Natural Language Inference with Self-Attention for Veracity Assessment of Pandemic Claims

Miguel Arana-Catania[1,2], Elena Kochkina[3,2], Arkaitz Zubiaga[3], Maria Liakata[3,2], Rob Procter[1,2], Yulan He[1,2]
[1] Department of Computer Science, University of Warwick, UK
[2] Alan Turing Institute, UK
[3] Queen-Mary University of London, UK

https://panacea2020.github.io

CONTENTS

01 Motivation

Motivation

- Misinformation detection especially relevant during the COVID-19 pandemic.

- Current approaches and datasets focus on a **single**: medium (TW, FB, websites), information domain (health, scholar), type of information (news, claims), or application (retrieval, verification).

Contributions

- **New dataset** containing **heterogeneous** pandemic claims and their information sources
- 2 novel **fact verification approaches**
- **Comprehensive experiments**, from information sources collection through information retrieval to veracity assessment.

02
PANACEA dataset

Data sources

Data Source	Description	Domain	No. of claims (False / True)
CoronaVirusFacts Database	Published by Poynter, this online source combines fact-checking articles from more than 100 fact-checkers from all over the world, being the largest journalist fact-checking collaboration on the topic worldwide.	Heterogeneous	11,647 (11,647 / 0)
CoAID dataset (Cui and Lee, 2020)	This contains fake news from fact-checking websites and real news from health information websites, health clinics, and public institutions.	News	5,485 (953 / 4,532)
MM-COVID (Li et al., 2020)	This multilingual dataset contains fake and true news collected from Poynter and Snopes.	News	3,409 (2,035 / 1,374)
CovidLies (Hossain et al., 2020)	This contains a curated list of common misconceptions about COVID appearing in social media, carefully reviewed to contain very relevant and unique claims.	Social media	62 (62 / 0)
TREC Health Misinformation track	Research challenge using claims on the health domain focused on information retrieval from general websites through the Common Crawl corpus (commoncrawl.org).	General websites	46 (39 / 7)
TREC COVID challenge (Voorhees et al., 2021; Roberts et al., 2020)	Research challenge using claims on the health domain focused on information retrieval from scholar peer-reviewed journals through the CORD19 dataset (Wang et al., 2020a), the largest existing compilation of COVID-related articles.	Scholar papers	40 (3 / 37)

Data sources used for the construction of our dataset.

PANACEA dataset

Category	LARGE	SMALL
False	1,810	477
True	3,333	1,232
Total	5,143	1,709

https://zenodo.org/record/6493847

Claim	Category	Source	Orig. data src.	Type
Stroke Scans Could Reveal COVID-19 Infection.	True	ScienceDaily	CoAID	
Whiskey and honey cure coronavirus.	False	Independent news site	CovidLies	
COVID-19 is more deadly than Ebola or HIV.	False	Australian Associated Press	Poynter	
Dextromethorphan worsens COVID-19.	True	Nature	TREC Health Misinformation track	
ACE inhibitors increase risk for coronavirus.	False	Infectious Disorders - Drug Targets journal	TREC COVID challenge	
Nancy Pelosi visited Wuhan, China, in November 2019, just a month before the COVID-19 outbreak there.	False	Snopes	MM-COVID	Named Entity, Numerical content

Example entries

PANACEA dataset

Information Retrieval and re-ranking.

- Pyserini - https://github.com/castorini/pyserini

$$\text{BM25}(q,d) = \sum_{t \in q \cap d} \log \frac{N - \text{df}(t) + 0.5}{\text{df}(t) + 0.5} \cdot \frac{\text{tf}(t,d) \cdot (k_1 + 1)}{\text{tf}(t,d) + k_1 \cdot \left(1 - b + b \cdot \frac{l_d}{L}\right)}$$

q query, d document, df documents frequency of term t, tf term frequency

[Robertson et al., 1994, Crestani et al., 1999, Robertson and Zaragoza, 2009]

- Pygaggle - https://github.com/castorini/pygaggle

- **MonoT5** [Nogueira et al., 2020]. T5 model [Raffel et al, 2019] trained on queries q and documents d with an Input sequence “Query:q Document:d Relevant:” and Output sequence “True/False”.
- Using pretrained model ‘castorini/monot5-base-msmarco’ Trained on MS-MARCO dataset [Bajaj et al, 2016]

PANACEA dataset

Category	Orig.	LARGE	SMALL
Similarity	0.67 ± 0.23	0.43 ± 0.13	0.37 ± 0.14
$\eta_{.90}$	0.99	0.60	0.56
False	14,739	1,810	477
True	5,950	3,333	1,232
Total	20,689	5,143	1,709

PANACEA dataset LARGE/SMALL statistics

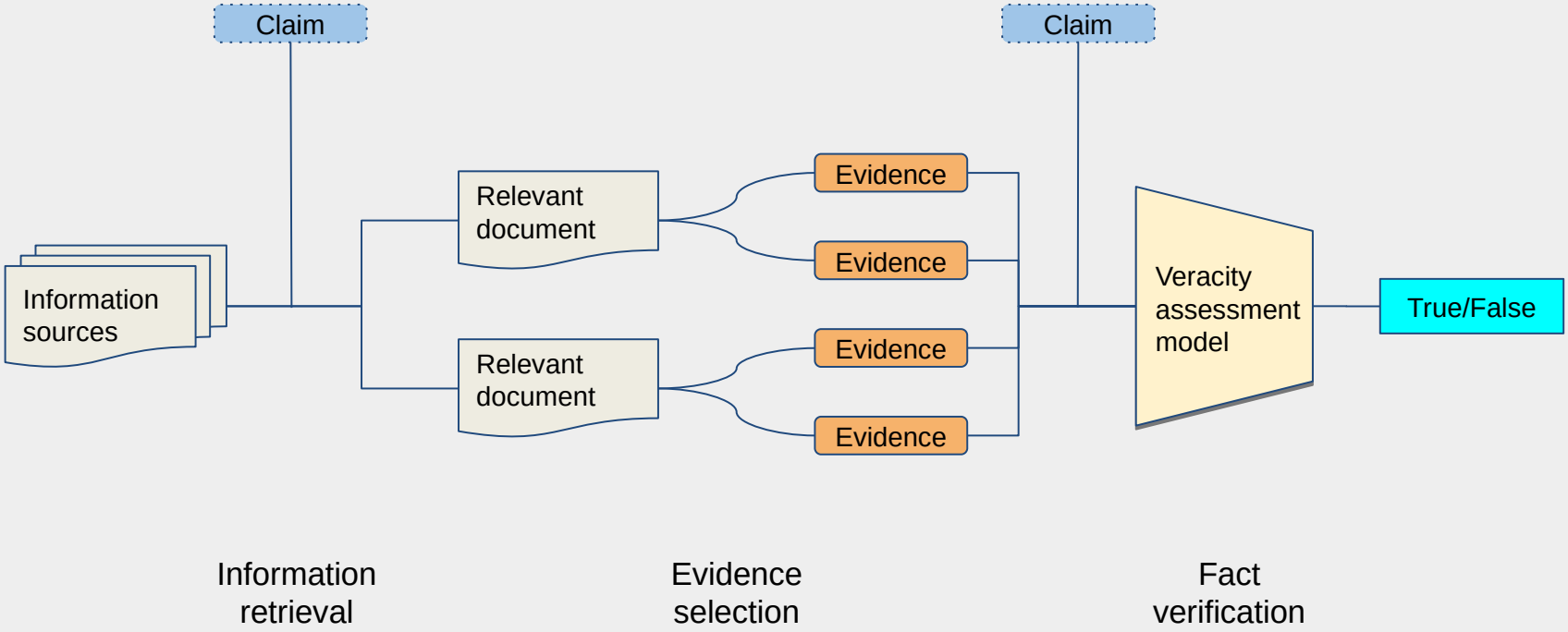
Similarity using BERTScore

Claim 1: Losing your sense of smell may be an early symptom of COVID-19.
Exclude from LARGE and SMALL: Loss of smell may suggest milder COVID-19. **Exclude from SMALL only**: Loss of smell and taste validated as COVID-19 symptoms in patients with high recovery rate.
Claim 2: COVID-19 hitting some African American communities harder.
Exclude from LARGE and SMALL: The African American community is being hit hard by COVID-19. **Exclude from SMALL only**: COVID-19 impacts in African-Americans are different from the rest of the U.S. population.

Claim de-duplication examples

03
Experiments

Experiments



04
Document retrieval

Document retrieval

Information sources:

(1) Centers for Disease Control and Prevention (**CDC**)

(2) European Centre for Disease Prevention and Control (**ECDC**)

(3) Online publisher of news and information on health **WebMD**

(4) World Health Organization (**WHO**)

	AP@5	AP@10	AP@20	AP@100
BM25	0.54	0.56	0.58	0.62
BM25+MonoBERT	0.52	0.55	0.58	0.62
BM25+MonoT5	**0.55**	**0.58**	**0.60**	**0.62**
BM25+RM3+MonoT5	0.51	0.53	0.55	0.57

Document retrieval results. Average precision for different cut-offs.

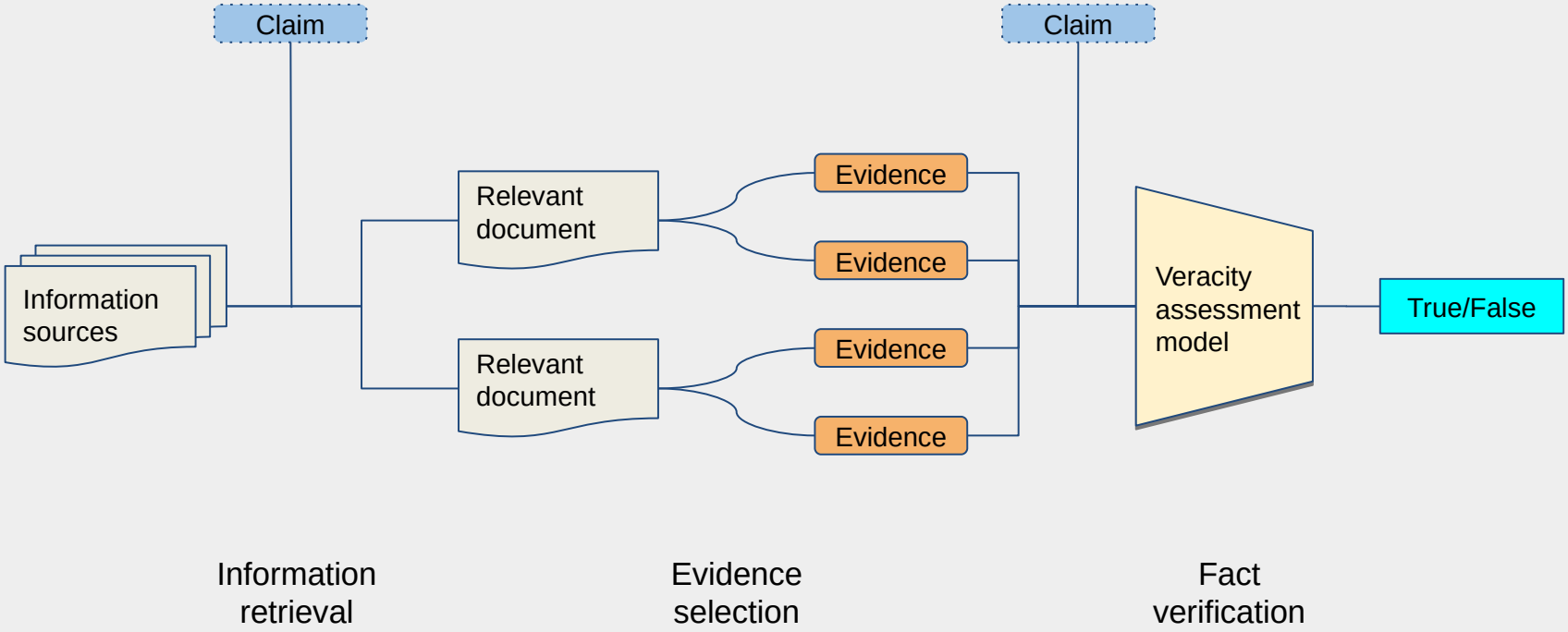
Information retrieval errors

Claim	Description
"*Sugar causes a cytokine storm in the lungs that promotes COVID-19*"	Retrieved documents are relating COVID and its cytokine storm effects, but without the specific mention of sugar, which does not cause a cytokine storm.
"*Barron Trump had COVID-19, Melania Trump says*"	Retrieved sentences such as "*Rudy Giuliani has tested positive for COVID-19, Trump says.*" with a similar structure and mentions but mistaking the family members and missing the key name.
"*Prince Charles tested positive for COVID-19 after meeting Bollywood singer Kanika Kapoor.*"	Documents mentioning Prince Charles positive COVID tests are obtained, but without any mentions to the singer.
"*Vice President of Bharat Biotech got a shot of the indigenous COVAXIN vaccine*"	Correct documents on the issue are retrieved. Similar sentences are retrieved such as "*Covaxin which is being developed by Bharat Biotech is the only indigenous vaccine that is approved for emergency use.*" or "*Bharat Biotech's Covaxin is the first Indian vaccine to receive approval to conduct Phase I/Phase II trials.*". However, being similar they give no information about the claimed situation. In the retrieved document, the sentence "*The pharmaceutical company, has in a statement, denied the claim and said the image shows a routine blood test.*" contains the essential information to debunk the original claim. But it is missed by the sentence retrieval engine as it is very different from the claim.
"*Masks can be sanitized in microwave*"	Correct documents are retrieved with similar sentences such as "*Claiming masks can be sanitized in microwave resurfaces*". However, sentences such as "*The study authors cautioned health care workers against trying to clean masks this way. Microwaves melted the masks, making them useless.*" or "*He also warns people against using microwaves or ovens to heat their masks.*" that are present in the retrieved documents but are not similar enough to the claim are missed.

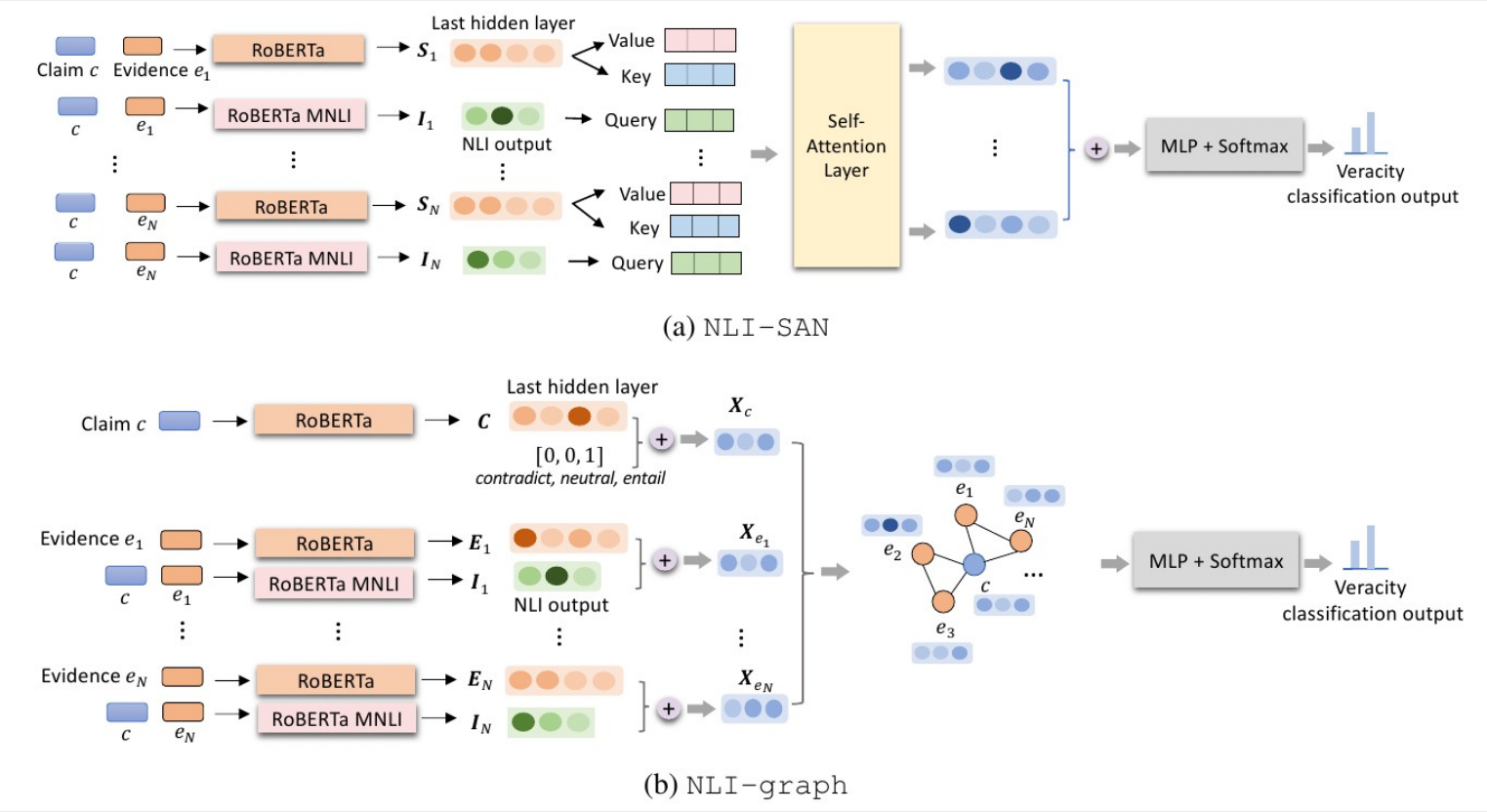
Examples of errors in document or sentence retrieval.

05
Veracity classification

Experiments

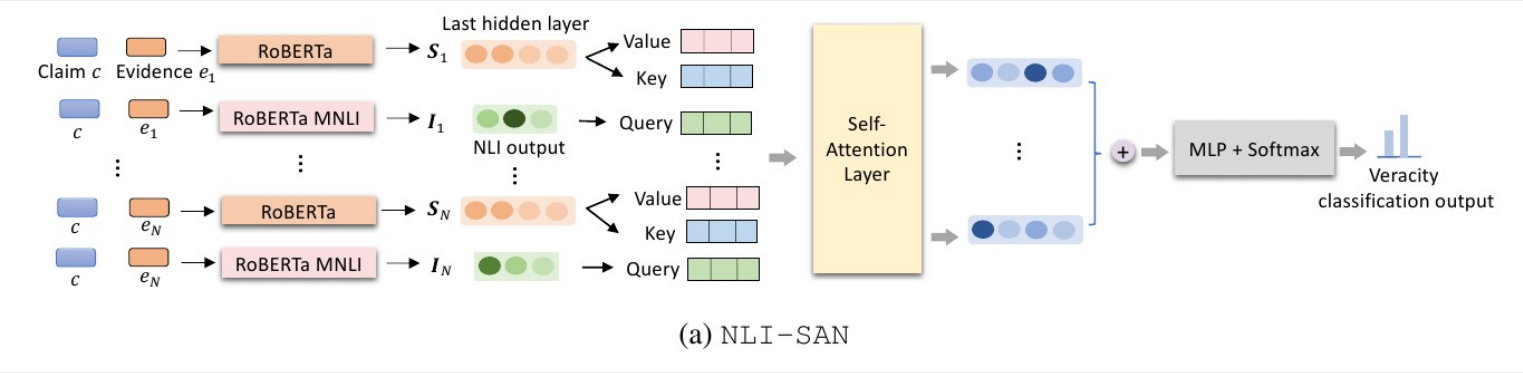


Fact verification approaches



NLI-SAN and NLI-graph proposed architectures for fact verification

Fact verification approaches



(a) NLI-SAN

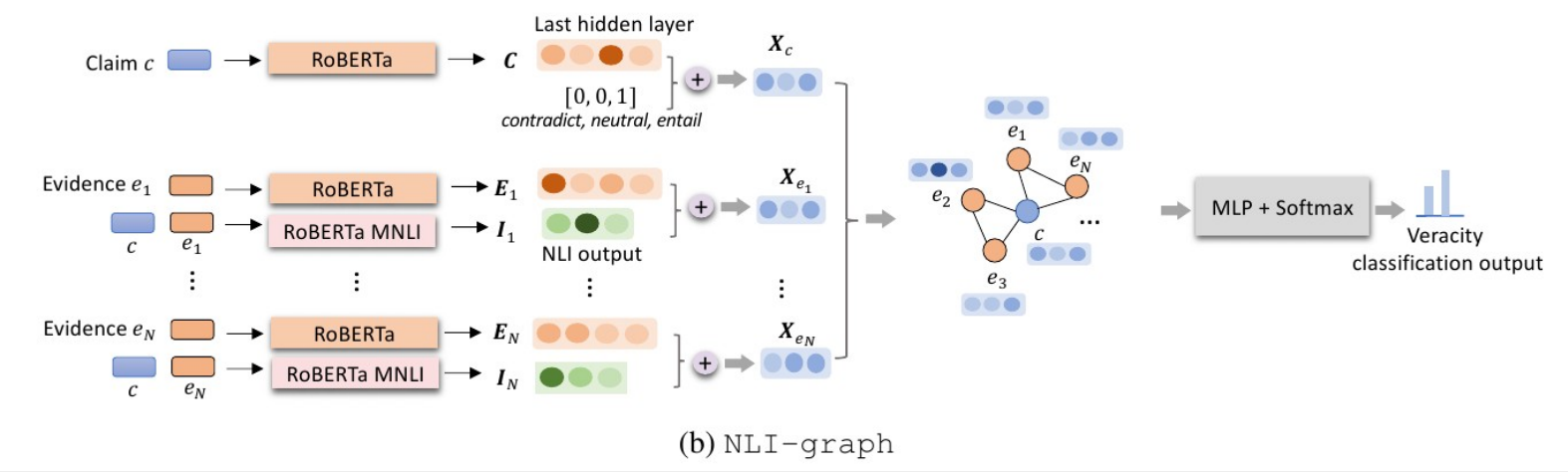
$$\mathbf{S_i} = \text{RoBERTa}(c, e_i)$$
$$\mathbf{I_i} = \text{RoBERTa}_{\text{NLI}}(c, e_i) \quad (1)$$

$$\text{Att}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}(\mathbf{Q}\mathbf{K}^\top / \sqrt{d})\mathbf{V} \quad (2)$$

$$\hat{\boldsymbol{y}} = \text{softmax}(\text{MLP}_{\text{ReLU}}(\text{O}_{\text{SAN}})) \quad (3)$$

NLI-SAN

Fact verification approaches



(b) NLI-graph

$$\mathbf{C_i} = \text{RoBERTa}(c); \quad \mathbf{E_i} = \text{RoBERTa}(e_i) \quad (4)$$

$$\mathbf{I_i} = \text{RoBERTa}_{\text{NLI}}(c, e_i) \quad (5)$$

$$\mathbf{X}' = \hat{\mathbf{D}}^{-1/2}\hat{\mathbf{A}}\hat{\mathbf{D}}^{-1/2}\mathbf{X}\mathbf{W}, \quad (6)$$

$$\hat{\boldsymbol{y}} = \text{softmax}(\text{MLP}_{\text{ReLU}}(\text{O}_{\text{graph}})) \quad (7)$$

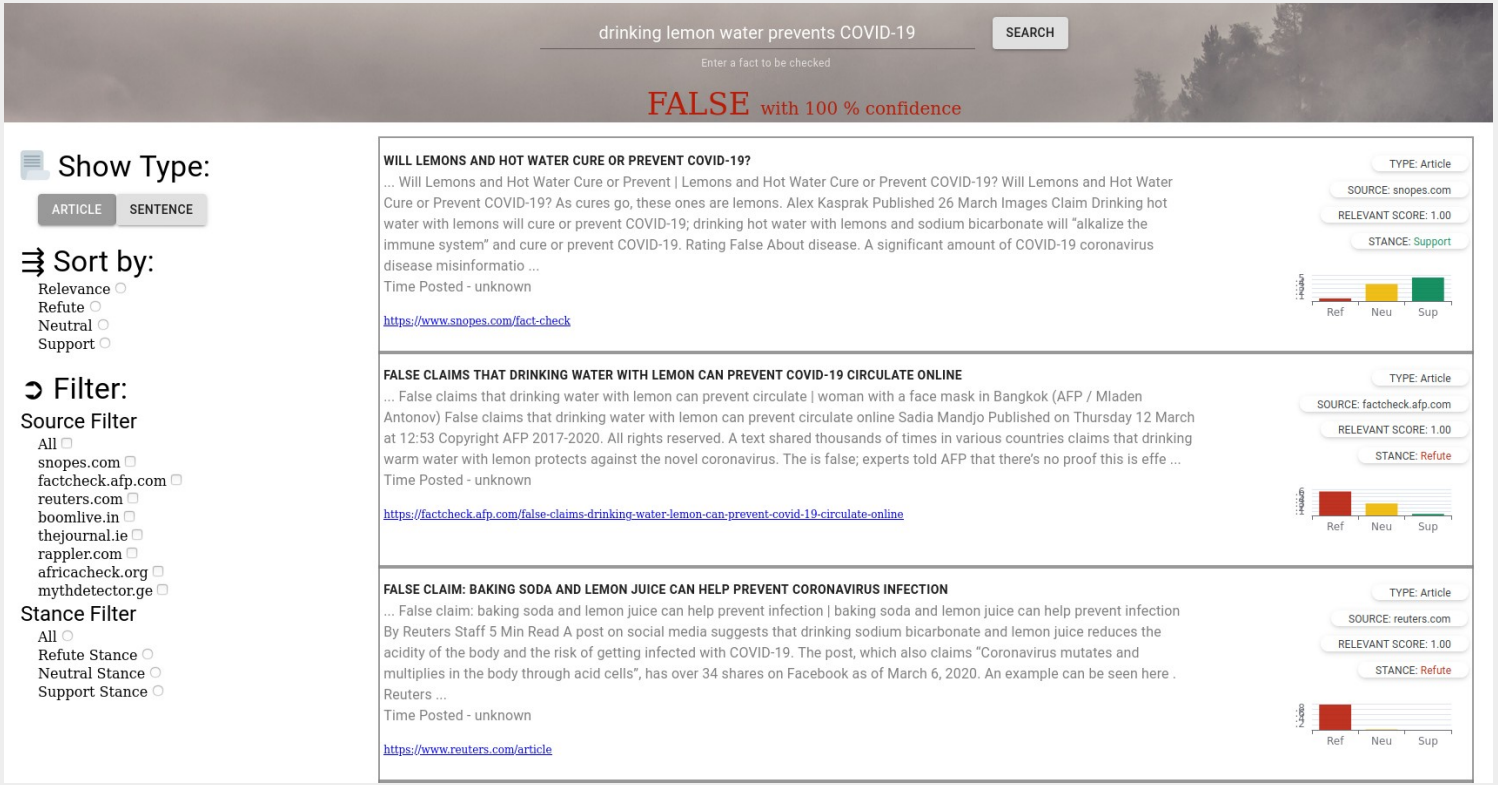
NLI-graph

Veracity classification results

Model	False			True			Macro F1
	Precision	Recall	F1	Precision	Recall	F1	
GEAR (Zhou et al., 2019)	0.81	0.60	0.69	0.85	0.94	0.89	0.79
KGAT (Liu et al., 2020)	0.89	**0.96**	**0.92**	**0.98**	0.95	**0.97**	**0.94**
NLI	0.48	0.24	0.31	0.75	0.90	0.82	0.56
NLI+Sent	0.91	0.87	0.89	0.95	**0.97**	0.96	0.92
NLI+PSent	0.87	0.72	0.79	0.90	0.96	0.93	0.86
NLI-SAN	**0.93**	0.89	0.91	0.96	**0.97**	**0.97**	**0.94**
NLI-graph$_{-abl}$	0.50	0.33	0.39	0.77	0.87	0.81	0.60
NLI-graph	0.89	0.83	0.86	0.94	0.96	0.95	0.90

Veracity classification results on the PANACEA SMALL dataset.

Demo site



Web implementation by Runcong Zhao, University of Warwick

06
Conclusions

Conclusions

- Novel PANACEA dataset: heterogeneous COVID-19 claims and fact-checking sources.
- Deduplication process of claims to ensure uniqueness.
- Information retrieval experiments using a multi-stage re-ranker approach.
- New NLI veracity assessment methods:
  - attention-based NLI-SAN
  - graph-based NLI-graph
- Discussion of challenging cases and ideas for future research directions.

Natural Language Inference with Self-Attention for Veracity Assessment of Pandemic Claims

Miguel Arana-Catania[1,2], Elena Kochkina[3,2], Arkaitz Zubiaga[3], Maria Liakata[3,2], Rob Procter[1,2], Yulan He[1,2]
[1] Department of Computer Science, University of Warwick, UK
[2] Alan Turing Institute, UK
[3] Queen-Mary University of London, UK

https://panacea2020.github.io