Classifying global warming tweets into subtopics and evaluating the public sentiment

Govindaraj Muthukrishnan
University of California, Berkeley
govin@berkeley.edu

Abstract

Twitter data contains key information that can be extracted and utilized to provide insight into a geographical region's public opinion about global warming. This paper proposes an approach to use Twitter data and provide this information in real-time - which is crucial to devise successful campaigns that educate and push for action. The paper provides a pipeline that incorporates both unsupervised and supervised learning to use an unlabelled Twitter dataset and train a reliable text classifier model. Once the model is able to effectively classify a given tweet into a category, it is then tied into a sentiment analysis model that can produce a score that helps in the understanding of a region's public opinion. The paper also provides motivation and cause for future research in regards to applications and further enhancement.

1 Introduction

Global warming, the long-term heating and shifting of Earth's climate system, is a current and serious problem for human beings that continues to worsen each day. There is a lot of scientific data, research and analysis in the environmental studies surrounding this. However, there is not enough research to effectively analyze the public sentiment of the causes, consequences and responses of climate change.

Two common responses to this are mitigation (preventing additional warming) and adaptation (adjusting society and its norms.) In both responses, a healthy and positive public outlook is necessary to bring proper change and subsequently eradicate this global issue.

All conventional solutions like spreading awareness, practicing eco friendly lifestyles and enforcing environmental legislation necessitate community acceptance of the importance and urgency. This acceptance in turn relies on the understanding and sentiment towards global warming and the subtopics it can be categorized into - for example: rising temperatures, sea levels, carbon emission, etc. Interpreting the current public opinion is very important for devising solutions to combat this universal problem.

We can use the public sentiment to generate successful regionally targeted campaigns - of both global warming and its specific subcategories. For example, a region that strongly believes in global warming is more likely to participate in training programs that teach how to practice safe, eco friendly habits and other ways to get involved and impact change. There may be another region where the community strongly disbelieves in climate change. Here, a more grassroot-level educational initiative may be required before pushing for action.

There is an abundance of Twitter data - tweets in particular - that provide useful information about global warming and climate change. In addition to news updates and article links, we can also find human responses such as likes, comments and retweets that reflect public opinion.

In this paper, 13 million tweets were collected and filtered for climate change related hashtags. Since these tweets are unlabelled - no assigned category - the data was first run through topic modelling, which identified clusters and generated a list of applicable topics. After choosing appropriate labels for the clusters, an effective classifier model was trained using various methods. Finally, the data was passed through Google's Sentiment Analysis API to retrieve the sentiment analysis score.

2 Background

There are few papers that provide research of global warming and it's public perception. However, there is no research of a complete pipeline that uses a dynamic unlabelled dataset. This paper references and builds on research work from a few different works.

Related to global warming, one such paper [1] takes an approach to track climate change opinions from Twitter data. The paper uses SVM and Naive Bayes classifiers that run against data subsets focused on the outcomes of positive vs negative sentiment polarity. Our paper focuses equally on the topic classification as much as the sentiment polarity itself. The dataset used in [1] was manually generated by five people who worked on labeling the Twitter data, choosing the label which at least three people agreed on (i.e., majority label), whereas our paper proposes a method for dynamic topic labelling.

[2] is also a paper related to climate change, that focuses on the sentiment analysis of tweets associated with global warming as a whole (without any descending classification by category.) [3] is a paper that works on improving LDA topic models for microblogs via tweet pooling and automatic labeling. In our paper, we use hyper parametrized LDA in the pipeline as the initial step. [4] presents BERTweet, a novel public model of pre-trained language for English tweets. Although we do not use the model itself, our paper derived upon its method to train and evaluate our classifier.

3 Methods

In this section, we cover the methods followed through the complete process beginning from data collection to the final sentiment analysis of topic-filtered tweets. 1

3.1 Data Extraction

We used dataset [5], which is a list that contains the IDs of a set of Twitter feeds. This list was was collected using the POST statuses/filter method of the Twitter Stream API, using the track parameter with the following keywords: #climatechange, #climatechangeisreal, #actonclimate, #globalwarming, #climatechangehoax,#climatedeniers, #climatechangeisfalse, #globalwarminghoax, #climatechangenotreal, climate change, global warming, climate hoax

Per Twitter's Developer Policy, tweet ids may be publicly shared for academic purposes; tweets may not. Therefore, we had to extract the data for modelling ourselves. Using an approved Twitter developer account, we were able to extract tweets using the tool Hydrator (Documenting the Now. (2020). Hydrator [Computer Software]. Retrieved from https://github.com/docnow/hydrator.) The extraction had to be done in batches because of the daily limit of 86,400 tweets. We collected about 13 million overall tweets.

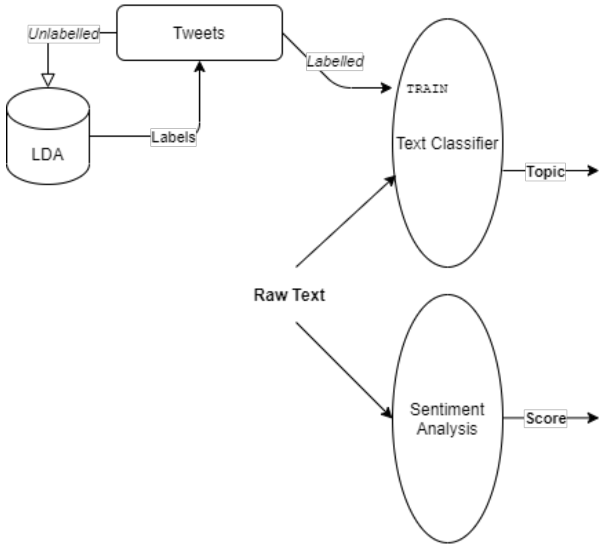

Figure 1: Design of Pipeline

3.2 Text Feature Extraction

Here, we used traditional methods to preprocess the text data. We used NLTK's WordNetLemmatizer to lemmatize and TweetTokenizer and tokenize the data used in topic modelling. For classification, we used the TfidfVectorizer on the text data. The implementation of bigrams did not test well for LDA, however we used it for classification.

3.3 Topic-Modelling

The prepared dataset has no labelling for us to use (and evaluate) for classification models. So we used Latent Dirichlet Allocation, a generative probabilistic model for collections of discrete data such as text corpora.

We passed 3 million processed tweets through the model - bifurcating them into chunks of 100 for faster computation, while also implementing parallel computing. We chose to use coherence and perplexity for evaluation. Perplexity explains how well the model represents the held-out data.

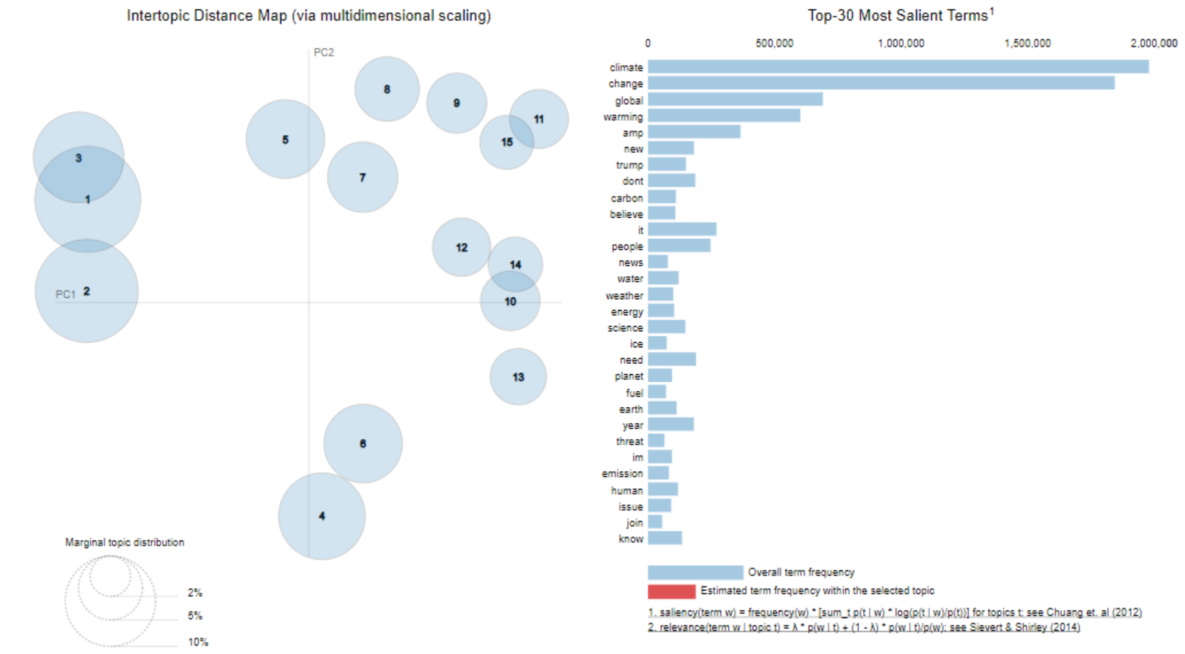

Figure 2: LDA Visualization

Topic coherence measures the degree of semantic similarity between features in a given cluster. In other words, it helps in assessing the human interpretability of generated topics.

After the popular topics were identified and chosen, we calculated the individual contribution of each tweet to each topic. Then we labelled the tweet with the most dominant topic - max of the contributions provided.

3.4 Classifiers

We used Naive Bayes as our baseline model as a text classifier. However, the assumption of independent predictors doesn't hold true for tweets. We then implemented a classifier model using BERT [6] - a transformer-based language representation model. The computational resources required were high for both 3 million and 1 million datasets. We also chose to run models with SVM and Logistic Regression. SVM has been proven to work well with text classification. Unlike Naive Bayes which is a generative classifier, SVM and Logistic Regression are discriminative classifiers. In logistic regression, we extended the max_iterations up to 1000 for the convergence criteria to satisfy. We continued experiments using another approach with Random forest - an ensemble-based learning algorithm that makes predictions by averaging over the predictions of several independent base models (decorrelated decision trees.)

3.5 Sentiment Analysis

Sentiment analysis is a technique that is used to elucidate opinions and subjective information. There are many well working pretrained models in existence through tremendous research. We have used Google's famous sentiment analysis API to help us identify a particular tweet's sentiment score and magnitude.

As Google explains, the score indicates the overall emotion of the text input while the magnitude indicates how much emotional content is present. Therefore, these values from tweets can assign the weight of opinions about specific global warming topics in the regions the tweets come from.

4 Results and discussion

4.1 Latent Dirichlet allocation

Originally, generated topic labels from LDA were not comprehensible from the clusters rendered. This was found to be due to retweets that got iden-

	Validation_Set	Topics	Alpha	Beta	Coherence
207	100% Corpus	16	asymmetric	0.61	0.423071
237	100% Corpus	19	asymmetric	0.61	0.418818
206	100% Corpus	16	asymmetric	0.31	0.415708
217	100% Corpus	19	0.31	0.61	0.415051

Figure 3: LDA Parameter Tuning

tified as unique tweets - which resulted in multiple text duplicates, subsequently skewing the individual topic contributions. After implementing an additional preprocessing step following feature extraction (against only the text column), we saw better generated topic labels as shown in the figure below. Topics that we decided to label the clusters. (using human judgement from the generated topic list)

```
ClimateChangeIsReal/FightClimateChange
Action/Efforts/Awareness
Destruction/Consequences
Belief/Sentiment
Arctic/Icecap
Energy/Emission/Carbon
PolarBear/Wildlife
Politics/Policy/Law
xxForeign
Paris Agreement
News/Media
```

We observed a coherence score of 0.355 and perplexity of -13.16. We tried to improve upon this score using hyper parametrization on the subsets of 3 million. Tuning 3 million required weeks for completion. So, we assessed the results on different 1 million subsets taken from the original. As seen in Figure 3, we observed fluctuating highest values around 0.453

4.2 Classifiers

In our baseline model, Naive Bayes performed with a 0.519 accuracy. We originally expected BERT to produce the best results in this text classification. However, results for the 1 million dataset came to be about 0.537. The computation time and resource utilization was too high even for 1 million to run subsequent tuning. We observed that SVM and Logistic Regression yielded significantly better results. Similar to BERT, Random Forest also did not yield good enough performance metrics compared to SVM and Logistic Regression.

Models and their accuracies:

```
LinearSVC                 0.745236
LogisticRegression        0.795301
MultinomialNB             0.519288
RandomForestClassifier    0.259242
```

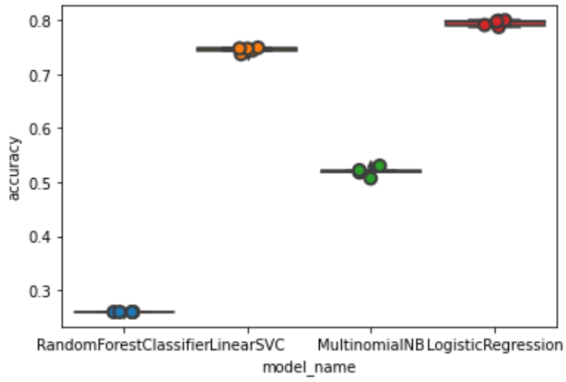

Figure 4: Text Classifier Models

	text	Sentiment Score	Sentiment Magnitude
0	#FEMA has no provision for slow moving disaste...	-0.8	0.8
1	RT @ClimateReality: Fossil fuels are dirty, da...	-0.8	0.8
2	RT @me_elfs: Really fascinating stuff @Benjami...	0.9	0.9
3	@dreamgirlbrooke global warming have a nice ev...	0.7	0.7
4	Professional Practice Guidelines for designing...	0.5	0.5

Figure 5: Sentiment Analysis

Upon further tuning of the Logistic Regression model, we were able to achieve a highest F1 score of 0.827259 and accuracy 0.829130

In our design, we are now able to predict the topics of a tweet. We can then pass it on to the Sentiment Analysis model which will provide us with the score and magnitude metrics. During our experiments, we observe the Google Sentiment Analysis API works quite well.

5 Conclusion

This paper provides a proof-of-concept for a full flowing pipeline. With this implementation, we can create applications that use this model as a back end component. It can be designed such that a single tweet (that we filter in the Twitter API for clauses related to climate change) gets correctly categorized to a topic and sentiment accordingly scored. Information from multiple tweets can be used to populate a geographic heat map by global warming topics - using the coordinates or user_location attributes of the extracted tweet.

The classifier accuracy results are not as high as we had intended to achieve. The tweet character limitation to 280 seems like it may have contributed to the limited results. It is worth researching the training of a classifier model with a larger corpus and only using tweets to evaluate.

Another major challenge is the fact that there is no clearly labelled dataset available. Despite this

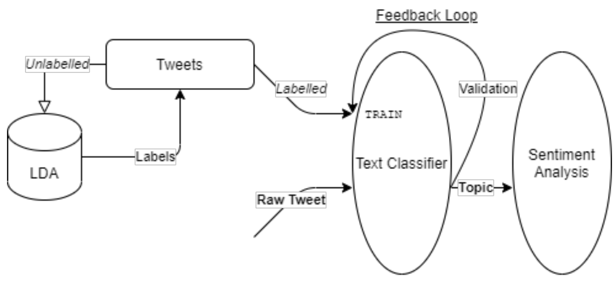

Figure 6: Future Research

pitfall, we were able to produce reasonable results. We can expect to further enhance this model by introducing a feedback loop that uses validation of the output labels and subsequently retrains the model. (as shown in Figure 6)

References

Xiaoran An, R. Auroop Ganguly, Yi Fang, B. Steven Scyphers, M. Ann Hunter, and G. Jennifer Dy. 2014. Tracking climate change opinions from twitter data. *In Proceedings of the Workshop on Data Science for Social Good Held in Conjunction with KDD 2014.* ACM, New York, NY

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., Passonneau, R. (2011). Sentiment analysis of Twitter data. *In Proc. WLSM-11s.*

R. Mehrotra, S. Sanner, W. Buntine, and L. Xie. Improving lda topic models for microblogs via tweet pooling and automatic labeling. *In SIGIR, 2013.*

Dat Quoc Nguyen, Thanh Vu, and Anh Tuan Nguyen. 2020a. Bertweet: A pre-trained language model for english tweets. *arXiv preprint arXiv:2005.10200*

Littman, Justin; Wrubel, Laura, 2019, Climate Change Tweets Ids, *https://doi.org/10.7910/DVN/5QCCUU, Harvard Dataverse, V1*

Blei, D. M., Ng, A. Y., Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research, 3(Jan)*, 993-1022.

Indra, S. T., Wikarsa, L., Turang, R. (2016, October). Using logistic regression method to classify tweets into the selected topics. *In 2016 International Conference on Advanced Computer Science and Information Systems* (ICACSIS) (pp. 385-390). IEEE.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

Devlin, J., Chang, M. W., Lee, K., Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805.*

Kirasich, K., Smith, T., Sadler, B. (2018). Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets. *SMU Data Science Review, 1(3)*, 9.

Jónsson, E., Stolee, J. (2015). An evaluation of topic modelling techniques for twitter.

Abdelwahab, O., Bahgat, M., Lowrance, C. J., Elmaghraby, A. (2015, December). Effect of training set size on SVM and Naive Bayes for Twitter sentiment analysis. *In 2015 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)* (pp. 46-51). IEEE.