

I'm out of breath from laughing! I think?

A dataset of COVID-19 Humour

Anonymous ACL submission

Abstract

Humor is a cognitive construct which predominantly evokes the feeling of mirth. The growing distress and agony during COVID-19 pandemic reemphasised the need to study humor and its role in our daily lives. In this paper, we present a well curated dataset of COVID-19 humor with rich and reliable annotation. Our dataset comprises over 4K reddit posts, memes and comments, annotated with labels namely *humor type*, *topic*, *maladaptive vs adaptive*, and the *target* of the humor.

1 Introduction

Humor is a cognitive construct which predominantly evokes the feeling of mirth. When combined with the right style of argument and addressed at the right audience, humour acts as social behavior regulator (Zekavat, 2021; Olah and Ford, 2021). Through humour, one can understand the emerging worldviews, evolving beliefs, and adjustments in socio-cultural conventions. Researchers have extensively studied the role of humor in diffusing stress (Hussein and Aljamili, 2020; Lemish and Elias, 2020), encouraging solidarity (Amici, 2020) and information sharing (Dynel, 2021) during COVID-19. Moreover, COVID-19 humour targeting certain ethnic and gender also revealed the widespread use of humour in masking online hate speech (Pérez, 2021). Through this work, we present a dataset of COVID-19 humor to enable research on societal behavior during the pandemic. Our main contribution is:

- a well curated dataset¹ of COVID-19 humor annotated with the type of humor (such as *pun*, *sarcasm* etc.), the topic (such as *illness*, *family* etc.), the target of humor (such as *women*, *elderly* etc.). We also labelled each post with inherent stereotypes if any.

¹link will be added upon acceptance

When you are laughing at all the corona memes but the laughing suddenly turns into coughing



Figure 1: A meme on COVID-19 pandemic

The rest of the paper is organized as follows. Sec. 2 discusses the prior work on humour during COVID-19. We describe our approach behind dataset creation in Sec. 3. We also provide a baseline model for humour detection and maladaptive humor detection in Sec. 3.3. We conclude our work in Sec. 4

2 Related Work

Hussein and Aljamili (2020) found humorous content such as caricatures and memes mocking the pandemic and unusual situations arising out it, helpful in relieving stress. Olah and Ford (2021) showed that people with self-enhancing humor style often engage in protective behaviors such as *wearing a mask*, *washing hands* during COVID19. Akram et al. (2021) found humor to be crucial in resolving feelings of anxiety on social media.

Bischetti et al. (2020) discovered that the appreciation of COVID-19 humour vary depending on demographic factors and personality traits. They further detected high aversion towards COVID-19 humor when compared to the usual humor. The increased aversion was contribution to dark themes including *death* and *elderly people*. Vurayai (2020) highlights the need to tackle disparaging COVID-19 jokes targeting women. Dynel (2021) performed multimodal analysis on a corpus of COVID-19 mask memes to track the evolving *voices* in the

Source	#posts	Example
<i>r/Jokes</i>	806	The World Health Organization has declared that dogs cannot transmit Coronavirus, and there is no reason to quarantine dogs anymore. W.H.O. let the dogs out.
<i>r/Onion</i>	103	Parents Can't Tell If Pandemic Inhibited Toddler's Social Skills Or If He's Just Taking After Dad
<i>r/NottheOnion</i>	543	Convict refuses parole, says jail safer than outside in pandemic
Reddit Comments	130	I'm out of breath from laughing!I think
Twitter	15	My son hasn't been able to go out since he was 13 because of the pandemic. He's a quaran-teen.
Total (Textual Humor)	1588	
<i>r/CoronavirusMemes</i>	909	Refer Fig. 1
Total (Textual + Memes)	2497	
Non Humorous	1563	Facing a COVID Crisis, Nepal Cries Out for Help
Total Samples	4069	

Table 1: An overview of collected data

highly shared content on social media.

Torres et al. (2020) collected 107 COVID-19 humor scripts each for English and Tagalog. These scripts were coded with the humor type (e.g. *self-deprecating humor*, *malicious or sexual humor*, *aggressive etc.*), humor targets (e.g. *people*, *countries*, *biblical characters etc.*), humor subjects (e.g. *appearance*, *food*, *country*, *date etc.*) and humor structure (e.g. *conversational*, *knock-knock*, *lyrical*, *narrative*, *one-liner*, *question and answer*).

Overall, there is an interesting collection of work on COVID-19 humor highlighting its varied purposes. (Torres et al., 2020) provided 107 humorous scripts in English with labels such as *humor subject* and *humor type* however, their analysis was qualitative and the size of sample is very small. Through this work, we provide a rich, well-curated dataset for COVID-19 humor in English.

3 Humor in COVID-19

3.1 Data Collection

The subreddits *r/Jokes*, *r/CoronavirusMemes*, *r/Onion* headlines and *r/NottheOnion* headlines were chosen to extract humorous posts. In order to extract only COVID-19 related jokes, keywords namely *pandemic*, *covid19*, *coronavirus*, *lockdown*, *quarantine*, *covid*, *corona*, *SARS* were used within the subreddit thread. We observed that the comments under humorous posts often tend to be humorous themselves. We thus also extracted these comments. PRAW (Python Reddit API Wrapper)

was used to extract the posts from Reddit. A total of 67 humorous Twitter accounts (Meaney et al., 2021) were queried using the above mentioned COVID-19 keywords to extract COVID-19 humorous tweets.

A total of 1563 non-humorous yet COVID-19 related posts, were collected from subreddits namely *r/Coronavirus*, *r/COVID19*, *r/Chinaflu*, *r/Pandemic*. The overview of the posts from different subreddit threads is provided in Tab. 1.

3.1.1 Data Reliability

A text is considered humorous if the intent of the text is to be humorous. Since we only considered reddit threads known for their humorous content, it is safe to assume that the collected data is humorous. Internet memes are often humorous piece of culture (Davies, 2008) and we thus assume all collected memes to be humorous. To analyse the quality of the collected textual data, we asked a team of human annotators to manually tag randomly selected 30 posts from each thread that are, *r/Jokes*, *r/Onion*, *r/NottheOnion* and the comments.

Critchley (2011) defined humor as jokes (spoken or written words) and actions which elicit laughter or generate mirth. Human annotators were asked to independently determine if a given post is humorous or not while keeping this definition in mind and the intent of the author. The inter-annotator agreement (Krippendorff's alpha) was found to be 0.932 for *r/Jokes*, 0.512 for *r/Onion*, 0.20 for

Topic (#posts)	<i>Example</i>	Target (#posts)	<i>Example</i>
Pandemic (487)	Coronavirus ruining your plans for 2020? Save them for 2022!. Cause 2022 is 2020 too.	General (625)	Why can't ants get coronavirus. Because they have little anty-bodies
Illness (152)	In a year, we're all going to remember this pandemic and laugh about it...at least the ones left...	Authority (149)	Trump wasn't lying about the coronavirus disappearing in April. He simply forgot to say 2021.
Life Changes (130)	It all started with a bat. Then toilet paper. Now we're going nuts in quarantine.. We really have gone bat, shit, crazy.	Anti-maskers/vaxxers (75)	It's primarily killing the unvaccinated. The vast majority of which are idiots.
Political & WHO (103)	Coronavirus isn't Trumps Fault. Ebola wasn't Obama's Fault. Sars wasn't Brush's fault..... And only a handful of cases of herpes were Clinton's fault.	Infected (66)	Don't worry, there are experts who are trained to deal with the coronavirus.. We call them coroners.
Love/sex (86)	I really hope coronavirus can't spread through sex. It would be so lonely being the last man on Earth.	Working Class (53)	Coronavirus. I bet the guy who invented hand sanitizer is rubbing his hands together now.
Pop culture (56)	Why did Princess Leia contract coronavirus?. Because she went to woo Han.	Couples (47)	This quarantine thing is not working out.. My wife and I are becoming friends and I almost told her about my girlfriend.
Boredom (56)	I finished 5 books during the quarantine. That's a lot of coloring.	Celebrities (37)	what do you call an artist during a global pandemic? . Vincent Van Cough
Groceries (51)	I went for an interview, they asked me where do I see myself in five years. Apparently, still using the toilet rolls I stockpiled during the Coronavirus outbreak is not the right answer	Women (32)	Japanese mayor says men should grocery shop during pandemic as women 'take a longer time'
Physical Appearance (48)	Covid lockdown has me gaining weight so I started a new diet. Its called Two weeks to flatten Your curves	Children (31)	This pandemic has given me a lot of time to get to know my kids better..or worse..
Media (39)	Fox News Limits Pandemic Coverage To Avoid Giving Coronavirus Notoriety It Craves	China (31)	What's the difference between kung pao and the coronavirus?. One's Chinese take out, the other takes out Chinese.
Hospital & Doctor (36)	"Raincoat" stolen by drunk man from hospital turns out to be PPE kit; tests COVID-19 positive	Religion (28)	Why are Muslim youth the best at dealing with pandemics? . Because as a Quran-Teen, you always keep Allah by your self.
Work/Jobs (29)	My boss says he can't afford to pay me during this pandemic... . But luckily he says I can work for exposure.	Criminals (24)	2020 is a weird year. 5 years ago i entered a bank in a mask and i got forced to quarantine for years, today i got praised for it.. weird times to be alive
Education (16)	OHSAA coronavirus rules: Students can wrestle, but can't shake hands	Family (23)	What's the difference between Coronavirus and my dad?. Coronavirus only takes five days to appear.
work from home (10)	When the lockdown started, all I did was masturbate and watch TV all day.. After 3 weeks it got awkward and my coworkers decided to tell me how to turn my camera off on Zoom.	Men (23)	There's a Pandemic Sperm Shortage, and Sperm King. Megadonors Are in High Demand
		Fans (19)	Why did everyone have Covid-19 at the KPop concert?. Because a symptom of Coronavirus is lack of taste.
		Health Workers (18)	Thankful for all the health care workers during this pandemic. Without dem, it would be a panic.
		Poor (12)	This is the first year that we didn't go to Hawaii because of coronavirus.... Every other year we don't go because we can't afford it.
		Elderly (9)	Why did the Coronavirus cross the road?. Cuz it likes to walk old people to the other side.

Table 3: Examples for each Category and Target

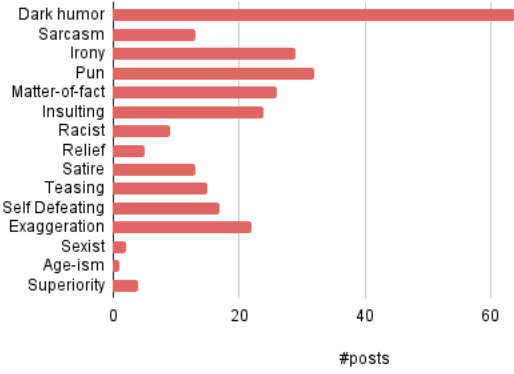


Figure 3: Distribution of Targets in Maladaptive Humor

Cambodian villagers trust magic scarecrows to ward off coronavirus

Why are Muslim youth the best at dealing with pandemics? Because as a Quran-Teen, you keep Allah by your self.

It may be noted that identified stereotypes overlapped with annotated *targets* and *categories* such as Gender, Race, Sexuality, Social-class, (Dis)ability, Age, Politics, Religion, Occupation and the Chinese population.

3.2.2 Challenges or Limitations

Humor is a complex creative construct with high subjectivity. The expressed comic content within a humorous text may evoke the feelings of mirth or confusion to a group and repulsion to some other. The use of certain words that are used to express surprise, shock or disbelief may be offensive to some as they are considered to be a part of 'rude language'. For example, in the text *TX Woman Dies Of COVID After Posting Facebook Rant: "You Don't Need Sanitizer, You Need Faith And Guns"*, the humor could be perceived by some as the irony in the situation and the dialogue, whereas others may find it non-humorous since the text deals with the death of a person.

3.2.3 Bias, privacy and Ethics

Human annotators were from Indian subcontinent, with ages in the range of 18-21, all 5 of whom identify as female. Keeping this in mind, the annotation may have a slight bias in terms of gender, age or geographical and cultural representation.

The data obtained from Reddit social media platform via the API are guaranteed to be anonymized. This ensures that the privacy of the authors of the humorous and non-humorous posts is respected and not revealed. As this work studies the differ-

ent maladaptive humor styles in the posts obtained, there may be offensive or potentially harmful content such as racist, sexist, ageist, trans/homophobic comments, comments insensitive to religious views, political views present in the dataset, which are essential to demarcate these offensive jokes from the adaptive ones.

3.3 Baseline Models

We finetuned a widely used pre-trained language model RoBERTa (Liu et al., 2019) for two tasks namely (a) Humor Detection and (b) Maladaptive humor detection. The default parameter weights were used and the model was fine-tuned for a maximum of 18 epochs. For Task-a, the dataset has 3K posts having textual humorous (1366) and non humorous posts (1) as provided in Tab. ?? whereas for Task-b, we only consider humorous posts. We consider 80% of posts for training and the rest 20% is equally divided for validation and test set.

We provide the performance evaluation for both tasks in Table 4. The model achieved an F-score of 0.86 for the task of humor detection and 0.51 for maladaptive humor detection.

Task-A: Humor Detection

Model	P	R	F-score
RoBERTa	0.88	0.84	0.86

Task-B: Maladaptive Humor Detection

RoBERTa	0.58	0.46	0.51
---------	------	------	------

Table 4: Baseline Models

4 Conclusion and Future Scope

Creating a reliable yet diverse dataset of humor in any language is a challenge. In this paper, we presented a well-curated dataset of COVID-19 humor tagged with a variety of relevant labels. We believe that this dataset will enable research on the evolving humorous content during the pandemic and its role in *information sharing* or even *behavior regulator*. One of the limitation is the size of the dataset. In future, we aim to augment this dataset using unsupervised techniques.

References

- Vikram Ahuja, Radhika Mamidi, and Navjyoti Singh. 2018. From humour to hatred: A computational analysis of off-colour humour. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 144–153. Springer.
- Umair Akram, Kamila Irvine, Sarah F Allen, Jodie C Stevenson, Jason G Ellis, and Jennifer Drabble. 2021. Internet memes related to the covid-19 pandemic as a potential coping mechanism for anxiety. *Scientific reports*, 11(1):1–8.
- Patrizia Amici. 2020. Humor in the age of covid-19 lockdown: An explorative qualitative study. *Psichiatria Danubina*, 32(suppl. 1):15–20.
- Luca Bischetti, Paolo Canal, and Valentina Bambini. 2020. Funny but aversive: a large-scale survey of the emotional response to covid-19 humor in the italian population during the lockdown.
- Karyn Buxman. 2012. Types of humor: The good, the bad, and the (sometimes) ugly. URL: <https://aath.memberclicks.net/assets/docs/HumorResources/the%20good-karyn%20aat%20hr.pdf>.
- Simon Critchley. 2011. *On humour*. Routledge.
- Christie Davies. 2008. Undertaking the comparative study of humor. *The primer of humor research, Berlin: Mouton de Gruyter*, pages 157–182.
- David JA Dozois, Rod A Martin, and Peter J Bieling. 2009. Early maladaptive schemas and adaptive/maladaptive styles of humor. *Cognitive therapy and research*, 33(6):585–596.
- Marta Dynel. 2021. Covid-19 memes going viral: On the multiple multimodal voices behind face masks. *Discourse & Society*, 32(2):175–195.
- Ahmed T Hussein and Lina Nabil Aljamali. 2020. Covid-19 humor in jordanian social media: A socio-semiotic approach. *Heliyon*, 6(12):e05696.
- Dafna Lemish and Nelly Elias. 2020. “we decided we don’t want children. we will let them know tonight”: Parental humor on social media in a time of coronavirus pandemic. *International Journal of Communication*, 14:27.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Rod A Martin, Patricia Puhlik-Doris, Gwen Larsen, Jeanette Gray, and Kelly Weir. 2003. Individual differences in uses of humor and their relation to psychological well-being: Development of the humor styles questionnaire. *Journal of research in personality*, 37(1):48–75.
- JA Meaney, Steven Wilson, Luis Chiruzzo, Adam Lopez, and Walid Magdy. 2021. Semeval 2021 task 7: Hahackathon, detecting and rating humor and offense. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, pages 105–119.
- John Morreall. 2012. *Philosophy of humor*.
- Andrew R Olah and Thomas E Ford. 2021. Humor styles predict emotional and behavioral responses to covid-19. *Humor*, 34(2):177–199.
- Ana Laura Pérez. 2021. “the ‘hate speech’ policies of major platforms during the covid-19 pandemic”. *UNESCO, Document Code: MTD/CI/2021/PI/0/REVI*, page 32.
- Rohit Revi et al. 2014. Understanding obscenity and offensive humour. *The European Journal of Humour Research*, 2(3):98–114.
- Joel Mayo Torres, Leila M Collantes, Emily T Astrero, Arceli R Millan, and Carlo M Gabriel. 2020. Pandemic humor: Inventory of the humor scripts produced during the covid-19 outbreak. *Torres, JM, Collantes, LM, Astrero, ET, Millan, AR, & Gabriel, CM (2020). Pandemic humor: Inventory of the humor scripts produced during the COVID-19 outbreak. The Asian EFL Journal*, 7(3.1):138–164.
- Simon Vurayai. 2020. The paradox of disparagement humor: An analysis of whatsapp jokes on women in the fight against the covid-19 pandemic. *Gender & Behaviour*, 18(3):15949–56.
- Massih Zekavat. 2021. Employing satire and humor in facing a pandemic. *Humor*, 34(2):283–304.

A Annotation Guidelines and Definitions

In this section, we provide the annotation guidelines and definitions followed during the data annotation.

A.1 Guidelines

The annotators were asked to follow the below steps to determine if a text is humorous:

- Read the entire text-discourse to establish a general understanding of the meaning.
- Determine if the text has implicit/explicit meanings of incongruous / conflicting / ambiguous / unexpected or strong emotions that make text humorous in the given context.
- Determine whether the text invokes a humorous reaction from you.

The definitions are provided in Tables A1, A2, A3, and A4.

To ensure that the annotators understand the task and definitions, a sample of pre-labelled 30 posts was provided for the test. Annotators were given the unlabelled posts for annotation only when they demonstrate clear understanding of the task over the test sample.

A.2 Annotator's profile

A team of five annotators were recruited for this task. All were from Indian subcontinent, with ages in the range of 18-21. All 5 of whom identify as female. All were highly proficient in English. No monetary incentive was provided for annotation.

Table A1: Key Terms in Annotation

Term	Definition
<i>Type of humor</i>	Humor style as provided in Table A4
<i>Topic of humor</i>	Key concept or idea on which the humor is based
<i>Target of humor</i>	a community/individual being targeted in the joke (who are being made fun of)
<i>stereotype</i>	a generalized belief about a particular category of people

Table A2: Adaptive vs Maladaptive Humor

Type	Definition
Adaptive Humor (1)	When someone uses humor that's affiliative or self-enhancing in nature, it's called adaptive humor. (Dozois et al., 2009)
<i>Affiliative Humor</i>	Benign humor that is used to enhance relationships with others.
<i>Self-enhancing Humor</i>	Benign humor that is used to enhance the self
Maladaptive Humor (0)	When someone uses aggressive or self-defeating humor, it's called maladaptive humor. (Dozois et al., 2009)
<i>Aggressive Humor</i>	Injurious humor that is used to enhance the self
<i>Self-defeating Humor</i>	Injurious humor that is used to enhance relationships with others

Table A3: Sub-Categories for Maladaptive Humor

Category	Definition
Teasing (T)	the act of persistently annoying someone, especially playfully and with silly jokes about them.(Buxman (2012))
Insulting (I)	It consists of offensive insults directed to a person or a group.(Ahuja et al. (2018))
Sexist (X)	suggesting that the members of one sex are less able, intelligent, etc. than the members of the other sex, or referring to that sex's bodies, behaviour, or feelings in a negative way.
Racist (R)	having the belief that some races of people are better than others or having general beliefs about other people based only on their race.
Age-ism (A)	It comprises of jokes that stereotype, prejudice, and discriminate against people on the basis of their age.

Table A4: Types of Humor

Type	Definition
Satire (S)	a way of criticizing people or ideas in a humorous way, especially in order to make a political point, or a piece of writing that uses this style.
Self-defeating (SD)	excessive self-disparaging humor; amusing others at one's own expense to facilitate integration or gain approval; allowing oneself to be the focus of others' humor. (Martin et al. (2003))
Dark humor (D)	a gruesome or a tragic topic such as death/illness is mixed with an innocuous topic which creates shock and inappropriateness.(Ahuja et al. (2018))
Pun (P)	Puns are usage of certain words in speech to suggest two or more meanings or different associations of a word, or using two or more words with different meanings that sound the same or nearly the same in order to produce humor.
Sarcasm (SR)	is often used to convey thinly veiled disapproval humorously. It is something that mocks or ridicules, usually by saying the opposite of what is actually meant.
Irony (IR)	humor that involves incongruity and disagreement between what is said and what is understood or what is said and what actually happened.
Exaggeration (E)	a statement that represents something as better or worse than it really is, i.e. an overstatement of the truth.
Relief (RE)	when the main purpose of the joke is to act as an outlet for built-up tension by finding humor in morally problematic, dehumanizing conjectures.(Revi et al. (2014))
Superiority (SU)	where the joke expresses feelings of superiority or dominance over other people/communities or over a former state of ourselves.(Morreall (2012))
Matter of-fact (obvious/facts) (M)	to introduce a statement in an unemotional and practical way, that gives more details about what has just been said, an explanation of it, or something that contrasts with it.