Universität Stuttgart
Institut für Maschinelle Sprachverarbeitung

CoVERT: A Corpus of Fact-checked Biomedical COVID-19 Tweets
Isabelle Mohr, Amelie Wuehrl, Roman Klinger

1. Motivation

COVID-19 Infodemic ➡ increased need for fact-checking biomedical claims on social media

2. Contributions

- Corpus of fact-checked biomedical tweets with NER annotations, verdicts and evidence texts
- Evidence-based fact-checking approach outperforms approach that uses language models (LMs) to generate evidence texts

3. Annotation Process

3.1 NER Annotation

- Manual annotations supported by biomedical NER model
- Annotating only NER of the 'causal' phrase
- NER annotations relevant for creating prompts

3.2 Fact-checking Annotation

- 3 annotators per tweet, crowdsourced on Prolific
- Label adjudication: majority label for each tweet (only 2 cases no agreement -> NEI)

Biomedical COVID-19 Tweet:
@handle @handle 1.) They take away the most important part of our comunication: mimic and body language in general. 2.) **The humid, warm environment behind the masks can cause bacterial pneumonia, which in turn can lead to Corona disease**. 1/2

Could you find a resource that confirms or refutes the claim in bold text?*
☐ Yes, the resource CONFIRMS this claim
☐ Yes, the resource REFUTES this claim
☐ No, I could not find any resource that confirms or refutes this claim

Enter the URL to the reputable resource you have found that substantiates or refutes the above claim. If you marked 'No (..)' previously, skip this field.
[Enter URL here]

Copy and paste the extract from this source that confirms or refutes the above claim. If you marked 'No (...)' previously, skip this field.
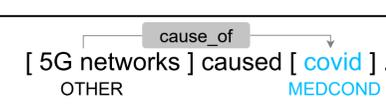[Enter text here]

Figure 1: Fact-checking annotation environment.

4. Corpus Statistics

	Annotation Category	Value	IA Score
Entities	Medical Condition	424	0.94 (F_1)
	Treatment	102	0.95 (F_1)
	Other	102	0.93 (F_1)
	Symptom/Side-effect	94	0.68 (F_1)
Relations	cause_of	253	
	causative_agent_of	30	
	not_cause_of	17	
Verdicts	SUPPORTS	198	
	REFUTES	66	
	NEI	36	
	Total (dataset size)	300	0.44 (k)

Table 1: Corpus statistics and inter annotator agreement (IA) on annotations.

Annotation	Example
NER Tweet	cause_of [5G networks] caused [covid] . OTHER MEDCOND
Verdict	REFUTES
URL	https://www.muhealth.org/ ...
Evidence	There are two types of conspiracy associated with 5G-COVID-19. One version suggests that radiation from 5G lowers your immune system ...

Table 2: Example instance in corpus

5. Verdict Prediction

- Explore performance of fact-checking pipeline using either LM-generated 'evidence' or evidence texts from the web
- Pipeline: <claim, evidence> pairs as input to textual entailment model
  - Multi-layer perceptron (MLP) component for classification
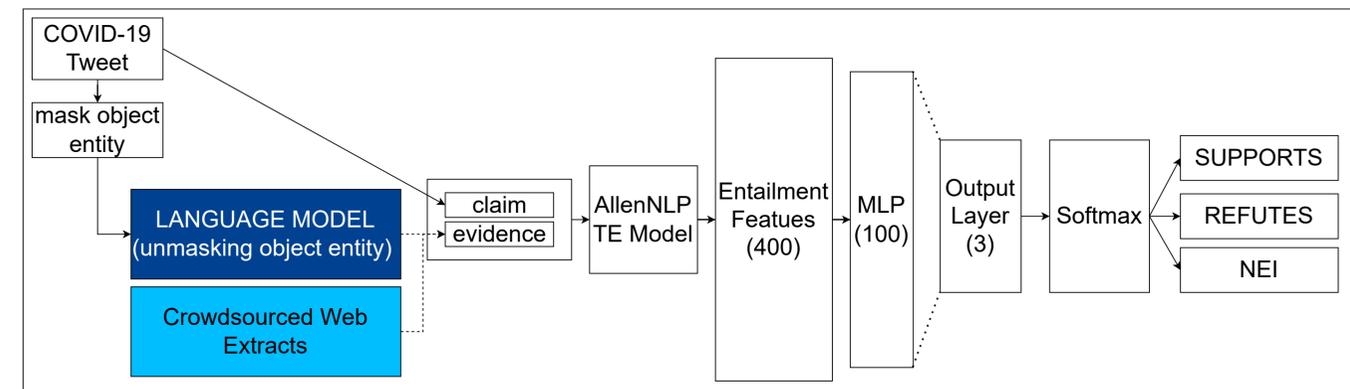

Figure 2: Full diagram of fact-checking pipeline. 'Evidence' is either generated by LM (dark blue) or taken from crowdsourced web extracts (light blue).

6. Results

	FEVER			Tweet + LM			Tweet + Ev.		
	P	R	F_1	P	R	F_1	P	R	F_1
MLP-FEVER	.52	.47	**.49**	.60	.61	**.60**	.61	.38	.46
MLP-Evidence	.64	.36	.46	.66	.68	.57	.68	.74	**.69**

Table 3: Fact-checking pipeline results.

- Improvement via fine-tuning MLP component with real evidence
- NEI category not modelled with CoVERT evidence
- Most errors in LM-generated evidence pairs are REFUTES misclassified as SUPPORTS

7. Take-aways

- Fact-checking annotation via crowdsourcing can achieve moderate agreement
- Fact-checking pipelines can benefit from real-world evidence over LM-generated evidence texts

Read the full paper:

✉ isabelle.mohr@ims.uni-stuttgart.de