Stance Detection on Climate Change Discourses

Augustine Malija

a.malija@mpp.hertie-school.org

Advisor: Slava Jankin

Hertie School

Finn Muller-Hansen

Prof. Dr. Jan Minx

The Mercator Research Institute on Global Commons and Climate Change (MCC)

February 3, 2021

1 Introduction

- Motivation

  The main motivation for writing this thesis is driven by a question it seeks to answer: To understand the divide in stance citizenry take towards policies mitigating climate change.

  The primary reason for choosing this goal is the theory that compared to continental Europe, the controversy in delivery of climate change information to United States citizenry leads for it to be less informed on climate change. It is then intriguing to find out country-wise differences in stances citizens take towards climate change discourses and if this difference is driven by the variation in strictness of legislation.

- Task

  To reach to the goals of this research the following tasks should be implemented

  a) To review state of the art theory about Stance Detection.

  b) To learn and understand Support Vector Machine (SVM) approach to doing Stance Detection Analysis.

  c) Get familiar with twitter data provided by The Mercator Research Institute on Global Commons and Climate Change (MCC) including identifying what countries it comprises of.

  d) To construct the SVM model architecture

  e) To prepare annotations ready to be subjected to training and test sets.

e) Experiment with the model

f) Write the thesis including results obtained from the experimentation stage

2 Related Work

A following synopsis of related papers related this study offers theoretical underpinning and shed light into the progression of this study.

Dilek Kucuk and Fazli Can determine stance on tweets posted about two opponent football clubs in Turkey – Galatasaray and Fenerbahce. They achieve their objective both through the use of explicit names of clubs and implicitly on tweets mentioning respective management or footballers themselves.

The paper is selected as a baseline because of its employment of SVM-based techniques in detecting stance from football tweet data.

One million tweets about popular Turkish Sports clubs – Galatasaray (Target 1) and Fenerbahce (Target 2) – published between August 18 and September 6, 2015 were compiled. Only Favor and Against Stance classes were considered. The Club stood as a stance target either in cases where the tweets mentioned the sports clubs in their entirety or in cases where management or some footballers were praised or criticized. From these annotations, 175 tweets were in favor of Target 1 while another 175 tweets were against it. On the other hand, 175 tweets were in favor of target 2 while another 175 tweets were against it. Therefore, the stance annotated dataset contains 700 tweets.

The target identification criteria comprised tweet features based on unigrams, bigrams, hashtags, external links, emoticons and named entities. Through the use of Natural Language Processing, SVMs for target 1 and two were formed. These SVMs were then trained and tested on the annotated datasets. Two SVM Classifiers – one for each target – using unigrams as features were trained. Prior to the extraction of unigrams, automated pre-processing was used to filter-out stopwords. A SVM implementation employing the SMO algorithm to train a classifier with a linear kernel was used. It provided results of the two classifiers using the metrics of precision, recall and F-Measure. To observe the contribution of hashtags on stance detection, the existence of hashtags in tweets was used as an additional feature to unigrams.

It proposes the use of n-gram based features of SVM classification for a favorable performance on stance detection problems. This is because using bigrams as the only feature leads to poor results. Conversely, unigram features lead to superior results compared to bigram features.

In a study by S.Mo Jang and P. Sol Hart, 5.7 million tweets mentioning either “climate change” or “global warming” emanating from four English speaking countries (US, UK, Canada and Australia) from July 1st 2012 to June 30th 2014 were accessed from Topsy. Topsy is a third party firm licensed by twitter that provides open access to actual tweets and metadata aligned to about 100 million active accounts. Retweets were also included because they show the perception extent for the importance of original tweets. To answer questions on geographical and partisan differences, tweets were filtered based on location and democratic-republican coded states.

A guiding assumption was that any tweet mentioning either "climate change" or "global warming" and "real" or "fact" at the same time was regarded to be engaging in legitimate climate change discussions. This is to identify uniqueness of public rhetoric that represents single themes of a more complex issue.

Top 500 tweets that were retweeted mentioning either climate change or global warming were looked into first. Through the analysis of such tweets, terms and phrases that commonly represented particular themes were gathered. Moreover, a series of word clouds showing the most used words in content was visually arranged so that the size of a word corresponded to the frequency of its appearance in a text. This procedure aided in identifying important keywords that reflect important themes – keywords with less than 1000 tweets per year and ones with too much noise were discarded.

Five climate change themes were selected as elaborated in this list; (1) real themes included the term "real" or "fact" (2) hoax themes included the term "hoax" or "lie" or "fraud" (3) impact themes included the term "impact" or "impacts" or "threat" or "threats", or "consequences", or "effects" or "affect" or "affects" or "disaster". (4) cause themes included the term "cause" or "causes" or "fuel" or "carbon" or CO2 or "human" and (5) action themes included the terms "action" or "act" or "stop" or "fight" or "policy" or "policies".

It found that tweets reflect much of the controversy observed in traditional media. The US presented a higher ratio of real themes than the rest of the countries. The same applied to hoax themes. The US registered a lower ratio of cause themes than the rest. This pattern was commonly observed concerning impact and action themes except that impact themes in the US and Australia were not significantly different.

There are systematic differences between red and blue states. Red states include real and hoax more than blue ones. In contrast, blue states focused more on impact and action than red states. But they are all equal in cause themes. Red states turned out to be using global warming more than blue states. But the use global warming reduced when tweets discussed climate change in terms of its impact.

It appears that all countries used climate change more frequently than they used global warming. Although American users showed preference of using global warming than climate change among the countries. Moreover, all countries were more likely to use the term global warming than climate change when tweets related to hoax themes.

Before diving into data and analysis Cody et al's study is backed by the following theoretical propositions. First is that climate change is anthropogenic – a result of human activity. They posit that most of what the general public learns comes from social media and twitter is a dominant source. They rely on findings that twitter activity positively correlates with proximity to the storm and physical damage. Also that individuals are likely to form close connections and groups on twitter if they have been affected by a natural disaster.

Cody et al collected 1.5 million tweets mentioning the word "climate" at least once from twitter's gardenhose API over the period of 6 years ranging from 14th September to July 14th 2014. They also include retweets to ensure a higher weighting of messages.

Their methodology is underpinned by the application of a hedonometer – an instrument that uses sentiment scores to calculate the level of happiness score for a large volume of text. 10,222 words of the most frequently used English words from four corpora were assigned happiness ratings using Amazon's Mechanical Turk online marketplace.

Each word was rated based on a rating from 1 – least happy to 9 – most happy subject to how the word made the raters feel. This rating was based on the time scales of day, week and month. Words earning scores between 4 and 6 are classified as neutral and thus were omitted from the study because of the ambiguity of describing them. The word climate earned a score of 5.8 and thus was not included at the time of calculating average happiness. For comparison, they also calculate average happiness for 5 related climate keywords.

To keep an eye on tweets including the word climate but not related to climate, a manual coding of a sample of 1500 tweets determined that 93.5 percent of the tweets were about climate change. This offered assurance that the removal of non-climate tweets did not alter results of the overall happiness score.

Cody et al's methodology relates to the Jang et al in the sense that it picked tweets with the word climate to form a dataset. It also included retweets to add on its stock of dataset. It diverges from Jang through offering and alternative approach of applying the hedonometer to calculate average happiness scores of words.

Williams et al's theoretical underpinning emanates from the argument that studying social media reveals social network structures in the context of online debates which occur to be highly polarized. Social network structures shape opinion and behaviour since they tend to have a strong influence on an individual's perception to phenomena like climate change. In this regard, many individual characteristics such as smoking and political opinion are congregated in social networks. Such homophily arises due to preferential connection to similar individuals and from peer influence. Influences such as musical preferences, health related behavior and emotional transfer are the primary triggers for causes of this homophily. Thus, such clustering determines an individual's adoption of a new attitude or behavior. For example, belonging to a certain online group leads to members of such a group having a similar linguistic construct when communicating within the group.

In tandem with Jang, twitter API was used to collect messages between 13th January 2013 and 30th May 2013 that included climate globalwarming climatechange representing tweets that discussed climate change while agw and climaterealists representing users that expressed skeptic views about climate change. Friend and follower connections were collected forming a bulk of 590,608 tweets from 179,180 users. To ensure data quality, 100 tweets for each hashtag were sampled and manually assessed for their relevance to the topic of climate change.

Follower, retweet and mention networks were constructed and analyzed using the NetworkXv1.8.1 module for Python. They were filtered and visualized as directed graphs using the ForceAtlas2 force directed layout algorithm in a sense that closely connected users were located near each other. A panel of three climate science researchers classified the most active users based on their expressed views towards climate change into "sceptic", "activist", "neutral" or "unknown" categories and "ambiguous" for users who the panel did not reach consensus.

High frequency of edges connecting users with similar or differing views was used as a measure for homophily. Furthermore, the Louvian method was used to algorithmically sense user communities – people interacting with each other more frequently than they do with others – based on social network. These community level interactions were visualized using nodes while edges were used to represent residual interactions.

Kim et al's theoretical arguments are clustered around user authority structures and homophily. They communicate that 90 percent of tweets containing @ - a form of referring a specific user – are conversational in context, making it opportune to study conversational aspects of tweeter. They argue that tagging and hashtags play a role of topic and community identification that extends a reach to users that are not connected. Notably, only 1 percent and 9 percent of users are the most active and highly active respectively while the rest 90percent only share very few tweets – making the 10 percent evangelists that reach the mass through hashtags. They go on to posit that homogenous minds tend to tweet among themselves more often compared to their contrasting minds. This has proven so when pro-life and pro-choice members within abortion groups tweeted to each other.

Their dataset comprised of 152,893 English tweets with an acronym IPCC between 17th September and 8th of October 2013 extracted from twitter's API. Author names, usernames and hashtags mentioned in the tweets were filtered from the dataset. Attention was paid to the most active and most frequent tweets in order to shy away the impact of spam tweets and noise that might arise during the course of analysis. To study conversational aspects of the messages, tweets mentioning usernames with the @ sign were extracted. They were converted to a network with Webometric Analyst then analyzed using Force Atlas to arrive at positions of nodes and layouts of the network. Attention was paid to frequently mentioned connections out of which a threshold of ten or more connections determined the most active networks. These were further checked for their relevance to climate change.

Kristin Demetrious's theory acknowledges twitter as a space for opinion leadership, demographically skewed and a platform to open and close public debate. Being a source for opinion leadership, twitter is a living barometer for political sentiment where elite users influence mainstream media that in turn report issues for a wider public. In comparison to other media platforms, twitter is the favorite to the young, urbanized, new users and well educated demographic groups. It outweighs others in terms of media content such as news, global reach and preference for forum/issue discussion. Public debates are amplified by twitter's subtle power to influence the unconscious assimilation of public discourses by readers – intertextuality. This occurs through weight, narratives and discourses carried in specific texts/quotes comprised in tweets.

The study's data comprises a spiked Australian twitter activity for the month of February 2017 scrapping tweets with clean coal as a key term during discussions about the AEFL campaign on clean coal. It also included tweets and retweets that exemplified intertextuality. This data was accessed from Tracking Infrastructure for Social Media Analysis – a project that tracks Australian twitter activity in real time to form a large deposit. Using tableau, the data was profiled and select periods of times were singled either due to high levels of engagement or key events in the course of the AEFL campaign. These revealed hashtags that were used to understand key themes discussing the issue of coal.

3 Data

Using twitter data provided by Mercator Research Institut on Global Commons and Climate Change, stance detection around climate change policy discourses will be done.

4 Methods

- Approach

  The following Null and alternative Hypothesis guide this analysis

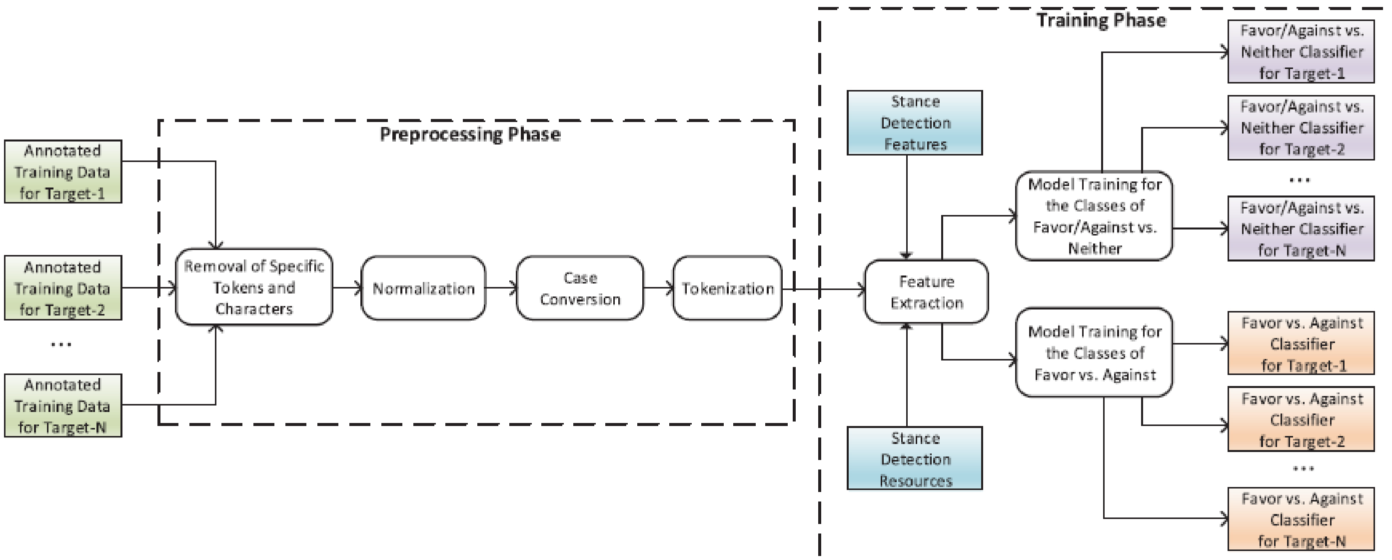
  H0: Better informed citizenry take a supportive stance towards climate change mitigating discourses.

  H1: Less informed citizenry take an opposing stance towards climate change mitigating discourses.

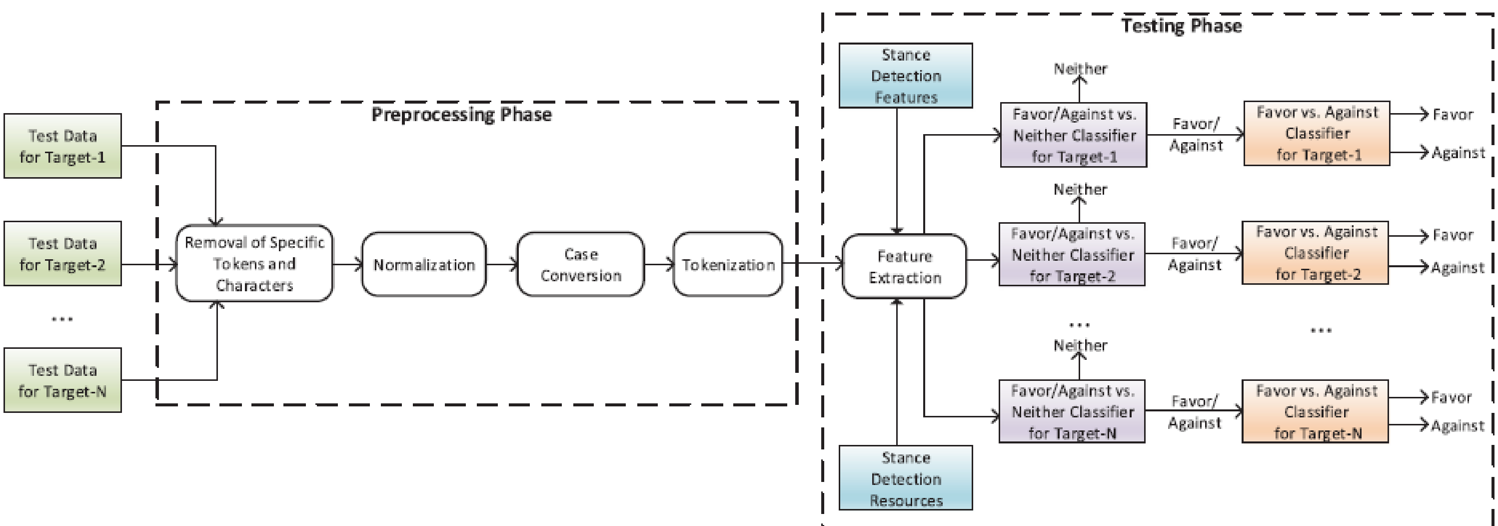
  To address this question, tweet data will be annotated into two stance classes: Favor and Against forming two different data sets. Through the application of Natural Language Processing with NLTK, these datasets will undergo pre-processing phase that includes removal of specific tokens and characters, normalization, case conversion and tokenization. They will then be subjected to training and testing phases as summarized in figure 1.

  The two phases of this pipelined classification scheme are geared towards two objectives. The first phase determines relevancy when inputs are classified as having a stance: (Favor or Against class) or not(Neither class). In the second phase, inputs that were classified as having a stance are further classified as favor or against towards the stance target.

  The explicit stance target will be either "Climate Change" or "Global Warming". This means tweets explicitly mentioning either of the two will be analyzed to see if they are in favor or against climate change. The explicit targets will include climate change influential panels such as "IPCC" and hashtags such as "ClimateAction", "trump", "carbon". It will also



(a) Training phase.



(b) Testing phase.

Figure 1: The Architecture of a Generic Stance Detection System

include top Climate Change influencers on twitter such as "Greta Thunberg", "Jerome Foster II", "Jamie Margolin" and "Leah Namugerwa"

From a tweet such as "Be a ClimateAction superhero" implicitly takes a stance that is in favor of climate change. A tweet going by "Wetlands are a secret power against climate change" explicitly takes a stance in favor of our target - Climate Change. An example of tweet against would be "It is our internal issue, don't speak GretaGoToSchool". This is implicitly takes a stance against climate change and thus will form Target 2.

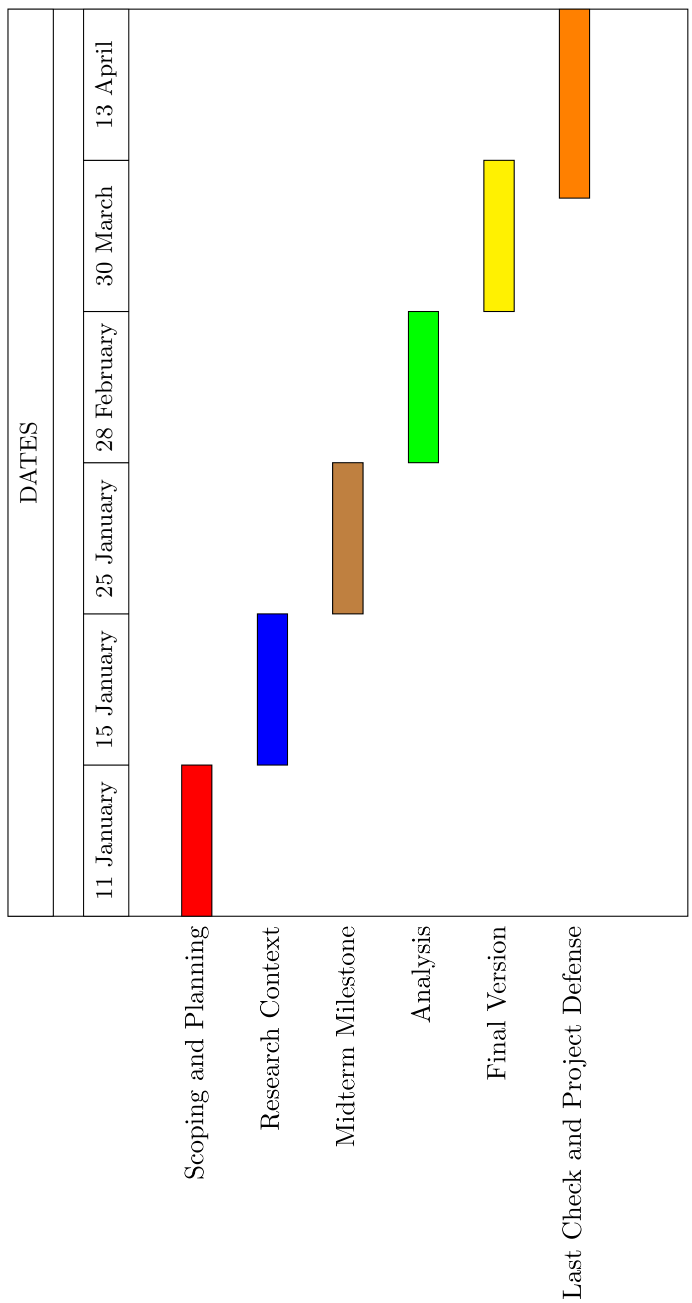
- The Baseline Study

  A study by Kucuck et al that applies the aforementioned SVM stance detection architecture will be used as a baseline.

- Successful Outcome

  SVM analysis of stances around climate change discourses has not been explored by many researchers. Thus, a successful outcome will be setting stage for deploying such an approach to explain phenomena around climate change.

5 Timeline


DATES
11 January
15 January
25 January
28 February
30 March
13 April
Scoping and Planning
Research Context
Midterm Milestone
Analysis
Final Version
Last Check and Project Defense

References

[1] Hywel.T.P Williams, James R. Mc Murray, Tim Kurz, and F.Hugo Lambert. *Network analysis reveals open forums and echo chambers in social media discussions of climate change*. Elsevier,2015

[2] S. Mo Jang and P. Sol Hart. *Polarized frames on "climate change" and "global warming" across countries and states: Evidence from Twitter big data*. Elsevier, 2015

[3] Emily. M. Cody, Andrew. J. Reagan, Lewis Mitchell, Peter Sheridan Dodds and Christopher M. Danforth. *Climate Change Sentiment on Twitter: An Unsolicited Public Opinion Poll*. Plos One, 2015

[4] Warren Pearce, Kim Holmberg, Lina Hellsten and Brigitte Nerlich. *Climate Change on Twitter: Topics, Communities and Conversations about the 2013 IPCC Working Group 1 Report*. Plos One, 2014

[5] Kristin Demetrious. *Twitter and the Struggle to Transform the Object: A Study of Clean Coal in the 2017 Australian Energy Policy Public Debate*. Journal of Public Interest Communications, 2019

[6] Dilek Kucuk and Fazli Can *Stance Detection: A Survey*. ACM Comput. Surv. 53, 1, Article 12 (February 2020), 37 pages.

[7] Dilek Kucuk and Fazli Can *Stance Detection on Tweets: An SVM-based Approach*. arXiv preprint arXiv:1803.08910, 2018