Using Tweets During a Natural Disaster to Gauge Public Sentiment on Climate Change

Alexa Tipton[1], Nicholas Van Nostrand[2], Nick West[3], Zhenning Yang[4]

***Abstract*— The purpose of this project is to analyze the influence of natural disasters, specifically hurricanes, on public opinion of climate change. Using data sets containing millions of tweets, sentiment will be classified using a developed machine learning model. These Tweets are collected with a streaming API using select keywords to filter relevant messages.**

I. INTRODUCTION

CURENT, Center for Ultra-Wide-Area Resilient Electric Energy Transmission Networks, is a National Science Foundation Engineering Research Center. The Twitter project originated in 2017. Using an enormous data set of tweets, work is done on analysing people's reactions to natural disasters. In addition, machine learning techniques are applied in novel ways to glean information from natural language. These data are used to gauge public sentiment toward government, utility companies, and media. The purpose of this project is to use these data to instead learn about public sentiment toward climate change, and to see how much natural disasters influence it, if at all.

II. DATA

A. Methods of Collection

One part of the CURENT project mentioned prior is sourcing the tweets that will then be analyzed. There are several methods in-place for gathering data: A web crawler to read tweets the same way a web browser would, Twitter's REST API, and Twitter's streaming API. Twitter's REST API is straightforward and yields a significant amount of metadata, but it is prohibitively slow due to rate limits. On the other hand, the web crawler has none of the aforementioned limitations, but it only has access to the data that is visible to users of the website. Therefore, for this project we will be focusing on the tweets obtained from the streaming API.

For this project, we will be using CURENT's collection of tweets obtained via the streaming API that contain climate change keywords. This collection is quite sizeable, having roughly 45 million tweets that were obtained in the time range spanning from about a year ago to today. We may also use CURENT's tweet collection that focuses on electric power keywords. This set is much larger, with almost 100 million tweets, but it is also much noisier, with previous analyses done with Doc2Vec suggesting that anywhere from 65% to 95% of the tweets may be irrelevant to electric power.

B. Restrictions and Challenges

The streaming API enables the project to track certain keywords or users, and obtain tweets within that tracked subset in real time. A stream can obtain up to 1% of tweets being sent at any given time. That does not sound like very many, but on average that is roughly 60 tweets per second, so the rate limit is negligible. Additionally, the tweets are in the same format as those from the REST API, so metadata is preserved. One downside to this method is that tracking keywords that are too vague can result in considerable noise. For instance, tracking tweets with the phrase "Python" may yield tweets about snakes, or tweets about programming. In fact, one goal of the project will be to separate relevant and irrelevant tweets, which we will discuss further below. Another limitation is Twitter's terms of service, which state that users of the API are forbidden from individually identifying tweets in such a way that they can be traced back to their author. In truth, this is not a major concern; our project does not depend on the public identity of any users, and this restriction would apply for other methods as well.

III. RESPONSIBILITIES

A. Alexa Tipton

Alexa's responsibilities will include writing documentation, evaluating data, and creating and testing machine learning classification models.

B. Nicholas Van Nostrand

Nicholas's responsibilities will include writing documentation, evaluating data, and using machine learning classification models.

C. Nick West

Nick will be responsible for gathering tweets, writing documentation, and using machine learning classification models.

D. Zhenning Yang

Zhenning will develop and test machine learning models, gather tweets, write documentation, and evaluate data.

*This work is supported by CURENT.

[1]A. Tipton is a CURENT employee and student in the Computer Science Department, the University of Tennessee, Knoxville, 37996

[2]N. Van Nostrand is a student with the Computer Science Department, the University of Tennessee, Knoxville, 37996

[3]N. West is a CURENT employee and student with the Computer Science Department, the University of Tennessee, Knoxville, 37996

[3]Z. Yang is a CURENT employee and student with the Computer Science Department, the University of Tennessee, Knoxville, 37996

IV. MILESTONES

- Collecting relevant tweets must be done first before any further work can be done. This should be completed, or at least well underway, within the first week of work.
- Tweets should then be sorted by relevance. Since the tweets will be gathered using specific keyword and time period parameters, it should not be necessary to sort between relevant and irrelevant tweets. However, tweets will be categorized by relevance to natural disasters, relevance to climate change, or both. A sample will be classified, and based on how many tweets belong in each category, it will be decided what topics merit further investigation. This should be decided within the following week once there are a sufficient amount of tweets to review.
- Once the topic of research is fully cemented, machine learning models will be created and trained. This will be the most lengthy step. Creation will likely take weeks, and as many tweets as possible will need to be manually classified in order to have a useful training set. The aim will be to have at least 400 manually classified tweets.
- After the model is created and trained, it will need to be cross-validated to ensure good results. This can be done manually and also with software, and both techniques will be used to ensure a useful result. This step will be performed often, then deferring back to the previous step ad nauseum until results are satisfactorily high. This will be a running task throughout testing and maintenance of the project.

V. EXPECTED OUTCOMES

- Collect a large set of tweets relating to climate change and natural disasters.
- Create at least one machine learning model to perform sentiment analysis on climate change tweets.
- Answer the research questions: Do natural disasters influence public opinion on climate change? If so, how do opinions change before, during, and after natural disasters?

REFERENCES

[1] https://curent.utk.edu/