Learning to listen: Using Tweets to observe emotional markup during extreme weather events

Anonymous ACL submission

Abstract

This work presents a data-driven approach to analyse the emotional markup of an area affected by extreme weather events. I use Twitter data collected before, during, and after Hurricane Sandy and Hurricane Harvey and use state of the art emotion recognition models to observe how emotion of people in a region changes on a macro-level as the threat of imminent danger increases. I discover interesting trends and find strong correlations between emotion markup and proximity to the extreme weather event. Based on this, I finetuned existing state-of-the-art classifiers to predict the proximity of a person to an extreme weather event based on the emotional markup of their tweet. I achieved 89-95% accuracy on unseen testing data for this task. The code and all the exploratory data is available at https://github-dev.cs.illinois.edu/atharva2/emosandy. A movie showing the time-lapse of Hurricane Sandy is available here: https://youtu.be/V7tuQaiyWCk.

1 Introduction

The ubiquity of mobile devices has ushered in a generation that expresses itself primarily using social media. This has opened up a lot of research opportunities to analyse the emotional markup of humans without any experimentation biases; we can now answer questions about human behaviour and cognition on an individual level without the use of any specialized apparatus. The topic of interest of this work is primarily to observing how the affective makeup of humans changes in extreme weather situations.

Moreover, extreme weather are of interest because they cause localized devastation. Before widespread social media, it was hard to grasp the imminent danger and despair of people affected by the disaster. However, with everyone having their own personal blog (Twitter), it is easier for people unaffected by the event to empathize with the victims. In this work, I also aim to empirically verify the claim that people unaffected by the event empathize those who are affected. To do this, I analyse data from Hurricane Sandy and Hurricane Harvey and find patterns that verify this claim.

Furthermore, I also aim to verify the claim that emotional markup is strongly correlated with proximity to an extreme weather event. I do this empirically by fine-tuning a classifier used to predict emotional markup to predict proximity to an extreme weather event. I obtain satisfactory results – with a minimum of 89% accuracy and a maximum of 95%.

Overall, the main contributions of this work are as follows:

- Present an exploratory analysis of the emotional markup of population affected by extreme weather events compared to that of the population that is not affected.
- Use state-of-the-art emotion classifiers alongside the hurricane data to predict the proximity of a user to a hurricane using only a social media text post (Tweet).

The rest of the paper is as follows: Section 2 discusses Related Work, Section 3 explains the methodology, Section 4 presents the exploratory results, Section 5 concludes the paper, and Section 6 discusses future work.

2 Related Work

The idea of using Twitter data to infer latent societal variables is not a new idea. There have been several works in the field of critical infrastructure resilience, risk management, affective computing, and other fields that have analysed this problem from different lenses.

2.1 Critical Infrastructure

The closes paper, in spirit, to mine from this body of literature is the work of Kryvasheyeu et al. [Kryvasheyeu et al., 2016]. In their paper, the authors presented a "Multiscale analysis of twitter activity before, after, and during Hurricane Sandy" that identified how "geographical and socio-cultural differences" manifested in Twitter data. In the process, they also identified relationships between Twitter activity in a region and damage inflicted by the hurricane in said region. My work builds upon the assumptions and observations from this work but asks how the relationship between emotional markup of people and twitter activity is correlated.

Another paper that was highly influential was the work of Jacob Heglund in his Master's Thesis [Heglund, 2020] on developing statistical learning and machine learning models for critical infrastructure resilience. His initial analysis demonstrated that the frequency of tweets in one hour in New York was highly correlated with the power outages experienced in the state during Hurricane Sandy. He leveraged this analysis to make time-series models that predicted the behaviour of power systems based on this. My work is similar to this in that I use machine learning models that leverage Twitter data from an extreme weather event to predict the proximity to the event instead of predicting power loss.

2.2 Risk Management

The paper that was closest, in spirit and in goals, to mine from this domain was the work of Venkata Neppalli et. al. [Neppalli et al., 2017] in identifying the polarity of sentiment expressed by users during disasters. The authors designed and developed sentiment classifiers and applied them to tweets from Hurricane Sandy. One of their key findings was that sentiment changed not only based on the geographical location, but also based on the distance from the epicenter. My work is similar to the work done here in that I analyse Twitter data and also notice a strong correlation between emotions and distance to epicenter. My work differs in that I use emotion classifiers instead of sentiment. Also, I do not make a novel sentiment/emotion classifier instead choosing to fine-tune existing architectures.

2.3 Affective Computing

There has been a lot of work in the natural language processing community, in general, in analysing the emotional markup of Twitter data. The one that I found most influential in my work was the work of Svetlana Kiritchenko et. al. [Kiritchenko et al., 2020] in analysing the "language and emotions associated with the state of being alone." They did this by collecting a corpus of tweets with keywords related to "loneliness" and analysed trends within the corpus. On a high level, the paper tries to answer the question : "how do humans express themselves when they feel isolated." In this work, I endeavor to ask a similar question from a different direction: "how do other humans connect with people who are isolated?" A storm is a localized event. I try to ask how the affective makeup of the unaffected population compares to that of the affected population.

3 Methodology

Through this research I aimed to analyse the emotional markup of two populations, and augment state of the art methods to predict proximity to the extreme weather event. Both of these tasks required a fairly sophisticated pipeline that involved *Data-set Collection*, *Data-set Preprocessing*, *Model Selection*, *Fine-tuning*.

3.1 Data-set Collection

The first step was to collect raw tweets for Hurricane Sandy and Hurricane Harvey. For Hurricane Sandy, Arkaitz Zubiaga et. al. [Zubiaga and Ji, In Press.] scraped Twitter to collect tweets from October 25th, 2012, to November 4th, 2012 using the keywords 'hurricane' and 'sandy' and all variations of the same. This amounted to nearly 15 million tweet ids. Figure 1 shows a sample of this dataset. Similarly, a dataset containing 7 million tweet ids was collected by Mark Phillips [Phillips, 2017]). Both these data-sets were scraped using Hydrator [Hydrator, 2020].

Because the data-sets are very old, a large portion of the tweets have expired and weren't recoverable. Figure 2 gives some more statistics about the number of tweets that were lost. Nevertheless, the number of tweets were sufficient for my analysis.

The hurricane tracks were obtained from IBTrACS [Knapp and Neumann].

	Sandy Text
0	So there's going to be a hurricane tonight and classes aren't cancelled!? Am I missing something here? #comeon #getittogetherguys
1	They named the hurricane that will be coming up the coast this weekend Frankenstorm.. #wtf
	Harvey Text
0	We are lucky to be safe from #hurricane #harvey. Our friends and family in South Texas are... https://t.co/t18E577hXQ
1	Pray for Texas! #harvey @ Houston, Texas https://t.co/LWW9cp7M0V

Table 1: Sample entries from raw data

3.2 Preprocessing

Because the dataset was too big to fit into my computer's RAM, I used Dask distributed data frames to multi-process the dataset in small enough batches that could fit in computer RAM. In each batch, I extracted the tweets that had a geolocation attached to it. This was done because the visualization part required latitude and longitude information. After this, an R-Tree was used to extract country, state, and county-level information about each tweet from the latitude and longitude. All duplicate tweet ids were removed. Any corrupted rows were also removed.

Overall, the following information was extracted from the raw dataset:

- `full_text`: The text for the tweet.
- `created_at`: The exact time the tweet was created in UTC.
- `id`: The twitter assigned unique id for the tweet.
- `lat`: The latitude of user location.
- `lon`: The longitude of user location.
- `name`: The name of the city extracted from the user location.
- `admin1`: The state extracted from the user location.
- `admin2`: The count extracted from the user location.
- `cc`: The country extracted from the user location.

After this, any unescaped HTML characters were decoded. No URL's, hashtags, etc were removed as this might have biased the pipeline downstream.

However, for the hurricane tracks, the missing entries were interpolated with the mean of the closest values. The timestamps were also normalized to UTC time.

3.3 Emotion Inference

To infer the emotive content of each tweet, I used the PyTorch implementation of DeepMoji [Felbo et al., 2017]. DeepMoji is one of the most performant models for emotion detection. It is based on an auto-regressive architecture that takes a string as input, and produces an emoji as output. This is done using a neural network model that converts each word to a word embedding and passes it through two bi-directional LSTM layers for capturing context followed by an attention layer with skip connections to the previous layers. Finally, a softmax layer converts the embedding into a class probability. The DeepMoji model was pretrained on 1.2 billion tweets. I specifically used the DeepMoji pretrained model because the model was trained using multi-lingual tweets and can hence do inference in other languages. This was very important since Hurricane Sandy originated near the Caribbean Sea and hence, a lot of the early tweets were not in English.

For processing the tweets, I used the same tokenization algorithm that was used to pretrain the DeepMoji architecture. For each sentence, the model predicted 5 emoji's. All 5 emoji's were retained. The sentiment of the emoji was also inferred using the emoji sentiment lexicon [Kralj Novak et al., 2015]. These rankings were not used for the final analysis.

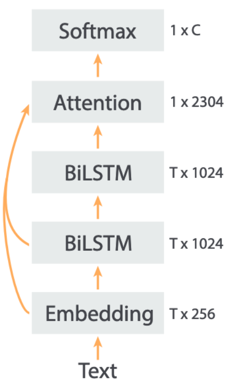

Figure 1: DeepMoji Architecture. Obtained from the original paper.

3.4 Visualization

After the inference, the data was passed to the visualization script. The visualization script first cleaned the data of any null values. Then, it binned the raw tweets based on their location. This was necessary because of two reasons. First, from a practical standpoint, plotting 50,000+ tweets was computationally infeasable. Each update would take too long. Second, the raw tweets ended up being too noisy and it wasn't possible to infer macro-trends from them. To alleviate this, as mentioned, the tweets were binned according to the corresponding S2Cell. S2 cells are a mathematical mechanism that helps computers translate Earth's spherical 3D shape into 2D geometry [S2Geometry, 2012]. Each S2Cell encompossas the same area and is hierarchically ordered. Figure 2 shows a visualization of level 14 S2Cell's in Manhattan. I specifically used S2Cell's as they allow flexible binning of geographical data.

Thus, the data was binned according to a default S2Cell level of 14. This was chosen empirically and allowed enough granularity bin tweets into manageable groups without obscuring too much information.

Each bin was labelled with the tweet that occurred the most in the bin, during the given time period. The Hurricanes were plotted based on the latitude and longitude. The radius of the circle was set to the intensity of the wind speeds of the hurricane during the time period.

After this, the results were plotted individually using plotly [Inc., 2015]. A Lambert conformal conic projection over the northern American continent centered on the -80 degree longitude was used.

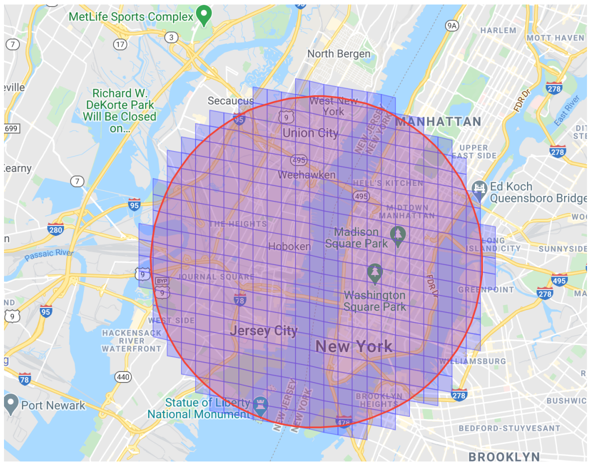

Figure 2: Level 14 S2Cell's over Manhattan. Obtained from RegionCoverer [S2Geometry, 2012]

3.5 Fine-Tuning for Proximity Inference

The final task for the dataset generation is to generate the dataset that will be used for fine-tuning the DeepMoji architecture to classify the proximity of a tweet to the hurricane. For this, I freeze all the DeepMoji layers except for the final softmax layer and retrain the model. To understand this, lets revisit the DeepMoji architecture. The architecture consists of an embedding layer, two bidirectional LSTM layers, an attention layer, and a softmax layer. The embedding layer is responsible for converting the input tokens into vectors that encode semantic meaning. This layer has already been "optimized" for Twitter data and since the Hurricane data is also from Twitter, I assume that no transfer learning is required here. By a similar argument, the BiLSTM layers, which are responsible for encoding word association and the Attention layer, which is responsible for eliminating the text-distance bias, can also be frozen during the training process.

Hence, I split the hurricane dataset into two parts: training and validation. Two-third of the dataset is used for training and one third of the dataset is used for validation. Moreover, for testing, I use the other hurricane dataset as an unseen test set (if training on Hurricane Sandy data, test on Harvey data and visa-versa).

To calculate the proximity of the user to the extreme weather event, I use the haversine distance between the tweet location and the last known location of the event epicenter. This yields:

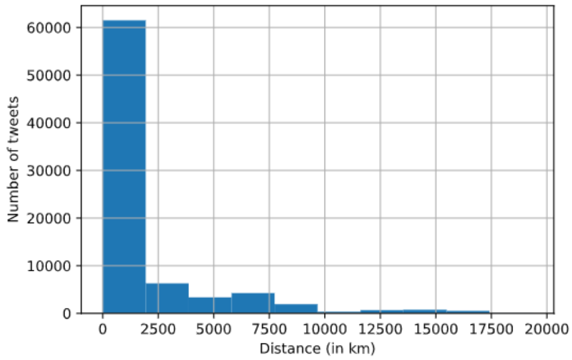

Figure 3: Distribution of Hurricane Sandy tweet proximity.

$$d = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\phi_2-\phi_1}{2}\right) + \cos(\phi_2)\cos(\phi_1)\sin^2\left(\frac{\lambda_2-\lambda_1}{2}\right)}\right) \quad (1)$$

After calculating the haversine distance between the tweet location and the last known location of the epicenter, the data is binned into discrete values such that the data is evenly split. This is done because the DeepMoji architecture is restricted to classification tasks. Figure 3 displays the distribution of tweets and their distance from the hurricane.

4 Experiments & Results

The visualization were made on an Intel i5-2415M CPU at 2.30 GHz with 16GB of RAM running Linux 18.04 (Buster). The DeepMoji inference was performed on a shared workshapce running Intel Xeon E5-2640 CPU at 2.50 GHz with 125GB of RAM. Due to compute time and resource restrictions, the inference was only run for 5 epochs for each binning. A high learning rate was used which allowed faster convergence and alleviated the low number of training epochs.

4.1 Exploratory Results

The results are presented below.

4.1.1 Emoji Scatter Plot

The following figure shows an interesting trend where the top 5 emoji's correspond to empathetic emotions (`stay strong`, `thoughts and prayers`, `love`) during the active hours of hurricane sandy from people outside the affected zone. This is especially true for October 29th and October 30th when the `thoughts and prayers` emoji peaks usage. This signifies that

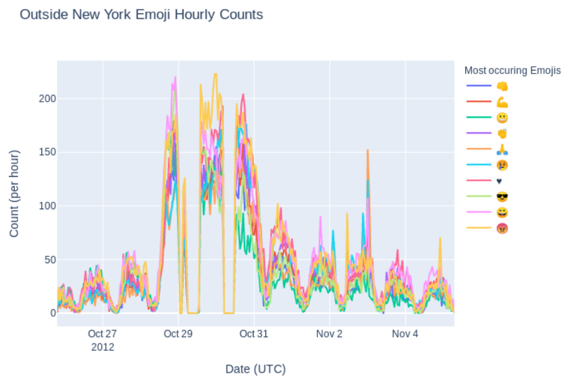

Figure 4: Line Plot of hourly count of emojis from outside New York

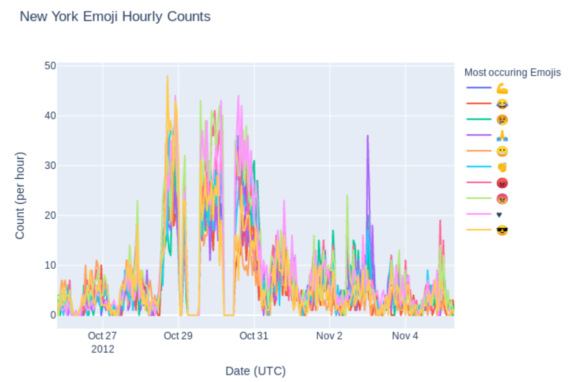

Figure 5: Line Plot of hourly count of emojis from within New York

people outside the area are empathetic to those who are suffering.

Moreover, if we look at 5, the top emotions are a mix of empathy and anger. The purple line, which signifies the frequency of the `anger` emoji, increases around the time when the hurricane was at its peak. People are also empathetic towards others in the city.

4.2 Emoji Map

A similar trend is shown in Figure 6 where during the worst part of the hurricane, people from across the United States were empathetic towards those affected. This is evident in that the emotions in the places that are farther away from the hurricane show empathetic emotions (`thoughts and prayers`, `crying face`, `sad face`, `covering eyes`, `love`, `heartbreak`).

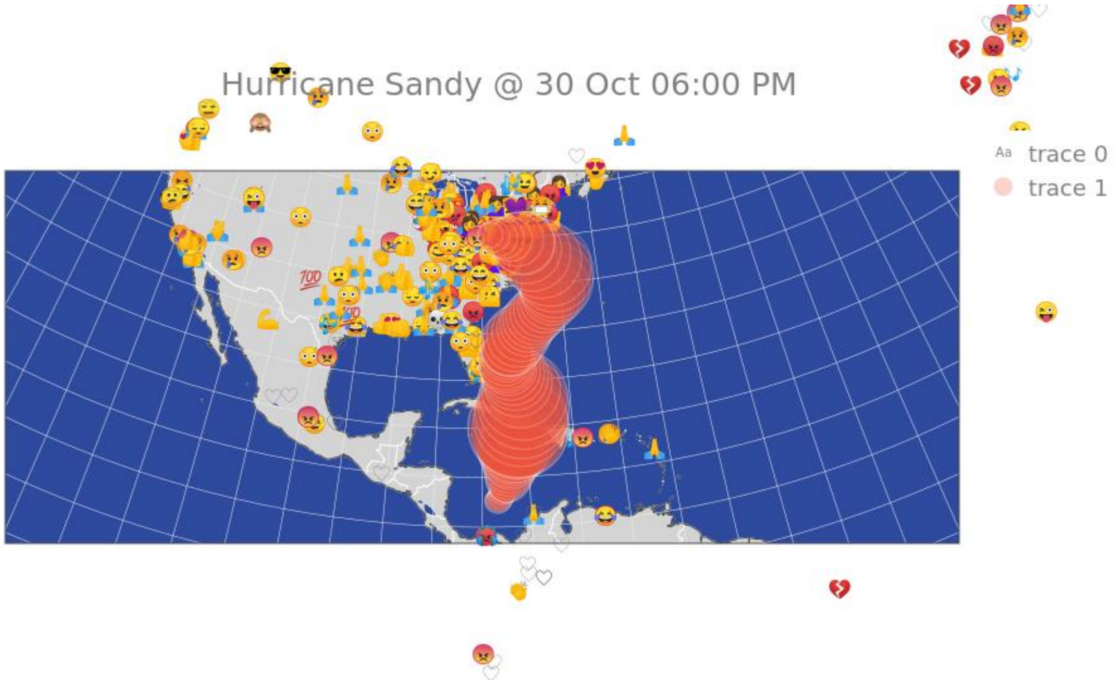

Figure 6: Visualization of Hurricane Sandy and Tweets. The red spheres mark the trajectory of hurricane sandy. Larger spheres mean higher intensity. The emojis are scaled using a custom heuristic to enhance readability.

4.3 Inference Analysis

After this exploratory analysis, I fine-tuned the DeepMoji classifier to predict the proximity of a person to the hurricane based on the emotional markup of the tweet. The results are shown in the following Table.

5 Discussion & Conclusion

In this work, I presented an empirical way to analyse the emotional markup of a population affected by Hurricane Sandy and by Hurricane Harvey. Towards that end, I found an interesting trend where people outside the affected area were more empathetic towards people inside the affected area. I also presented a way to augment a state-of-the-art emotion classifier to instead predict the proximity of the text to the extreme weather event. The high accuracy of the model thus signifies a strong correlation between emotional markup and proximity to the hurricane.

6 Future Work

There are a lot of interesting that I can think of to extend this research. Firstly, I did not explore the usage of metrics provided by Twitter such as retweets and favourites. I think an analysis of how these website specific metrics effects the emotional markup of the text would be interesting.

Epochs	4 bins	5 bins	6 bins
0	0.9700	0.950	0.380
1	0.9800	0.980	0.960
2	0.9850	0.985	0.970
3	0.9850	0.985	0.970
4	0.9850	0.985	0.970
Test Accuracy	**0.9205**	**0.900**	**0.891**

Table 2: Validation Accuracy for Hurricane Harvey. Last row is test accuracy on unseen data.

Epochs	4 bins	5 bins	6 bins
0	0.78	0.563	0.7239
1	0.867	0.836	0.8144
2	0.87	0.851	0.8162
3	0.879	0.851	0.8164
4	0.87	0.851	0.8164
Test Accuracy	**0.949**	**0.9435**	**0.9415**

Table 3: Validation Accuracy for Hurricane Sandy. Last row is test accuracy on unseen data.

Secondly, in this work, due to computational constraints, I did not train a new classifier. It would have been interesting to quantify the performance of the DeepMoji model compared to a classical machine learning model such as an SVM or Hidden Markov Model. Thirdly, the dataset can be extended by querying services that cache twitter data such as Crimson Hexagon. This would allow a richer representation of the macro-level trends observed. Fourthly, auto-regressive models that are made for classification can also be used for generating text. It would have been interesting to see the sentence tokens that correspond with different degrees of proximity to the epicenter.

7 References

References

Yury Kryvasheyeu, Haohui Chen, Nick Obradovich, Esteban Moro, Pascal Van Hentenryck, James Fowler, and Manuel Cebrian. Rapid assessment of disaster damage using social media activity. *Science Advances*, 2(3), 2016. doi: 10.1126/sciadv.1500779. URL https://advances.sciencemag.org/content/2/3/e1500779.

Jacob Scott White Heglund. Statistical and machine learning models for critical infrastructure resilience. Master's thesis, University of Illinois at Urbana-Champaign, 2020.

Venkata K Neppalli, Cornelia Caragea, Anna Squicciarini, Andrea Tapia, and Sam Stehle. Sentiment analysis during hurricane sandy in emergency response. *International journal of disaster risk reduction*, 21: 213–222, 2017.

Svetlana Kiritchenko, Will E. Hipson, Robert J. Coplan, and Saif M. Mohammad. Solo: A corpus of tweets for examining the state of being alone, 2020.

Arkaitz. Zubiaga and Heng Ji. Tweet, but Verify: Epistemic Study of Information Verification on Twitter. *Social Network Analysis and Mining*, In Press.

Mark Edward Phillips. Hurricane harvey twitter dataset. https://digital.library.unt.edu/ark:/67531/metadc993940/, 2017.

Hydrator. Documenting the now. https://github.com/docnow/hydrator, 2020.

M. C. Kruk D. H. Levinson H. J. Diamond Knapp, K. R. and C. J. Neumann. The international best track archive for climate stewardship (ibtracs): Unifying tropical cyclone best track data.

Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2017.

Petra Kralj Novak, Jasmina Smailović, Borut Sluban, and Igor Mozetič. Sentiment of emojis. *PLOS ONE*, 10(12):1–22, 12 2015. doi: 10.1371/journal.pone.0144296. URL https://doi.org/10.1371/journal.pone.0144296.

S2Geometry. S2 cells, 2012. URL https://s2geometry.io/devguide/s2cell_hierarchy.html.

Plotly Technologies Inc. Collaborative data science, 2015. URL https://plot.ly.