INTRODUCTION

- Properties of humor
  1. Producer + perceiver
  2. Positive emotional reactions (laughter)
  3. Highly individualistic & cultural specific

- Lack of data annotated with humor

- Our goal
  - Generate reliable labels of perceived humor
  - Without the need for extra human annotations

- Facebook reactions

- Covid-related Facebook posts
  - Keywords: covid-19, coronavirus, corona, covid 19, sars-cov-2, covid, sars cov 2
  - Language: English
  - Post type: text-only
  - 2M posts retrieved, 785K posts after cleaning

DATA COLLECTION

- Examples of a humorous post and a non-humorous post

Users tend to respond to humor with 😂 reactions
A high percentage of 😞 reactions towards a certain post
↓
HUMOR

DEFINING HUMOR SCORE (HS)

- Posts with higher haha percentage -> more humorous
- Discount unpopular post
- Humor Score (HS)
  - The percentage of haha reaction, with a popularity stretcher
  \[ HS = \frac{h}{t} \times \tanh\left(\frac{t}{50}\right) \]
  
  \( h \) = # of haha reactions, \( t \) = total # of reactions, 50 = popularity stretcher

DEFINING NON-HUMOR SCORE (NS)

- Need to retrieve negative samples for binary humor detection
- Posts with the lowest HS are too sad/unlaughable
- We want not only sad posts but also general non-humor posts
- Posts with high Non-humor Score (NS)
  - Posts whose reaction distributions are closest to an average Facebook post
  \[ NS = -\log(\tanh(\frac{t}{50})) + \sum_{r \in R} \left(\frac{S(r) - O(r)}{|R|}\right)^2 \]
  
  \( R \) = the set of Facebook reactions, \( S = \% \) of reaction \( r \) in the standard distribution, \( O = \% \) of reaction \( r \) in the observed post,
  \( t \) = total # of reactions, 50 = popularity stretcher

HUMOR ANALYSIS

- Our humorous posts have
  - Human centeredness: singular first-person pronouns, total pronouns
  - Negative polarity: anger words, negations, negative sentiment
  - Less detailed and more abstract writing style
  - Lower complexity

EXPERIMENTS

- Experiment settings
  - Continuous: HS is used as ground truth of humor
  - Binary: Positive - high HS posts; Negative - high NS posts
- Models
  - RoBERTa-base
  - BERTweet: RoBERTa + Tweet
  - BERTweet-covid: BERTweet + 23M COVID-related Tweets
- Human labels
  - Used not as gold standard, but as a baseline

FUTURE WORK

- Retrieve general humorous posts without topic constraints
- CHoRaL: Collecting Human Reaction Labels
- Reactions to other emotions in the post (sad, angry)