UN & climate change: does a growing concern of UN countries regarding climate change result in sustainable actions?

Richard Chou

University assignment
University of Amsterdam, Science Park 904, 1098 XH Amsterdam

Abstract. This paper describes exploratory data analysis that examines if countries that address climate change more frequently during the yearly UNGA do more to stop global warming. The UN speeches are enriched with climate data from The World Bank to explore correlations. Preliminary results show no significant correlation between emissions and the frequency of climate change-related mentions.

Keywords: Climate Change · Sustainability · Exploratory Data Analysis · UNGA · Topic Modeling

1 Introduction

Heads of state or government from around the world have gathered each year since 1946 to address major international problems at the United Nations General Assembly (UNGA). Each country gets to make their most important statements during a speech. The text from these speeches is available as data from 1970 onwards [1]. Climate change is a significant subject discussed by the UNGA. 1992 is the starting point for the analyses; this year, the United Nations Framework Convention on Climate (UNFCCC), aimed to reduce CO2 emissions, was created and signed by 158 states [2]. The Kyoto Protocol operationalised the UNFCCC in 1997. Countries committed to reducing the emissions of greenhouse gasses by (on average) 5% between 2008 and 2012 [3]. The Paris Agreement in 2015 is another example of a United Nations effort to reduce the effects of climate change [4].

The timeline of examples portrays the vital role the UN plays in addressing climate change. This report describes the exploratory data analysis that tries to determine a correlation between a countries' frequency of climate change-related mentions during the UNGA and their actual impact on the environment. The UNGA speeches data gets enriched with climate change data made available by The World Bank to look for this correlation [5]. The main focus of the research is the possible correlation with greenhouse gas emissions. Still, other climate change-related data, for example, renewable energy consumption, will also be analysed.

This analysis is relevant because it tries to find out if countries that seem to be concerned and active on climate change, based on their mentions during the UNGA, are taking measures to prevent climate change. The report first presents the hypotheses and questions to analyse. Then the methodology section explains how the data is cleaned, combined and analysed. After that, the results and discussions section shows and discusses the analyses' outcome before concluding within the conclusion section.

1.1 Hypothesis and Questions

The hypothesis is: "Countries that address climate change more frequently during the UNGA take more measures to prevent climate change". Based on this hypothesis, the following questions are formulated:

- What is the correlation between a country's frequency of climate change-related mentions during the UNGA and their change in CO2 emissions?
- Is there a correlation between the topics discussed in a nation's speech and its sustainability plans? (Do more mentions of sustainable projects also mean a higher production of renewable energy?)
- Which countries mention the most climate change-related topics? What's the correlation between these countries' frequency of climate change-related mentions and their actual impact on the climate?

2 Methodology

2.1 Data collection & pre-processing

The General Debates (GD) during the annual sessions of the UN General Assembly (UNGA) presents an opportunity for leaders to give their government's perspective on major world issues. The GD statements (in English) of UNGA sessions from 1970 to 2020 are textualised into a UN General Debate Corpus (UNGDC) and available for download in the public domain as a 'gzip archive'.

The corpus was downloaded and extracted. The speeches were organised into individual text files, country-wise, under a folder structure where each folder represented an annual session from 1970 to 2020. Python code was used to read the textual data files to prepare for analysis as part of the pre-processing activity. As a next step, checks were made to see if there were any empty data values. Further, exploratory data analysis was performed on the UNGDC dataset. Some of the analysis was to check for the top 100 words (not including the stop words) mentioned in all the speeches, to check mentions of 'Climate', 'Environment', and other words related to Climate Change.

The other dataset that was used for this analysis was the climate change dataset, which The World Bank provided. It is a dataset containing countries from everywhere in the world starting in 1960 till 2020. The dataset covers climate systems, exposure to climate impacts, resilience, greenhouse gas emissions, and energy use per country.

The dataset was not made analytics-ready for the analysis. For example, the years were stored as separate columns instead of one column containing all the years per country, while the topics were stored in separate rows within one column. These were the main problems encountered regarding this dataset, which did not help when merging with the UN debate dataset because there were no common columns to merge it on.

With these problems in mind, our team started with the data cleaning part by transposing the year columns with the individual column, which contained all the climate change topics with their respective values. After transposing the dataset, the years that were stored as separate columns had the string datatype. This was changed by changing the years to an integer type. The dataset was subsetted by having a start date of 1992 till 2018. This was due to the growing concern of climate change during these years. Next to that, most of the countries had empty values in 2019 and 2020. That is why the years 2019 and 2020 were excluded from the cleaned dataset. In the end, the datasets were multi-indexed on the "country" and "year". This concludes our data preparation activities for the climate change dataset.

After cleaning and pre-processing the UN and World Bank datasets, they contain indexes on ISO 3166 country code and year. The datasets are also unique on this index, making it possible to join them using a Pandas join function. After joining the datasets, the duplicate country code and year columns are dropped to make the column names unique again.

2.2 Modelling

To search for the frequency of occurrence of 'climate change' related keywords in the UN speeches, an appropriate model was needed to describe the vocabulary of 'climate change' words and measure the presence of 'climate change' words. A "bag-of-words" model provides an approach to tokenise the words and find out the frequency of each token. The "bag-of-words" model was chosen for implementation as it is simple and provides a lot of flexibility for customising as compared to other models. A "bag-of-words" model was implemented in Python to search for the frequency of 'climate change' related keywords in the UN speeches. Next, the frequency of occurrence of these keywords was used as a feature to train classifiers for further analysis and visual plots. The keywords that were primarily used for classification were: "co2", "carbon", "climate", "global warming", "gas", "green", "sustainable", "Kyoto protocol", "pollution", "emissions", "fossil fuels", and "greenhouse".

The investigation also included identifying a cause and effect relationship between the frequency of climate change keywords in UN speeches and the country's CO2 emissions as per the World Bank report. As there is just one independent variable, Simple Linear Regression was used as a regression analysis method to model the relationship between the independent variable (frequency of climate change keywords) and the dependent variable (CO2 emissions).

3 Results and Discussion

Q1: What is the correlation between a country's frequency of climate change-related mentions during the UNGA and their change in CO2 emissions?

According to the following scatter plots, it is hard to find a correlation between the frequency of climate-related mentions and the CO2 emissions. Most of the data points are grouped around the y-axis, meaning countries that rarely mention the topic of climate change can emit either a lot or nearly no CO2 (Figure 1).

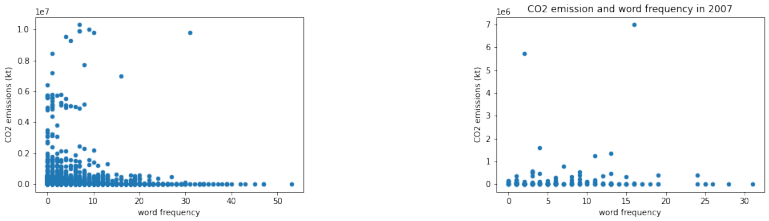

Fig. 1: Scatterplot for the CO2 emissions and word frequency of the UN speeches

To investigate in more depth the correlation, it is subsetted data from 2007 for CO2 emissions. The findings suggested no correlation for the year 2007. As seen on the scatter plots, countries with a low 'word frequency' emit a small amount of CO2. Furthermore, countries that have a high 'word frequency' do not necessarily outperform the ones with a low frequency.

Based on the findings, it is hard to find a correlation between the frequency of climate-related words and the emission of CO2. Many countries barely mention climate change, and yet their emission can be high or low. Overall, no clear correlation could be found between the frequency of climate change-related mentions by a country and their emission of CO2.

Q2: Is there a correlation between the topics discussed in a nation's speech and its sustainability plans?

Based on the correlation plot below, there is not much of a correlation between the word frequency and the renewable electricity output from each country (stored as Renewable electricity output (% of total electricity output)) (Figure 5). The correlation score is -0.17 (rounded) between these two variables.

A scatter plot of the entire dataset shows the distribution of these two variables. From this scatter plot, it cannot be derived if there is a (positive or negative) linear relationship between the two variables (Figure 2).

The vast majority of the countries pick up sustainable initiatives without emphasising them too much during the debates. Interestingly, some of the countries that frequently mention it during the speeches do not produce any form of renewable energy.

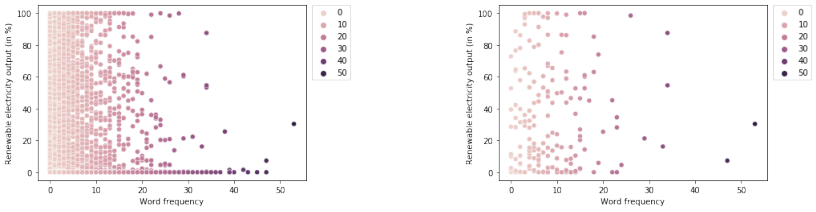

Fig. 2: Scatterplot between word frequency and renewable energy output (in %) on the whole dataset (left) and in 2015 (right)

When subsetting it on the latest year where Renewable electricity output data is available for most countries, in this case, 2015, it does not show much of a pattern (Figure 2).

Based on the findings above, the conclusion is that there is no correlation between the mentions of sustainability problems/plans and their actual action to create renewable energy. Countries that have incorporated sustainability and climate change statements in their speeches do not immediately increase renewable energy output.

Q3: Which country mentions the most climate change-related topics over time?
Sorting countries by the frequency of climate change-related words shows that from 1992 to 2018, the top three countries that mention climate change-related words are Micronesia (542), Tuvalu (502) and Madagascar (490). Micronesia and Tuvalu are island nations. Madagascar is one of the most species-rich countries globally [6]. Therefore, these countries are sensitive to climate change which, in turn, leads to their close attention to climate change-related topics.

The scatterplot shows a negative linear correlation between CO2 emissions (per capita) and word frequency for Tuvalu and a positive linear correlation for Madagascar. There seems to be no clear correlation for Micronesia (Figure 2).

In each linear regression, the dataset was randomly split into a train set and a test set, where the size of the test set is 30% of the total number. The train set was used to fit the linear regression model and the test set to evaluate the model (Figure 4).

Figure 4 indicates a negative linear correlation between CO2 emissions (per capita) and word frequency for Tuvalu and a positive one for Madagascar. As word frequency increases, CO2 emissions (per capita) decrease in Tuvalu. In contrast, the opposite is true in Madagascar.

Then the models were evaluated by R2 and MSE. The R2 scores of the linear regression models for Tuvalu and Madagascar were 0.68 and 0.73, respectively, and the MSEs for both models were close to 0, indicating that the estimate of the coefficient was reasonable.

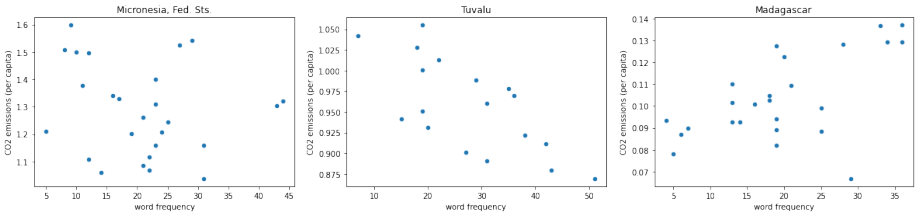

Fig. 3: Scatterplot of CO2 emissions (per capita) vs word frequency for Micronesia (left), Tuvalu (centre), and Madagascar (right)

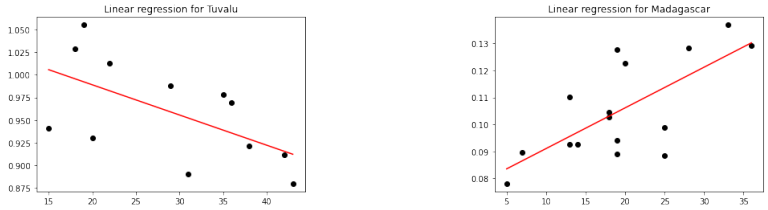

Fig. 4: Linear regression for Tuvalu (left) and Madagascar (right)

3.1 Limitations

Due to the small size of the sample of data, it is hard to draw a sound conclusion for most countries. Further, our estimates and evaluation were very dependent on the split of train and test sets. Finally, the analysis did not consider negative mentions of climate change in speeches, which might have biased some results.

4 Conclusion

This paper combines the UNGDC dataset and the World Bank's climate change dataset to investigate the relationship between the country's mention of climate change-related topics in the annual UN speeches and their actual impact on the environment. The study has shown that as the frequency of mentioning climate change-related keywords in the UN speeches increased, the CO2 and greenhouse gas emissions did not decrease. One of the more significant findings to emerge from this study is that there is no relationship between the mentions of sustainability problems in the UN speeches and countries taking actions to increase the creation of renewable energy. The most prominent finding from this study is that the top three most vocal countries about climate change in the UN speeches are also some of the most impacted due to climate change. However, even in these countries most related to climate change, we can see different measures and results in controlling their impact on the environment. These findings suggest that the hypothesis, "Countries that address climate change more frequently during the UNGA take more measures to prevent climate change", can be rejected.

References

[1] A. Baturo, N. Dasandi, and S. J. Mikhaylov, "Understanding state preferences with text as data: Introducing the un general debate corpus," *Research & Politics*, vol. 4, no. 2, p. 2 053 168 017 712 821, 2017.

[2] *From stockholm to kyoto: A brief history of climate change*, https://www.un.org/en/chronicle/article/stockholm-kyoto-brief-history-climate-change, Accessed: 2021-10-01.

[3] *What is the kyoto protocol?* https://unfccc.int/kyoto_protocol, Accessed: 2021-10-01.

[4] *The paris agreement*, https://www.un.org/en/climatechange/paris-agreement, Accessed: 2021-10-01.

[5] *Climate change - the world bank*, https://data.worldbank.org/topic/climate-change, Accessed: 2021-09-29.

[6] J. U. Ganzhorn, P. P. Lowry, G. E. Schatz, and S. Sommer, "The biodiversity of madagascar: One of the world's hottest hotspots on its way out," *Oryx*, vol. 35, no. 4, pp. 346–348, 2001. DOI: 10.1046/j.1365-3008.2001.00201.x.

Appendix A

Additional plots

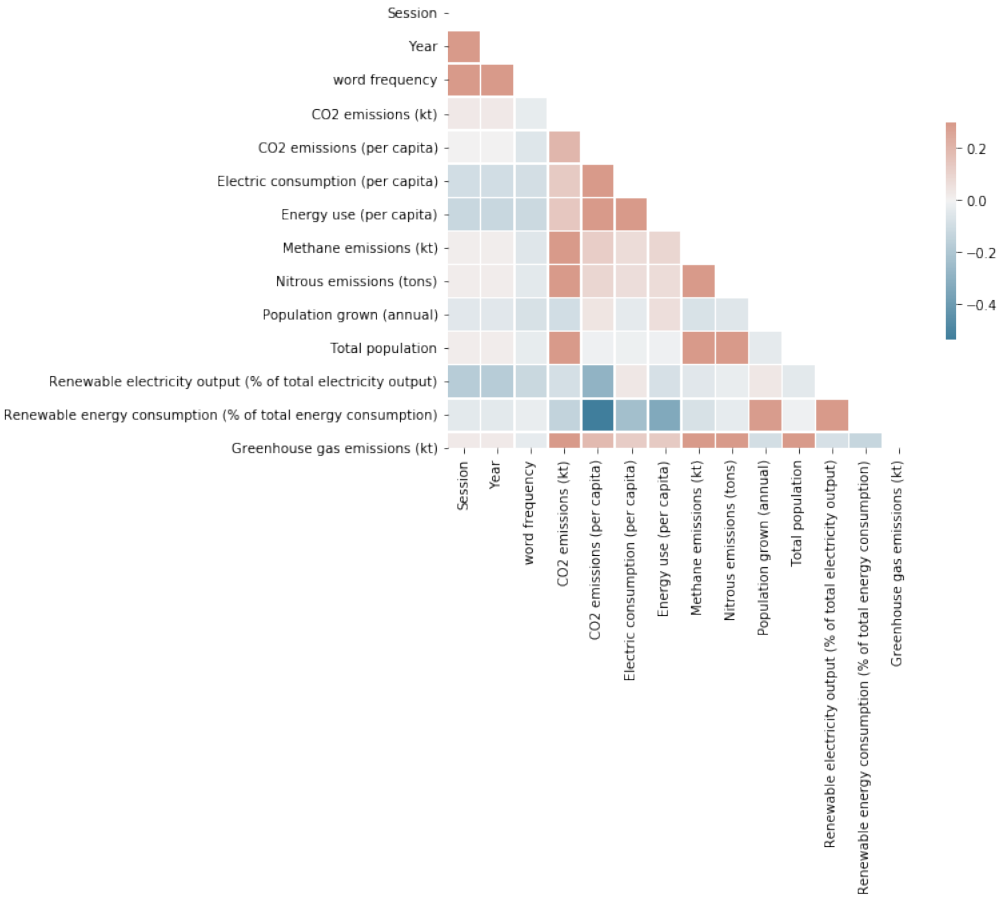

Fig. 5: Correlation plot