



NEWSCLAIMS: A New Benchmark for Claim Detection from News with Attribute Knowledge

Revanth Gangi Reddy¹, Sai Chetan Chinthakindi¹, Zhenhailong Wang¹, Yi R. Fung¹, Kathryn Conger², Ahmed Elsayed², Martha Palmer², Preslav Nakov³, Eduard Hovy^{4,5}, Kevin Small⁶, Heng Ji¹





¹UIUC ²CU Boulder ³MBZUAI ⁴DARPA ⁵CMU ⁶Amazon Alexa



CODF

1) Introduction

- NEWSCLAIMS extends the claim detection task to extract additional attributes relating to the claim, such as the claimer, claim object, etc.
- ► We introduce the notion of *claim object*, which is an entity that identifies what is being claimed with respect to the topic of the claim.
- Mainly contains claims about COVID-19 from 143 news articles.

2) QA-based Claim Detection Framework (COLING 2022)

- The same extractive QA model solves multiple sub-tasks within claim detection, without the need of any task-specific training data.
- ► This involves:
 - Filtering claims relating to specific topics
 - Identifying claim objects associated with such topics
 - Attribution for identifying the claimers making these claims
- Realized by using directed questions to help solve connected sub-tasks.

News Article

The claim originated in a Jan. 26, 2020 press release titled "Vitamin C Protects Against Coronavirus," from the International Society for Orthomolecular Medicine, which promotes large doses of nutritional supplements. The article was then republished in full the next day HealthImpactNews.com, a Red-rated network of health sites. From the WHO, there has been no information nor evidence to suggest that the new coronavirus could by transmitted by mosquitoes.

Claim Detection Claim Sentence 1 Topic: The claim originated in a Jan.

Vitamin C Protects Claim Span Against Coronavirus Detection

26, 2020 press release titled "Vitamin C Protects Against Coronavirus,", from the	from the virus	Stance Detection
Claim Sentence 2 From the WHO, there has been no information nor	Topic: Transmitting	Claim Obje Detection
evidence to suggest that the new coronavirus could by transmitted by mosquitoes.	the virus	Claimer Detectio

Affirm Vitamin C International Society for Orthomolecular Medicine

Figure 1: An example from NewsClaims.

Topic	Claim Sentence
Origin	The genetic data is pointing to this virus coming from a bat reservoir
Transmission	The virus lingers in the air indoors, infecting those nearby
Cure	Vitamin C is an effective treatment for COVID-19
Protection	Taking a hot bath prevents you from getting COVID-19

Table 1: Examples showing the claim object in **bold** for some claims in NEWSCLAIMS

- Claim detection approaches need to be able to be ported to new scenarios, without access to much training data.
- Further, the claimer detection subtask within NewsClaims requires considerable document-level reasoning, making it harder for existing attribution models which mainly involve sentence-level reasoning.

3) Claim Topic Filtering and Claim Object Detection



Figure 4: Proposed QA-based claim detection framework.

4) Document-level Claimer Detection

- Claims within news articles come from various types of sources.
- Claimer detection involves identifying whether the claim is made by a journalist or who the source is, in case it is reported in the news article.

Claim Topic Filtering

- Topic relevance is measured using answer confidence from QA model
- Input to the QA model comprises of the claim sentence along with the question corresponding to the individual topic
- ► For NLI baseline, topic relevance is via the entailment score

Model	Ρ	R	F1
ClaimBuster	13.0	86.5	22.6
ClaimBuster + NLI	21.8	53.3	30.9
ClaimBuster + ClaimQA	30.7	43.4	36.0
Human (single)	52.7	70.0	60.1
Human (3-way)	60.2	83.5	70.0

Table 2: Performance for identifying claim sentences.

Premise : There is beir	Context : There is no evidence that eating garlic prevents you from being infected with the coronavirus
Hypothesis 1 (prot	Question 1 (protection) : What can protect from the virus?
Hypothesis 2	Question 2 (origin) : What is the origin of the virus?
Hypothesis	Question 3 (cure) : What can cure the virus?
Hypothesis 4 (trans	<i>Question 4 (transmission)</i> : What can transmit the virus?
Figure 3: Hy	Figure 2: Questions corresponding to

ootheses corresponding to





individual topics.

s is about protection from the virus

evidence that eating garlic prevents you from

his is about <mark>cure</mark> for the virus

sion) : This is about transmission of the virus

rigin) : This is about origin of the virus

nfected with the coronavirus

Claim Object Detection

- Claim object detection involves identifying what is being claimed in the claim sentence with respect to the topic.
- The answer span from the QA model for the question corresponding to the claim topic is used as the claim object.

Approach	Model	Туре	F1
Prompting	GPT-3	Zero-shot	15.2
Prompting	T5	Zero-shot	11.4
In-context learning	GPT-3	Few-Shot	51.9
Prompt-based fine-tuning	T5	Few-Shot	51.6
ClaimQA	BERT	Zero-shot	57.0
Human	-	-	67.7

Table 3: F1 score for claim object detection.



Figure 6: Two-step ClaimQA pipeline for identifying the claimer.

Model	Overall F1	Reported	Journalist	
SRL	41.7	23.5	67.2	
PolNeAR	42.3	25.5	65.9	
ClaimQA	50.1	39.8	64.4	
Human	85.8	81.3	88.9	
Table 4: F1 for identifying the claimer.				

Model	In-sentence	Out-of-sentence
SRL	35.8	2.4
PolNeAR	38.9	2.7
ClaimQA	46.2	29.0

Table 5: F1 for claimer detection for when it is present within or outside the claim sentence.