CSC110 Final Project: Pathos and the Pandemic — The Spread of Misinformation Along With COVID-19

Bryson Lo, Helena Sabina Glowacki, Emma King Wai Ho, Chin Chin Jim

Tuesday, December 13, 2021

Information for Whom It May Concern

Please note our use of the scikit learn library, which seems to have problems installing on m1-chip Macbooks (apologies to any TA grading this with an m1 macbook!)

UTsend Info:

Claim ID: sFrHbPV9F6DXrSB4

Claim Passcode: ycVt4nMDjqdzyFyw

Problem Description and Research Question

Fake news has been present long before the invention of social media; however, the recent rise of social media and unverified new sources immensely increases our exposure to fake news. A study done in 2018 by 3 scholars from MIT showed that fake news spreads substantially faster than real news on Twitter, a social media platform commonly used today (Dizikes, 2018). After certain catastrophic events such as natural disasters and national tragedies, fake news has the tendency to appear as forms of propaganda, fear-mongering, and control of public opinion. Shortly after the beginning of COVID-19, misinformation about the pandemic was spread, only exacerbated by the prevalence of social media communication in our current society. As harmless as some rumours can be, others have the potential to create dangerous consequences to public health and order. Our research question is **What factors of a tweet contribute to the spread of misinformation during the COVID-19 Pandemic?** We chose to investigate this topic because the rise of misinformation regarding the COVID-19 that appeared through social media and other mediums is alarming. This is problematic because fake news is misleading and causes widespread panic. The spread of fake news likely contributes to the public making choices that are harmful to their personal interest and society as whole. According to an article published by BBC News in May 2020, many people died because of drinking alcohol-based cleaning products – for example methanol – believing online claims that it would kill the virus (Coleman, 2020). Information about the efficacy and intentions behind vaccine roll outs also caused mass hesitation among the public, and many civilians today are still not vaccinated due to these fears. Evidently, the spread of misinformation can be dangerous to public health and safety. Certain groups of the population seem to be more vulnerable to the guise of fake news. The goal of our group's project is to investigate the factors in the spread of misinformation, and how they relate to its victims. We will be looking at data obtained from 10000 true tweets and 10000 fake tweets from Twitter in the year 2020 (which includes keywords, emotions, topic, etc.) to train a machine learning model that classifies whether or not a certain social media post is fake.

Our group has decided to create this model using data on keywords and emotions of tweets instead of our original predictors of age and education on top of other social media factors. Furthermore, as per TA's advice, rather than classifying social media or news as a whole, we narrowed down our population to one social media platform, Twitter, in the year 2020.

Dataset Description

Dataset 1 Link: https://doi.org/10.3886/E120321V11 For this investigation, we are planning to use a dataset from `OpenICPSR`. We will make adjustments to the size and format of the data to make it more usable for our code. This dataset contains the tweet_id, user_id, tweettimestamp, keyword, country/region, valence_intensity, fear_intensity, anger_intensity, happiness_intensity, sadness_intensity, sentiment, emotion of each tweet. In our model, we are using these variables

Dataset 2 Link: https://www.kaggle.com/arashnic/covid19fakenews We will also be using the 2020 Twitter observations of the Kaggle dataset provided by user Möbius, a data scientist working in the field of healthcare. This dataset includes the link, title, tweet_id, users' engagement, and whether each tweet is real or fake news. Out of this dataset, we are using the tweet_id to help classify whether observations of our `OpenICPSR` dataset are real or fake news.

Computational Overview

To begin, we noticed that our `OpenICPSR` data set only had the tweet id, not the tweet itself. To fix this, in **tweet_scraping.py**, we used **Twitter API** and the **Tweepy** library to gather the tweets' content and create a new csv with tweet ids and their respective tweets. Next, we added a new variable to our data set called "tweet_content" and renamed the data set to "emotions.csv", which can be downloaded at our UTSend Drop off.

As the `COVID-19 Fake News Data` from Kaggle was separated in many .csv files separating real and fake, and tweets and articles, in textbftweet_scraping.py, using the **Glob** and **Pandas** libraries we ran through all the files and combined the real and fake tweets to create a new clean data frame.

Like before, the new `COVID-19 Fake News Data frame`did not contain the tweet content, only the user id. So similarly to the `OpenICPSR` data set, in **tweet_scraping.py** we used Twitter API and the **Tweepy** library to fetch the content of the tweets. Next, in **tweet_scraping.py**, we filtered the data frame so that it only contained successfully fetched tweets and made a new .csv file called **kaggle.csv** with the the original data frame(with information about whether or not the dataset is real or fake) with a new variable of tweet content.

With all our data formatted and ready, we trained 7 emotion attribute classifiers with the Twitter Dataset with Latent Topics from OpenICSPR. We trained one classifier for each emotion attribute in the dataset, using the text content as input and the categories of each given emotion attribute as output.

We began by shuffling the data and split 80% of the observations for training and 20% for testing. Then, we computed a CountVector, which stores the number of instances each unique word appears in each tweet across all of our training data. Next, we use the TfidfTransformer provided by the sci-kit learn library to return a sparse matrix that gives us information about the weighting of importance of words that appear in our tweet contents. Using this sparse matrix, we finally use it as input to fit a Support Vector Machine model to our desired output categories of each given emotion attribute. To optimize the accuracy of each model, we used GridSearchCV, a class provided by sci-kit learn that allows us to train models with a combination of many different parameters and obtain the best performing one. After optimization, our emotion attribute classifiers had an average of 73% accuracy on testing data.

We then used these trained `emotion_attribute_classifiers` to classify emotion attributes of the tweets from the `COVID-19 Fake News Dataset` from Kaggle and generated 'valence_intensity', 'fear_intensity', 'anger_intensity', 'happiness_intensity', 'sadness_intensity', 'emotion' and 'sentiment' values for each tweet in the dataset. We then again shuffled the data and split 80% of the observations for training and 20% for testing. As we have knowledge of whether each tweet of the `COVID-19 Fake News Dataset` is true or false, we used the predicted emotion attributes as input and used whether the content is true or false as the targeted output to train a logistic regression model which would classify whether tweets are real or fake, also returning a percentage of how sure the function is about its prediction. Similarly to before, we also optimized this logistic regression model using GridSearchCV and obtained a final optimized accuracy of 67% on testing data.

In order to make the use of our trained classifier more user-friendly, we created `user_classifier.py`. The `UserClassifier` class is used in the **main.py**. The user either inputs the content of one tweet as a string or the content of multiple tweets as a list of strings. If the user chooses to input one tweet, the `single_predict` method will run and return a dictionary containing two pieces of information, the first being a prediction of whether or not the tweet is real or fake and second being the probability of its prediction. If the user chooses to input multiple tweets, the `batch_predict` method returns a **Numpy** array, which contains a list of predictions of whether or not the tweet text is fake or real.

Instructions for obtaining data sets and running your program

Our **main.py** file contains 5 steps. The first 3 steps are commented out.

Our first step creates our processed data files. Before running our first step to check data processing. All our datasets can be downloaded from links in the Dataset Description section.

Please make sure that all .csv files downloaded from Kaggle are stored in a folder named 'kaggle_data' on the same directory level of all python files. Please also move the .csv file downloaded from OpenICSPR titled

'COVID19_twitter_full_dataset' to the same location (same directory level) as where the python files are stored. Please also make sure that 'emotion.csv' and 'kaggle.csv' are not in the same location as the python files when running step 1 as step 1 will create these 2 files.

As mentioned previously, we used the given data sets, `COVID19 Twitter Dataset with Latent Topics`, `Sentiments and Emotions Attributes` from OpenICSPR and `COVID19 Fake News Data` from Kaggle, to create **emotion.csv** and **kaggle.csv** respectively. These new data sets are samples of their original data sets, but with a new variable added: tweet_content.

To run steps 2 and 3, which trains our classifiers and stores them in a file, simply uncomment them once 'emotion.csv' and 'kaggle.csv' are created. Please note that step 2 will take around 1 hour to run as the training and optimization of classifiers take around 10-15 minutes each and we have 7 classifiers to train. Alternatively, the files can be downloaded from UTsend as shown below.

UTsend Info:

Claim ID: sFrHbPV9F6DXrSB4

Claim Passcode: ycVt4nMDjqdzyFyw

To obtain our processed datasets and trained models, please visit UTSend and download four files: emotion.csv, kaggle.csv, emotion_classifiers.zip and rf_classifier.joblib. Before running our program, please unzip the zipped emotion_classifiers.zip file and store its contents in a folder named 'emotion_classifiers' at the same directory level of all python files. Please also move the 3 remaining files emotion.csv, kaggle.csv and rf_classifier.joblib to the same location (same directory level) as where the python files are stored.

Steps 4 and 5 allow our user to run this program, upon running **main.py** in the python console, the program will ask the user:

"Would you like to test a single tweet or a list of tweets? (single/list)"

For the program to run efficiently the user should input "single" or "list". Otherwise, the program will return "User did not enter correct input. Expected input 'single' or 'list'. Please try again."

If the user inputs **single**, the program will ask the user to "Please enter your tweet", and return the prediction results.

If the user inputs **list**, the program will ask the user to "Please enter your first tweet (If there are no more tweets enter 'end'):". The program will continue to ask the user to enter a tweet until the user enters 'end'. Once the user has entered 'end', the program will return the predictions for the batch. If the user submits 'end' the first time the user is asked, however, the program will print this message: "There were no tweets to test. Please try again".

Description of changes we made to our project plan

Our original plan was to investigate the factors in the spread of misinformation and how they relate to victims. We were planning to look at data of education levels and age groups of users, and social media factors to determine how they relate to fake news. In our original plan, we would use **scikitlearn** library to train binary classifiers that will classify real and fake news according to this set of selected variables. Although our core idea of classifying between real and fake news remains unchanged, we took the advice our TA gave us on our proposal and went in a different direction than planned. Instead of focusing on news categories and how misinformation is spread, we decided to focus on intensities, emotions and sentiments as variables and use these variables to create a classifier that determines whether each tweet is true or not. We also focused on one social media platform, twitter, and retrieved the content of the tweets using **API**. These changes added complexity to our code as we had to do more than calling methods from machine learning libraries. We also believe this produces a more useful algorithm as our program classifies the validity of the tweet solely from its content and disregards how other people react and perceive it.

Discussion of the results of our program

Our program takes in the content of tweet(s) as inputs and returns either a dictionary of whether the tweet is true or false based on our classifier, and the probability of that outcome, as we are aware that our classifier is not 100% correct. This is useful in everyday life as our emotions hugely affect what we do and what we say on social media platforms. Therefore, it would be useful to have this function that helps us determine to an extent what we see from these platforms, as a lot of news we obtain from it is fake. Projects similar to this one can be done as part of data analysis on public health and awareness studies, for example, misinformation regarding other topics such as vaccines, mental health, and social discrimination.

Because of the human nature of emotions and sentiments, as well as meaning lost in translation when it comes to cultural expressions of language, it is difficult to create any kind of computer program that can analyze these topics

with 100% accuracy. Firstly, placing emotion to text computationally cannot be fully accurate, as human judgement would probably be the most efficient way to identify emotions. Some idioms and underlying messages can also only be recognized by a person. Thus, we lose some accuracy when placing emotions onto the string contents of our tweet collection. Furthermore, emotions can be used when presenting both false and true information. Passion, as well as both negative and positive sentiments being present in a true statement is very possible, as the topic of the pandemic itself can be extremely emotional. Thus, it is hard to fully tie emotion with lack of accuracy in a statement. Additionally, it is challenging to identify sarcasm or other tones of voice through computations. For example, what if a tweet was "subtweeting" and quoting a separate false fact? We were aware that our project would not be able to provide full certainty on the emotions and truthfulness of a text. We developed an approach that would help us overcome these barriers to the best of our abilities within the scope of this course.

Some limitations we encountered along the way is that we had to combine the data from `COVID19 Fake News Data` from kaggle and `COVID19 Twitter Dataset with Latent Topics, Sentiments and Emotions Attributes` from OpenICSPR, as neither dataset contains both emotion intensities and whether the tweet is real or fake. Therefore, we needed both datasets to train our classifier instead of using just one.

Fake news is present in almost all subject matters we encounter on social media. Going forward, we hope to use the same idea to train programs that classify tweets on other topics, in addition to COVID19. We could also improve our algorithm by having it trained on other datasets that use information from other social media platforms, for example Facebook and Instagram.

Our project has sought to provide insight into the use of emotion and sentiment in relation to misinformation, and how false facts often pander to these human feelings in order to gain credibility. Our results confirmed our worrying suspicions — that emotion can, and often is, used in connection with false information. We cannot be sure whether this is because people who wish to spread false facts utilize manipulation tactics that cater to the human mind, or because people who have a lot of fake news to share simply do so with much emotion. At the very least, with this in mind, we hope to be more intentional in the future when receiving information, and aim to think critically — What is the source of this information? What is the broader implication of these facts? Are there any biases in how the information was collected, analyzed, and distributed? Hopefully, with projects similar to ours, further awareness can be spread on the dangers of misinformation and how to prevent its rise.

References

arashnic "Möbius". (n.d.). COVID-19 Fake News Dataset: Help researchers to identify fake contents. Melbourne. Retrieved November 2, 2021, from https://www.kaggle.com/arashnic/covid19-fake-news.

Coleman, A. (2020, August 12). *'Hundreds dead' because of Covid-19 misinformation.* BBC News. Retrieved November 4, 2021, from https://www.bbc.com/news/world-53755067.

Dizikes, P. (2018, March 8). *Study: On Twitter, false news travels faster than true stories.* MIT News — Massachusetts Institute of Technology. Retrieved November 4, 2021, from https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308.

Garneau, K., & Zossou, C. (2021, February 2). Misinformation during the COVID-19 pandemic. Ottawa. Retrieved November 2, 2021, from https://www150.statcan.gc.ca/n1/pub/45-28-0001/2021001/article/00003-eng.htm.

Gupta, R., Vishwanath, A., & Yang, Y. (2021, November 4). COVID-19 Twitter Dataset with Latent Topics, Sentiments and Emotions Attributes. Ann Arbor, MI. Retrieved November 17, 2021, from https://doi.org/10.3886/E1 20321V11.