UNIVERSITY OF WATERLOO
Faculty of Mathematics

SENTIMENT ANALYSIS AND CLIMATE CHANGE

University of Waterloo
Waterloo, Ontario

Prepared by
Vincent Talbot
3B Applied Mathematics
ID 20481891
December 17, 2015

92 Keating Drive
Elora, Ontario
N0B 1S0

December 17, 2015

Dr. Chris Bauch, Professor
Dept. of Applied Mathematics
University of Waterloo
200 University Avenue West
Waterloo, Ontario
Canada N2L 3G1

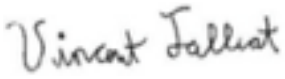
Dear Dr. Bauch:

This report entitled, “Sentiment Analysis And Climate Change”, is my fourth report and was written for my 3B academic term. The purpose of this report is to analyze the effects that major climate change events have on climate change opinion through sentiment analysis.

As a research assistant at the University of Waterloo, I was tasked to create programs to run simulations and analyze data to provide insight on the effects of major climate change events.

The Faculty of Mathematics requests that you evaluate this report. The report and your evaluation will be submitted to the Math Undergrad Office for evaluation. This report was written entirely by me and has not received any previous academic credit. I would like to thank Mrs. Kathleen Talbot for proofreading my report and Dr. Chris Bauch for providing insight and guidance for my report. I would also like to thank Mr. Demetri Pananos for providing guidance on machine learning algorithms. I received no other assistance.

Sincerely,



Vincent Talbot
ID 20481891

Table of Contents

List of Figures

Summary

This report entitled , “Sentiment Analysis And Climate Change”, analyzes the effects that major climate change events have on climate change opinion. By applying sentiment analysis to tweets regarding climate change, public perception of global warming can be determined, analyzed and correlated with current events. Insight into these major events could positively aid in the battle against climate change.

The conclusions drawn from this report include:

- Further research and analysis should be done to determine the full extent of how major climate change events effect individual opinion.
- Both weekly and monthly cycles occur in the daily number of Pro and Anti climate change tweeters.
- The Pope’s climate change encyclical is the most significant event to raise awareness in the measured time period.

1.0 Introduction

Climate Change is an important area of discussion and research in the world today. It involves the change of weather patterns for an extended period of time. These changes include the shift of regular patterns and a difference in the frequency of extreme weather events. It can be caused by biotic factors, plate tectonics, solar radiation, the oceans and volcanic eruptions. The existence of anthropogenic climate change has been a highly contested topic in past years.

1.1 Climate Change

By studying factors such as vegetation, ice cores, sea level and more, scientists can determine previous climate conditions and changes over time. These studies have led the majority of the scientific community to believe that anthropogenic climate change exists. The steep rise of greenhouse gas output of humans since the industrial revolution, the rapidly expanding world population and depletion of natural resources are some of the main reasons why humans are blamed for climate change. These human activities correlate with the rise of global temperatures, sea level, ocean temperatures, ocean acidification and frequency of extreme weather events. Scientists believe that humans have a major impact in climate change and must change their ways to stop it. (Crowley, 2000) In this paper these scientists and people who share the same beliefs will be referred to as Pro.

Though climate change is widely believed to exist, there still are people who do not accept its validity. These people call into question the legitimacy of the conclusions that many scientists have determined. They believe that scientists skew data and distort facts to fabricate global

warming. Many hold the belief that the world's climate is not changing, and any variations noticed in weather patterns are just natural fluctuations. There are also people who believe the climate is changing, but that humans have no effect on this process. They profess that the Earth has experienced fluctuations in climate throughout its long history, and any changes that humans see is just symptoms of this process. Other groups of deniers may not believe in climate change for religious reasons. The group of people who disbelieve climate change will be referred to as Anti.

1.2 Machine Learning and Sentiment Analysis

With the improvements of computer technology, new methods to study and analyze public opinion have become available. Machine learning is based on the study of algorithms that allow computers to learn from and formulate predictions about data. During this process, the computer takes sample inputs and constructs a model around them to help create informed decisions or predictions. The formulation involves computational statistics and optimization to maximize efficiency and accuracy. Machine learning can generally be divided into three major categories: supervised learning, unsupervised learning and reinforcement learning. This paper will focus on support vector machines (SVMs), which is a type of supervised learning.

Supervised machine learning is when the computer is provided with sample input vectors and their corresponding outputs. The computer learns from these examples and creates a function to map any new inputs to their complementary outputs. SVMs will use sample inputs that are assigned to two possible outcomes, and appoint new data to one of these two categories. Multi

Class SVMs also exist, allowing for more than two identifier classes. There are many applications for supervised machine learning and SVMs, but this paper focuses on its use for sentiment analysis.

Sentiment analysis is the use of text analysis and language processing to extract intuitive information from a source. More specifically, it can be used to identify the tone, opinion or sentiment of a piece of text. For example, it can be used to identify whether a review is positive or negative. An et al. concluded that applying sentiment analysis to social media provides viable insights into citizens' opinions on climate change. By comparing data from traditional survey techniques and data mining from Twitter, they found that the information gathered from both sources are equally reliable and credible. It is acknowledged that Twitter users may not represent all social groups, but with its rising popularity, it includes an extremely wide variety of these groups. An et al. also concluded that there is a correlation between short-term fluctuations of negative sentiment tweets and major climate events. These major events can cause a sudden change in sentiment polarity, but since they looked at sentiment as a whole and not on an individual basis, there is great uncertainty in this overall sentiment.

Cody et al. applied sentiment analysis to twitter as well. By rating keywords in the tweets on a happiness scale of one to nine, they determined the overall happiness of each tweet. They found that tweets regarding climate change are generally less happy than average tweets. They also found that climate change activists are more prevalent on Twitter than deniers, leading to the conclusion that most Twitter users agree with the scientific consent that climate change is real.

1.3 Purpose

The purpose of this paper is to use sentiment analysis to analyze tweets regarding the climate change opinions of Twitter users. The plan is to look at how major climate change events effect the number of climate tweets and if these events are successful in swaying the opinions of individual people. By comparing different types of events such as natural disasters, weather patterns, political events and other occurrences, it may be possible to determine which type of events have the largest and lasting impact on people.

2.0 Methods

The first step in this project was to collect data. Using the Twitter API, a computer program collected and stored data about tweets that contained any of the keywords “climate change”, “global warming” or “warming planet”. Data collection spanned the timeframe of April 7, 2015, to October 29, 2015. The data included the tweet, the date and time of the tweet, the number of retweets, the user’s screen name, the user’s id, the location of the user and the number of friends and followers of the user. Approximately six million tweets were collected during this time period, with any non-english tweets being filtered out.

The second step involved the creation of a machine learning algorithm to classify the tweets. There were three classifications of tweets: Pro, Anti and Neutral. The Pro class included tweets for people who believe in the global warming issue and are raising awareness. The Anti class contained the people who believe global warming is false, or is not a pressing issue. The Neutral class included tweets that had little to do with global warming. For example, the algorithm would

catch a tweet that said, “We need to change the climate of our workplace.” This tweet and others with no direct connection to global warming would be included in this class.

The machine learning algorithm used was SVMs. By manually searching through a few thousand tweets, 464 Pro, Anti and Neutral tweets were found for the training set. Due to the similarity of the tweets, it was sometimes difficult to discern between Anti and Pro tweets, especially due to the use of sarcasm. Despite this, the program had a success rate of approximately 80 to 85%. To test this, the program was trained with a random 70% of the training set and then tested on the other 30% multiple times. Each training set always had an equal amount of Pro, Anti and Neutral tweets. Over the testing period, the program correctly guessed the sentiment of unique tweets 80 to 85% of the time. After testing was completed, the program took in the millions of tweets that were collected and returned the sentiment of each tweet.

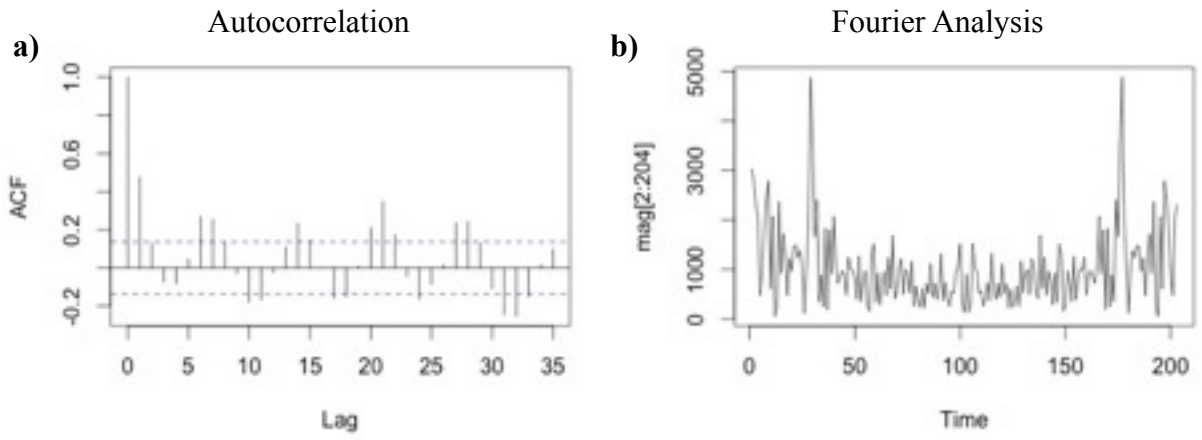
The next step was to determine which events were causing the daily fluctuations of Pro and Anti global warming tweets. Multiple news websites such as Global News, CBC news, Democracy Now and others were used to determine which events might be causing influxes in tweets. These events included extreme weather conditions, natural disasters, political debates, business deals, charity events and more. The daily tweets were examined as well to confirm which events in a day caused the most substantial number of tweets. Through this process, a rough timeline of climate change events was created to compare with the Tweets Time Series (**Appendix A**).

3.0 Analysis

The tweets were run through the program and assigned the identifier Pro, Anti or Neutral. The amount of each type per day was graphed accordingly. (**Appendix A**) Some notable dates include May 28, June 18, June 23, August 3 and August 7. These were the dates that Barak Obama answered climate questions on twitter, the Pope released his climate change encyclical, various extreme weather events occurred globally, severe wildfires began in California and the republican presidential candidates had their first debate, respectively. The date the Pope released his climate change encyclical coincides with the largest number of climate change tweets in a single day.

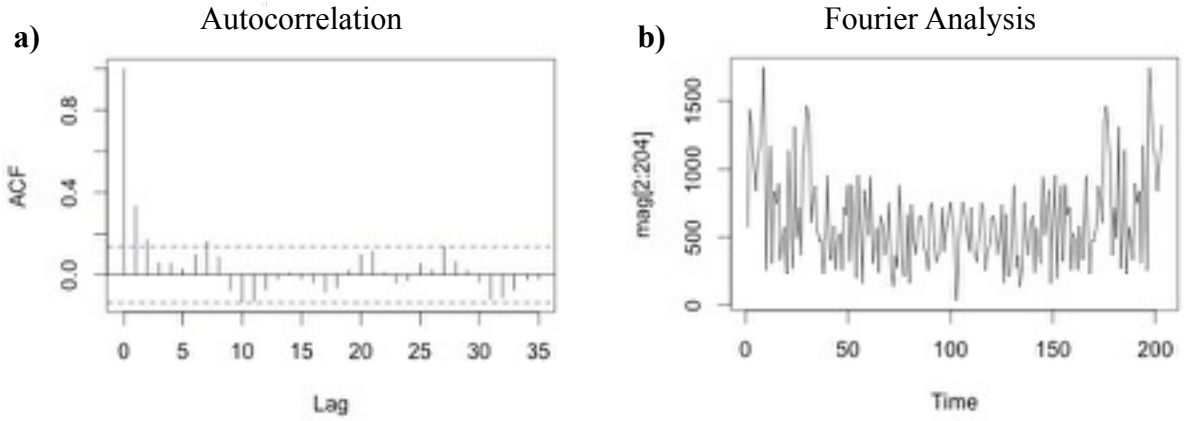
To find trends, significant spikes and cycles in the data, autocorrelation analysis and Fourier analysis were applied to the data. In **Figure 1**, the autocorrelation analysis of the Pros show that there is significant correlation on a weekly basis, and the Fourier analysis shows there may be a monthly cycle that repeats itself. **Figure 2** shows that the same may be said for the Antis, but these relations are less significant.

Figure 1 - Pro Analysis



a) is the autocorrelation analysis of the Pros and b) is the Fourier analysis of the Pros

Figure 2 - Anti Analysis



a) is the autocorrelation analysis of the Antis and b) is the Fourier analysis of the Antis

The coefficient of variation was calculated to be 41% for the Pros, and 58% for the Antis. This shows that the Pros have a lower degree of variance than the Antis and suggest that the number of Pro tweets does not fluctuate as much as the number of Anti tweets, which is unintuitive by looking at the time series.

4.0 Discussion

Further research needs to be done to understand the extent of how major climate change events affects the general population's opinion on climate change. It is clear that these events can raise the number of climate change tweets and awareness, but the scope of this impact is still unknown. Supplementary examination is needed to determine which type of event causes the greatest impact. Do natural disasters or political events raise more awareness? Early research shows that these events have the potential to sway people's opinion of climate change. If that is true, it is important to establish which events change people's minds, do Antis or Pros switch more often and how long do they stay switched for. These are some of the questions that remain

to be answered. The continuation of research in this area may lead to significant insights and conclusions which will positively aid in the battle against climate change.

5.0 Conclusions

Further research and analysis should be done to determine the full extent of how major climate change events effect individual opinion.

Both weekly and monthly cycles occur in the daily number of Pro and Anti climate change tweeters.

The Pope's climate change encyclical is the most significant event to raise awareness in the measured time period.

References

An, X., Ganguly, A. R., Fang, Y., Scyphers, S. B., Hunter, A. M., & Dy, J. G. (2014). Tracking climate change opinions from Twitter data. In *Workshop on Data Science for Social Good*.

Boussalis, C., & Coan, T. G. (2013, September). Signals of doubt: text-mining climate skepticism. In *Workshop", London School of Economics* (Vol. 27, p. 28).

Cody, E. M., Reagan, A. J., Mitchell, L., Dodds, P. S., & Danforth, C. M. (2015). Climate change sentiment on Twitter: An unsolicited public opinion poll. *arXiv preprint arXiv: 1505.03804*.

Crowley, T. J. (2000). Causes of climate change over the past 1000 years. *Science*, *289*(5477), 270-277.

O'Connor, R. E., Bord, R. J., & Fisher, A. (1999). Risk perceptions, general environmental beliefs, and willingness to address climate change. *Risk analysis*, *19*(3), 461-471.

Appendix A - Tweets Time Series

This is a graph of the number of tweets each day for 206 days. The x-axis is number of days and the y-axis is number of tweets. The red represents the Pro tweets, blue represents the Anti tweets and black represents the Neutral tweets. The starting date is April 7, 2015, and the ending date is October 29, 2015.