Sentiment Analysis of Twitter Discourse in the Context of UN Climate Conferences

James Hughes, Ameer A Saleem, Vignesh Sridhar, Holly Taylor.
Supervised by Janique Krasnowska and Fernando A Zepeda

17 March 2023

Abstract

The following document entails a student-led end-to-end data science project, the goal of which being to investigate the changes in perceptions towards climate change over time. The project involved a review of the current literature and available data, sentiment analysis, neural network model training and evaluation of the developed model, as well as projections for the future.

Contents

1 Introduction

We discussed many different avenues for research and project focus initially. Below are detailed a few of these ideas and their resolutions:

- Analyse pledges made by political leaders and conduct some data-backed investigation into how much they keep to their word.
  Pros: Plenty of data available, can be investigated in many different scopes (e.g. national vs global, a few select politics vs advocates for a particular party, etc.).
  Cons: The language used by such politicians is constrained, which could make it difficult to train a model. Also, it would be difficult to associate a metric with keeping promises, as this could be subjective and it is unlikely that a database encapsulating this with objective labels exists. The scope of individuals to investigate is also somewhat constrained, regardless of the aforementioned scope.

- Investigate the changes of attitudes of the public towards climate change following different climate summits.
  Pros: A genuinely interesting question, the implications of which could be used to forecast how attention on climate change could evolve through the years.
  Cons: Very ambitious scope for a student-led project- the title needs refined in order for the problem to be well-defined.

After considering the benefits and drawbacks of each of the project ideas we eventually settled on the key questions outlined below.

Twitter is a widely-used social media platform which can be used to sample the public's reaction to international events. Our study used historical data of Tweets occurring across the courses of multiple COP conferences from COP24 to COP26. We used a human-labelled dataset [1] to construct a model to classify the sentiment of a Tweet, limited to *negative*, *neutral*, or *positive*, which we then used to label our original dataset. Additionally, we manually categorised the Twitter accounts in our original data - limited only to the top 100 users ordered by the maximum likes of any of their tweets in the dataset. Our analysis was then centered around the following key questions:

1. Who are the most influential voices in the Twitter discourse that surrounds the COP events?

2. Is there a relationship between the sentiment of a tweet and how much exposure it receives?

3. How do these questions change over time, both across the course of each event, and across the years?

From here, we conducted a literature review to ascertain which of these avenues would be the most fruitful to pursue.

2 Literature review and pre-requisites

2.1 Related Work

[Insert information about the literature on the NLP Teams Channel]

2.2 Scraping Tweets

The literature suggested that online tweets would prove useful as an abundant data source for the project. While Twitter has an API for scraping tweets, we settled on the use of snscrape, which appeared easier to implement.
Twitter is a prevalent platform with immense usage, meaning that scraping within an arbitrary timeframe will yield a lot of irrelevant data. It makes reasonable sense to focus on time intervals before, during and after COP summits; this is the time when most 'notable figures' would be discussing climate change and voicing their respective opinions on it the most. This also extends to the public, providing us a greater density of climate-related data. Since Twitter does not allow making the text of tweets public, any Twitter-related dataset must undergo a process called hydration.

2.3 COP Conferences

COP, standing for 'Conference of the Parties', is the main decision-making body regarding any major decisions regarding climate change. COP meetings are organised by the United Nations and involve almost every country in the world. The goal of such summits is to discuss what major governmental authorities can achieve in the fight against climate change. COP meetings usually take place every year and aim to review the global progress made in mitigating the impact of climate change. The first COP summit took place in Berlin in March 1995. At the time of writing, the most recent COP summit, COP 27, took place in Egypt in November 2022.

2.4 Word Embeddings

In order to get a computer to understand and interpret the words in each tweet is to first convert them into vectors in higher dimensional space. The idea is to give words with similar meaning/context a similar vector value- for example, we may want to give the words 'lion' and 'leopard' similar vector representations in a bank of words containing the names of all the animals in the world. There are a few choices as to how this can be achieved, such as Word2Vec or GloVe's pre-trained word embeddings. We went with the latter for this project.

This begs the question of this technique's utility. Once the words of a dataset have been word embedded, we can implement the clustering algorithm to have the computer try and identify patterns in the words for us. For the sake of this project, we investigated clustering with different numbers of clusters in the EDA portion of the project.

We encountered another potential limitation when manually categorising the data in our scraped tweets. A lot of tweets were indeed related to climate change, but sometimes they discussed an individual rather than the topic of climate change itself, for example praising Greta Thunberg for her efforts in this space. This is not directly related to our goal of attributing sentiment labels to these tweets *with regards to stance on climate change*.

2.5 Clustering techniques

We briefly outline a few different clustering techniques below:

1. Centroid-based clustering: these types of algorithm work by assigning the optimal location for the center of each cluster, known as the *cluster centroid*). The position of each centroid is updated by finding the average of the vectors that belong to that cluster, terminating once a pre-specified tolerance has been reached. The K-Means algorithm is the most popular centroid-based clustering technique.

2. Hierarchical clustering: this algorithm creates a tree of clusters, with a larger number of clusters corresponding to additional branches in these trees. As the name suggests, this method of clustering is best suited to hierarchical data.

3. Density-based clustering: this algorithm creates clusters based on the density of points, allowing for varying shapes and sizes of clusters. However, this form of clustering does not generalise well into higher dimensions.

4. Distribution-based clustering: this approach assumes that the data follows a probability distribution. The closer the data points are to the centre of the cluster, the higher the probability they belong to that individual cluster. This method of clustering ought not to be used if the underlying data distribution is unknown (or doesn't exist).

3 Data

3.1 Main Twitter Datasets

We used the python library *snscrape* [citation] to acess historical tweet data via the usual Twitter API query. For the datasets relating to each conference we limited the query to English Tweets containing the string "COP" (case-insensitive). Each tweet was collected along with important meta-data such as information about the author account, and metrics such as number of likes and replies. The date range for each tweet is shown in table [table].

We then used Pandas to create a DataFrame summarising each user in the dataset. The top 100 users by maximum tweet likes were labelled with one of nine pre-defined categories [table].

3.2 Implementation

We began by focusing our attention on dates close to/during COP summits. We scraped tweets containing the string '#COP–' within time frames of a few weeks at a time during periods of high COP discussion and activity, yielding tens of thousands of tweets in each scrape. Our goal at this stage was to explore patterns at both tweet and word level.
We started by accumulating all the tweet data in each scrape and clustering at the word level, with both K-means clustering and agglomerative clustering. The most interesting insights were as follows:

- The most frequently used words in each scrape were what are called *stop words*. In the context of NLP, these are words that carry very little information, as seen in Figures 1a and 1b. This was something we would have to take care of during data pre-processing.

- For K=__, the K-Means clustering algorithm was able to independently group together tweets pertaining to different categories of *people*, i.e. separate clusters related to politics, climate change activism and environmental science. This gave us the idea to categorise the tweets by the occupation of the user (or their relation to the topic of climate change).

- The dendrograms from the agglomerative clustering showed that...

3.3 Sentiment Analysis

We implemented a neural network model in Tensorflow to learn the patterns of label associations with the tweets.

In order to train the model, we required a dataset with sentiment labels attributed to each tweet. This could be done by a computer with clustering,

	Word	Count
10	the	5138
66	to	3545
23	#cop24	2869
34	and	2464
12	of	2058
21	at	1690
42	in	1573
47	a	1491
56	climate	1301
26	for	1270

(a) Enter caption here

	Word	Count
20	the	3959
7	to	2510
51	#cop25	2208
11	and	2054
75	of	1612
22	in	1380
59	at	1142
126	a	1062
0	for	1030
54	is	1012

(b) Enter caption here

Figure 1: The top 10 most common words in the COP24 and COP25 tweet scrapes

but defeats the purpose of using the dataset for training, since we would have no feasible way of validating so many rows of labelled data. There exists a Climate Change Twitter Dataset that, among many other things, attributes a value in the interval $[-1, 1]$ with 1 denoting positive sentiment, -1 denoting negative sentiment and 0 indicating a neutral stance on the topic. We decided on using this dataset **[cite properly later]** consisting of 1.6 million tweets across 15 different languages. We started with just the tweets written in English, with the project potentially extending to tweets in other languages if time allowed. Note that the tweets in these datasets also require hydration to use.

We trained five neural network models in TensorFlow on the English tweets sentiment dataset. The testing and training accuracies are shown in Figure 2. From this, we opted with our first model, consisting of architecture shown in Figure 3.

Out of the three categories, random chance yields a success rate of 33%, which we certainly want our model to exceed. On the other hand, our training data contained class imbalance, with a proportion split of $34.9\%, 46.0\%$ and 19.1% for positive, neutral and negative tweet labels respectively. This led to us deciding on a baseline accuracy of 45% when comparing the models' accuracy values.

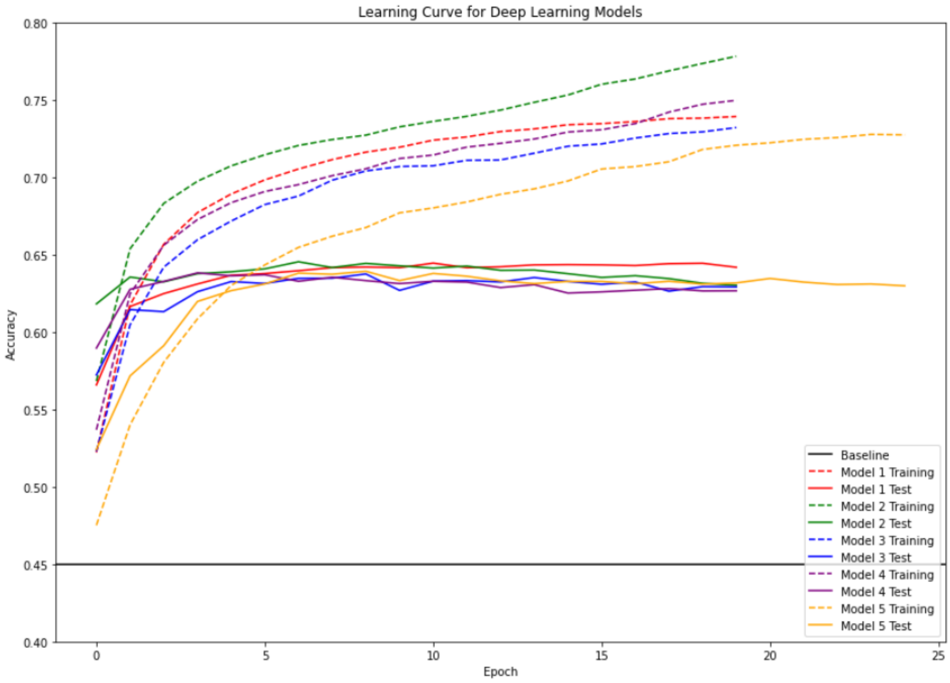

Figure 2: A plot of the training and testing accuracies of the five neural network models.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, None, 32)          640000

 global_average_pooling1d (G  (None, 32)               0
 lobalAveragePooling1D)

 dense (Dense)               (None, 3)                 99

=================================================================
Total params: 640,099
Trainable params: 640,099
Non-trainable params: 0
```

Figure 3: Model 1 architecure.

4 Analysis

4.1 COP24

4.2 COP25

4.3 COP26

5 Conclusion

5.1 Ambiguity of Sentiment

Suppose that one of the tweets in our scrape was the following: " 'I'm so happy that everyone is working so hard to solve the issue of climate change', said no-one ever.".[1] The tweet indicates a negative sentiment with regards to climate change. However, the use of the word 'happy' could potentially cause our model to classify this as a tweet with positive sentiment. This would not only reduce the accuracy of our network, but the tweet itself is inherently destructive to the model training process; it may suggest to the network that the word 'happy' ought to be synonymised with negative sentiment rather than positive (which it ought to in isolated context). Therein lies an inherent drawback in our pipeline, as sarcastic tweets are difficult to deal with, and there's no feasible way for us to filter them out of the dataset.

Another issue we came across during the course of the project was related to the reliability of the sentiment labels attached to the tweets in our dataset. For example, we initially trained our neural network on the Climate Change Twitter Dataset. However, we had some difficulties on getting our prediction accuracy above a certain amount. Upon closer inspection of the dataset, we as a team found that we didn't totally agree with some of the labels given, hence why we tried another dataset to much more satisfactory results. Depending on the project, it is certainly worth sampling a few records of data to convince yourself that you agree with the data you're using.

5.2 Next steps

References

[1] Sara Rosenthal, Noura Farra, and Preslav Nakov. SemEval-2017 task 4: Sentiment analysis in Twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 502–518, Vancouver, Canada, August 2017. Association for Computational Linguistics. doi: 10.18653/v1/S17-2088. URL `https://aclanthology.org/S17-2088`.

[1] Note that this made-up sentence does not necessarily reflect the opinions of any of the collaborators of this project.