Comparing Models on COVID-19 Tweets

Eliana Mugar & Sam Rigor

Najoung's Feedback

Sounds interesting! A few concerns that would need to be addressed in the final report:

- What exactly is your novel contribution here? e.g., what would you find that you wouldn't find with running an experiment with a random subset of the COV19Tweets dataset? I'm assuming the diversity in opinion, source/user, and timeframe would be what differentiates it, but I am not sure how this ties into the overarching goal that you describe in the final paragraph.

- How exactly would you go about filtering the dataset to 500-1000 tweets, meeting the constraints that you described? The process and design decisions should be well documented in the final report.

Background

- In recent years, younger people have been using social media as a news source
- Though social media can make information accessible to a wide audience,

Previous Work

- In recent years, researchers have already developed datasets to train ML models to detect misinformation
  - “ANTi-Vax: a novel Twitter dataset for COVID-19 vaccine misinformation detection”
  - “CoAID: COVID-19 Healthcare Misinformation Dataset”

Task Description & Metrics

Are models able to correctly identify whether or not social media posts contain misinformation regarding COVID-19?

- Models used (both from HuggingFace):
  - spencer-gable-cook/COVID-19_Misinformation_Detector
  - roupenminassian/TwHIN-BERT-Misinformation-Classifier
- Both of these models use binary classification
  - **0** = NOT misinformation
  - **1** = misinformation
- That being said, we can use accuracy, precision, recall, and F1 to evaluate model performance

Dataset Description

- Tweets mentioning COVID-19 in some way
  - Example terms: “nCov,” “Pfizer,” “vaccine,” “corona”
- Existing datasets are too large → randomly select N=1,000 tweets from the data

Data Sources

COVID-19 Tweet datasets with different scales

nanyy1025/covid_fake_news

label (string)
"fake"
"real"
"fake"
"real"

arbml/COVID_19_Disinformation_ar

q6_label (string)
"no_not_harmful"
"yes_panic"
"no_not_harmful"
"yes_rumor_conspiracy"
"yes_xenophobic_racist_prejudices_or_hate_speech"
"yes_rumor_conspiracy"

justinqbui/covid_fact_checked_polifact

```
true -> true
mostly-true -> true
half-true -> misleading
barely-true -> misleading
false -> false
pants-fire -> false
full-flop -> false
```

Limitations

- There are multiple COVID-19 Tweet datasets
  - However, they are often large— hard to load into Colab and Jupyter Notebook
- Annotations may be incorrect— need to find a viable way to hand-label whether the Tweet is misinformation or not

Bibliography

https://www.sciencedirect.com/science/article/pii/S0033350621004534

https://arxiv.org/abs/2006.00885

https://www.kaggle.com/datasets/gpreda/covid19-tweets

https://huggingface.co/datasets/arbml/COVID_19_Disinformation_ar

https://huggingface.co/datasets/justinqbui/covid_fact_checked_polifact

https://huggingface.co/datasets/nanyy1025/covid_fake_news

https://huggingface.co/spencer-gable-cook/COVID-19_Misinformation_Detector

https://huggingface.co/ChrisEsworthy/Covid_Misinformation_Model