

Detecting COVID-19 Misinformation

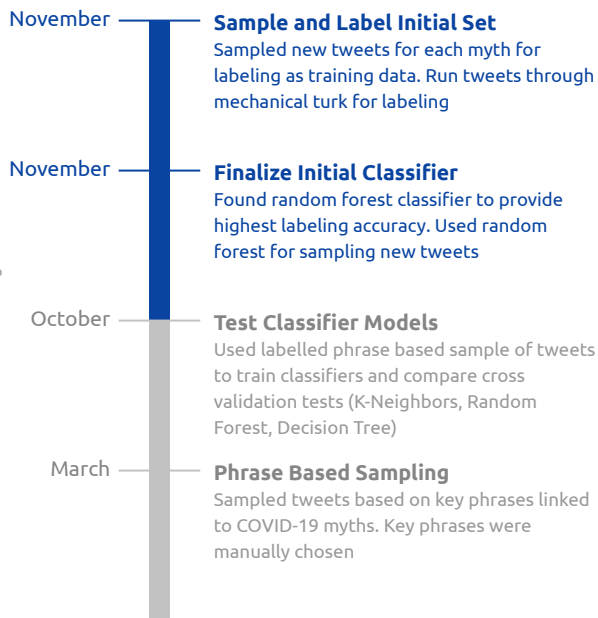
Alexander Chen, December 10, 2020

Overview

- Use tweets related to COVID-19 from 2020 to develop iterative machine learning approach to gather high quality data for training a high quality classifier to detect popular myths in tweets

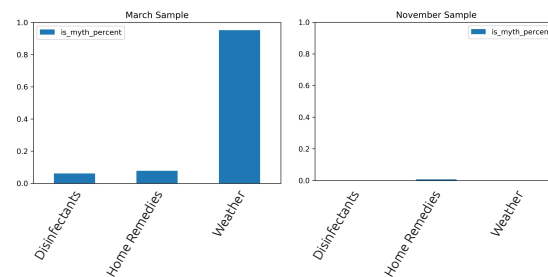
Myths Analyzed with Examples

- **Disinfectants**
 - "I see that QAnon accounts are recommending that people drink bleach to avoid the #coronavirus. As a PhD-holding public bear speaking on behalf of the @USER01 I can 100% promise you this: if you drink enough bleach, you absolutely WILL NOT have to worry about viruses. At all. #Q"
- **Home Remedies**
 - "Can regularly rinsing your nose with saline help prevent infection with the new #coronavirus? #Covid_19 #virus #COVID19italia #covid19UK #Covid19usa #Covid19fr #WuhanCoronavirus #2019nCoV #Wuhan"
- **Weather**
 - "Will Warmer Weather Stop the Spread of the #Coronavirus? Don't Count on It, Say Experts #USA #Trump #Republicans URL Removed"



Results

- Tweets were sampled poorly, leading to nearly 100% "no" labels for each myth. Methodology needs to be improved



Next Steps

- Adjust tweet sampling methodology to require a minimum confidence threshold
- Check future random sampling before labeling for general accuracy
- Collect new sampled tweets for training and improving existing classifiers
- Consider similarity metric to identify high quality tweets for labeling

Ground Truthing for Detection of Covid-19 Misinformation

Alexander Chen, May 10, 2021

Overview

- > Use tweets related to COVID-19 from 2020 to develop iterative, efficient machine learning approach to gather high quality data for training a high quality classifier to detect popular myths in tweets
- > **CURRENT STEP:** Create an approach that can build a weak classifier for sampling relevant tweets for a high quality classifier

Myths Analyzed

- > [Disinfectants](#)
 - o + : 98, - : 691
- > [Home Remedies](#)
 - o + : 105, - : 684
- > [Weather](#)
 - o + : 457, - : 471
- > [\(New\) Masks](#)
 - o + : 177, - : 63
- > [\(New\) Origins](#)
 - o + : 196, - : 54





Myth Comparison

Disinfectants

"Drinking/swallowing disinfectants can kill coronavirus"

Weather

"The effects of hot or cold weather on preventing the spread of coronavirus"

 Phrase Based Sampling	<ul style="list-style-type: none">• Examples: "bleach", "drink disinfectant"• Low Accuracy: 6% (POS: 27, NEG: 413)	<ul style="list-style-type: none">• Examples: "warm weather", "heat kills"• High Accuracy: 95% (POS: 417, NEG: 23)
 ML Sampling	<ul style="list-style-type: none">• Trained classifiers with oversampled phrase based samples• Higher Accuracy: 35% (POS: 71, NEG: 128)	<ul style="list-style-type: none">• Trained classifiers with phrase based samples• Lower Accuracy: 17% (POS: 40, NEG: 198)
 Takeaways	<ul style="list-style-type: none">• Phrases typically lead to high precision, but not always• ML algorithms typically lead to higher coverage, but lose precision• Hybrid approaches are needed to generate high quality ground truth data for noisy domains like social media	
 Next Steps	<ul style="list-style-type: none">• Augment phrase based sampling method by exploring embedding spaces• Continue training new classifiers to build towards a weak classifier for sampling tweets• Apply approach onto other myths to test generalizability and scalability of process	