and Schütze, 2019] Nora Kassner and Hinrich Schütze. Negated and misprimed probes for pretrained language models: Birds can talk, but cannot fly, 2019.

[KhudaBukhsh *et al.*, 2020] Ashiqur R. KhudaBukhsh, Shriphani Palakodety, and Jaime G. Carbonell. Harnessing code switching to transcend the linguistic barrier. In Christian Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 4366–4374. ijcai.org, 2020.

[Liang *et al.*, 2019] Shuailong Liang, Olivia Nicol, and Yue Zhang. Who blames whom in a crisis? detecting blame ties from news articles using neural networks. In *Proceedings of the AAAI*, volume 33, pages 655–662, 2019.

[Marshall, 2011] Samantha Marshall. The water crisis in kenya: Causes, effects and solutions. *Global Majority E-Journal*, 2(1):31–45, 2011.

[Narula *et al.*, 2011] Kapil Kumar Narula, Ram Fishman, Vijay Modi, and Lakis Polycarpou. Addressing the water crisis in gujarat, india. 2011.

[Oz and Bisgin, 2016] Talha Oz and Halil Bisgin. Attribution of responsibility and blame regarding a man-made disaster: #flintwatercrisis. *CoRR*, abs/1610.03480, 2016.

[Palakodety *et al.*, 2020a] Shriphani Palakodety, Ashiqur R. KhudaBukhsh, and Jaime G. Carbonell. Hope speech detection: A computational analysis of the voice of peace. In Giuseppe De Giacomo, Alejandro Catalá, Bistra Dilkina, Michela Milano, Senén Barro, Alberto Bugarín, and Jérôme Lang, editors, *ECAI 2020 - 24th European Conference on Artificial Intelligence*, volume 325 of *Frontiers in Artificial Intelligence and Applications*, pages 1881–1889. IOS Press, 2020.

[Palakodety *et al.*, 2020b] Shriphani Palakodety, Ashiqur R. KhudaBukhsh, and Jaime G. Carbonell. Mining insights from large-scale corpora using fine-tuned language models. In Giuseppe De Giacomo, Alejandro Catalá, Bistra Dilkina, Michela Milano, Senén Barro, Alberto Bugarín, and Jérôme Lang, editors, *ECAI 2020 - 24th European Conference on Artificial Intelligence*, volume 325 of *Frontiers in Artificial Intelligence and Applications*, pages 1890–1897. IOS Press, 2020.

[Palakodety *et al.*, 2020c] Shriphani Palakodety, Ashiqur R. KhudaBukhsh, and Jaime G. Carbonell. Voice for the voiceless: Active sampling to detect comments supporting the rohingyas. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence*, pages 454–462. AAAI Press, 2020.

[Peffley, 1984] Mark Peffley. The voter as juror: Attributing responsibility for economic conditions. *Political Behavior*, 6(3):275–294, 1984.

[Pennington *et al.*, 2014] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[Qin *et al.*, 2007] BQ Qin, XD Wang, XM Tang, Sheng Feng, YL Zhang, et al. Drinking water crisis caused by eutrophication and cyanobacterial bloom in lake taihu: cause and measurement. *Advances in Earth Science*, 22(9):896–906, 2007.

[Rodell *et al.*, 2009] Matthew Rodell, Isabella Velicogna, and James S Famiglietti. Satellite-based estimates of groundwater depletion in india. *Nature*, 460(7258):999, 2009.

[Sarkar *et al.*, 2020a] Rupak Sarkar, Sayantan Mahinder, and Ashiqur R. KhudaBukhsh. The non-native speaker aspect: *Indian English* in social media. In *Proceedings of the 6th Workshop on Noisy User-generated Text (W-NUT 2020)*, page To appear. Association for Computational Linguistics, November 2020.

[Sarkar *et al.*, 2020b] Rupak Sarkar, Sayantan Mahinder, Hirak Sarkar, and Ashiqur R. KhudaBukhsh. Social media attributions in the context of water crisis. In *Empirical Methods in Natural Language Processing (EMNLP), 2020*, page to appear, 2020.

[Schindler and Donahue, 2006] David W Schindler and William F Donahue. An impending water crisis in canada's western prairie provinces. *Proceedings of the National Academy of Sciences*, 103(19):7210–7216, 2006.

[Seckler *et al.*, 1999] David Seckler, Randolph Barker, and Upali Amarasinghe. Water scarcity in the twenty-first century. *International Journal of Water Resources Development*, 15(1-2):29–42, 1999.

[Shaver, 2012] Kelly G Shaver. *The attribution of blame: Causality, responsibility, and blameworthiness*. Springer Science & Business Media, 2012.

[von Medeazza, 2006] Gregor Meerganz von Medeazza. Desalination in chennai: What about the poor and the environment? *Economic and Political Weekly*, 41(11):949–952, 2006.