CHoRaL: Collecting Humor Reaction Labels from Millions of Social Media Users

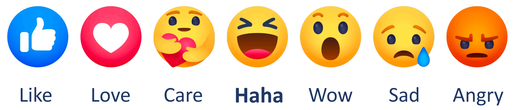
Zixiaofan Yang, Shayan Hooshmand, Julia Hirschberg
Columbia University

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



Like Love Care **Haha** Wow Sad Angry

DATA COLLECTION

- 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

# of Posts	784,965
# of Poster Accounts	264,685
# of User Reactions	126,839,984
# of Haha Reactions	6,525,247

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

Got kicked out from the hospital because I told the Covid patients to stay positive.	295 3
Too many families have lost a loved one to COVID-19. This is a disease that doesn't discriminate — we need to come together & meet this moment.	253 39 34 10 7 4 1

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

$$\mathrm{HS} = \frac{h}{t} * \tanh(\frac{t}{50})$$

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

$$\mathrm{NS} = -\log(\tanh(\frac{t}{50}) * \sum_{r \in R} \frac{(S(r) - O(r))^2}{|R|})$$

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
- Humorous posts have more emojis in general
  - Top 1 humorous emoji: ``Face with Tears of Joy'' 😂😂😂
  - Humorous posts have fewer heart emojis 💜💙💚💛, but more 💔

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

	Continuous		Binary	
	F1	AUC	F1	AUC
Human	-	-	0.867	-
RoBERTa	0.869	0.939	0.868	0.937
BERTweet	0.879	0.947	0.881	0.950
BERTweet-covid	0.880	0.948	**0.883**	**0.951**

FUTURE WORK

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