







Zero-shot Fact Verification by Claim Generation

Liangming Pan, Wenhu Chen, Wenhan Xiong, Min-Yen Kan, William Yang Wang

ACL 2021

Fact Verification





Donald Trump

stated on April 27, 2020 in comments made during a White House briefing:

"We've tested more than every country combined."

HEALTH CHECK

CORONAVIRUS

A DONALD TRUMP



Donald Trump's claim that US tested more than all countries combined is Funts on Fire wrong

OF TRACE THAT IS A SENSET!

If the State of the State of

We've send her fire eve.

"We've send her fire eve.

"We've send her fire her pointify orthined." During an analysis of the send of the se

seaped of the interest cases to size and desire for face and due desired on earliers. Some search of medicals, the face people from their frames seed of and the first 15 sound except desired "energy contented"?

We exceed to the face in the face of the face of the face of the search of the face in the face of the face of the face of the search of the face in the face of the face of the search of the their search of the face of the face of the face of the search of the search of the face of the face of the search of the desired or the face of the face of the face of the search of after a substitute of the face of the face of the search of the desired or substitute of the face of the f

contract contracts, as the section has been, done than "of contracts contracts," as the section of the "of dependence, we not contract and "o to contract of the selects" contracts to request within you fection by the section of the selection of the section of the selection of and the "of the selection of the selection of and the "of the selection of the selection of The numbers."

execution device on the plant health experts a feet of the beauty and a plant of the plant of th

FEATURED FACT-CHEC

Passo, lemma and the D.K. both Welderstern and No Data plane that would alone 6 million. Either map, 1% no the U.S. has done. Big picture, a relay adhanced name of methy by "wwayness," all maps consisten" amid generate ingless tools and a manufacel of market.*

Transp is also flusting on the energy figure, global health expensed.

"The highest number of two twices and of that it is not twenty for any particular energies in hunter within a manny." well become for any, and proposition or the laster freely final medical contents of the contents

We inhalise them has be right projection than may off impart execution. There of the measure is the solid bear in the right production of the solid bear in the foreign frame or classific. The solid bear in the continue, frame or classific. The continue of the solid florar and other could invarious. The continue of the protection for CROSS is assigned with the remaind souther of protection for CROSS is assigned with the remaind souther of the processing of the salest's population fair has been count for internal countries. The continue of the countries of the countries of the salest's population fair has been countries for its ...

framego promet our four Germany, Nisted, Margan and Canada, the and Stander Good Nager processings of the projection flow of Intellectual Association of the Committee of the Committee of of paratite out mode. A lone processing reduction a various is of paratite out mode. A lone processing reduction a various is a paratite out mode. A lone processing reduction a various is to intellect quitte or processing and processing of the control to intellectual and supplies of the consuming for intellectual for all to intellectual processing and the control for the control of the control of the control for the control of the control of the control for the control of the contro

Our rating

here is no resemble may to conclude that the iteratives systems to not convenient may be about the state of t

Fact Verification: the FEVER dataset

Claim: The Rodney King riots took place in the most populous county in the USA.

[wiki/Los_Angeles_Riots]

The 1992 Los Angeles riots, also known as the Rodney King riots were a series of riots, lootings, arsons, and civil disturbances that occurred in Los Angeles County, California in April and May 1992.

[wiki/Los_Angeles_County]

Los Angeles County, officially the County of Los Angeles, is the most populous county in the USA.

Verdict: Supported

Split	SUPPORTED	REFUTED	NEI
Train	80,035	29,775	35,639
Dev	6,666	6,666	6,666
Test	6,666	6,666	6,666

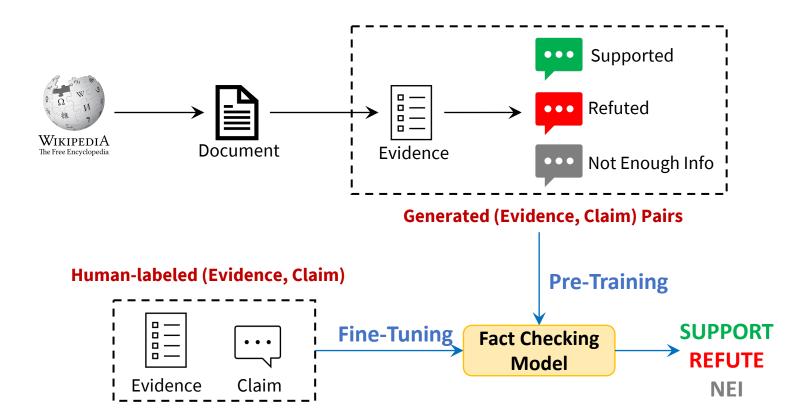
Labels:

Supports, Refutes, Not Enough Info (NEI)

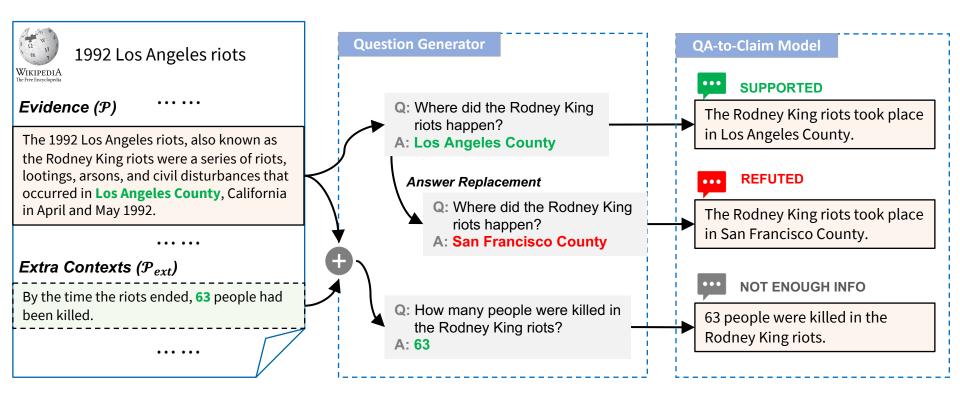
Motivation

1) Training fact checking models require large scale human-annotated (Evidence, Claim, Label) data, such as FEVER.

- It is time-consuming to ask humans to annotate the training data.
- Do we really need human annotation to train fact checking models?
- Could we automatically generate (Evidence, Claim, Label) to train the fact verification model?



Claim Generation with QG



Examples of Generated Claims

Evidence

Budapest is cited as one of the most beautiful cities in Europe, ranked as the most liveable Central and Eastern European city on EIU's quality of life index, ranked as "the world's second best city" by Conde Nast Traveler, and "Europe's 7th most idyllic place to live" by Forbes.

Alia Bhatt received critical acclaim for portraying emotionally intense characters in the road drama Highway (2014), which won her the Filmfare Critics Award for Best Actress, and the crime drama Udta Punjab (2016), which won her the Filmfare Award for Best Actress.

Generated Claim

SUPPORTED claims

Budapest is ranked as the most liveable city in central Europe. Budapest ranks 7th in terms of idyllic places to live in Europe.

REFUTED claims

Budapest ranks in 11th in terms of idyllic places to live in Europe. Budapest is ranked the most liveable city in Asia.

NEI claims

Budapest is one of the largest cities in the European Union. Budapest is the capital of Hungary.

SUPPORTED claims

Bhatt won the Filmfare Award for Best Actress in Udta Punjab. Bhatt received the Filmfare Critics Award for her role in Highway.

REFUTED claims

Alia Bhatt won the Best Original Screenplay award in Highway.

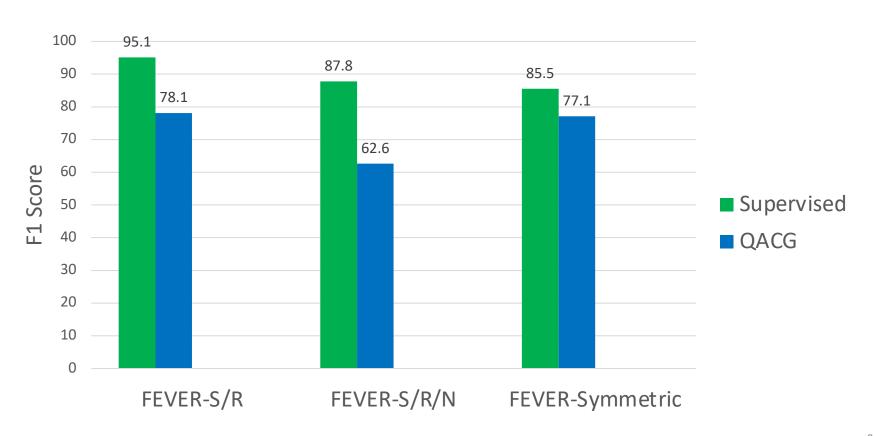
2 States (2014) won Alia Bhatt the Filmfare Award for Best Actress.

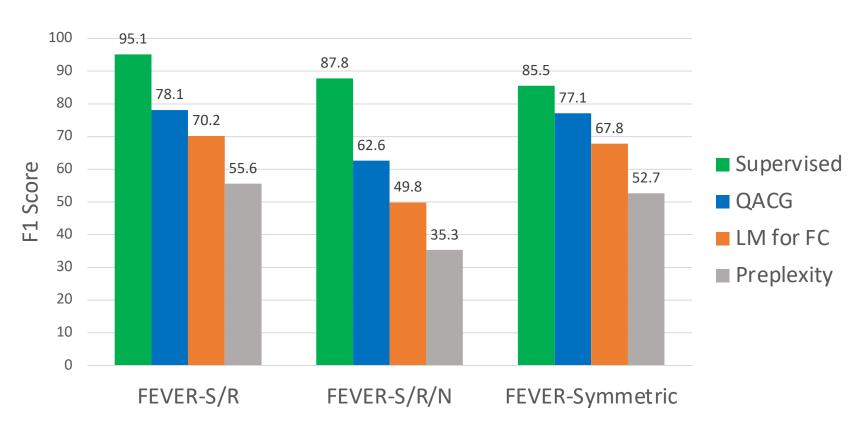
NEI claims

Alia Bhatt made her acting debut in the 1999 thriller Sangharsh. Bhatt played her first leading role in Karan Johar's romantic drama.

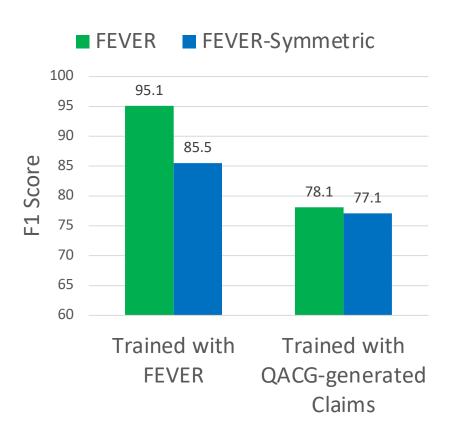
Model		FEVER -Symmetric	FEVER-S/R	FEVER-S/R/N
		$P/R/F_1$	$P/R/F_1$	$P/R/F_1$
Supervised	S1. BERT-base (Devlin et al., 2019)	81.5 / 81.3 / 81.2	92.8 / 92.6 / 92.6	85.7 / 85.6 / 85.6
	S2. RoBERTa-large (Liu et al., 2019)	85.5 / 85.5 / 85.5	95.2 / 95.1 / 95.1	88.0 / 87.9 / 87.8
Zero-shot	U1. Random Guess	50.0 / 50.0 / 50.0	50.0 / 50.0 / 50.0	33.3 / 33.3 / 33.3
	U2. GPT2 Perplexity	52.7 / 52.7 / 52.7	55.6 / 55.6 / 55.6	35.3 / 35.3 / 35.3
	U3. MNLI-Transfer	62.2 / 55.5 / 58.7	63.6 / 60.5 / 61.8	41.4 / 39.6 / 40.7
	U4. LM as Fact Checker (Lee et al., 2020b)	71.2 / 64.5 / 67.8	77.9 / 65.6 / 70.2	64.3 / 54.6 / 49.8
	U5. QACG (BERT-base)	73.2 / 73.0 / 72.9	74.2 / 74.0 / 74.1	56.5 / 55.7 / 55.9
	U6. QACG (RoBERTa-large)	77.3 / 77.0 / 77.1	78.1 / 78.1 / 78.1	64.6 / 62.0 / 62.6

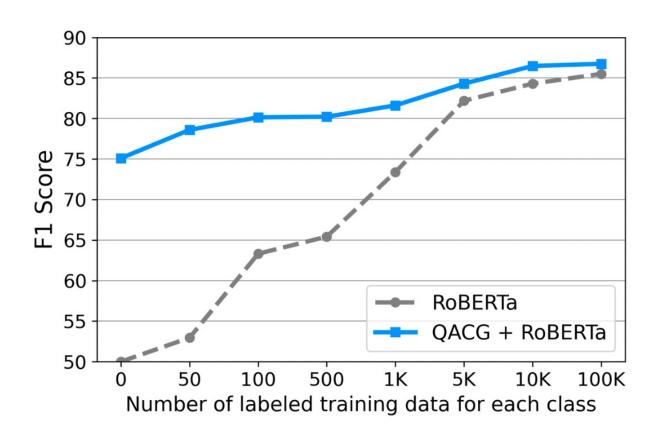
Table 1: Fact verification performance for supervised models and zero-shot models on three different settings.





- When transferring from FEVER to FEVER-Symmetric, the F1 drops by 10.0
- However, if we train the model with the generated data, the transfer only results in 1.0 F1 drop.
- Training with generated claims brings more robust fact checking model (preliminary results)





Discussions

- Our model requires a good question generator.
- To generate deep claims, you need to generate deep questions.

Model	Answerable Rate	FV Performance $P/R/F_1$
NQG++	63.0%	62.2 / 62.4 / 62.3
BART	89.5%	76.3 / 76.0 / 76.1

Table 2: Answerable Rate: the ratio of answerable questions generated by the NQG++ and the BART model. FV Performance: the zero-shot fact verification performance on the FEVER-Symmetric.

Evidence:	Roman Atwood is best known for his vlogs,	
	where he posts updates about his life.	
Claim:	Roman Atwood is a content creator.	
Evidence:	In 2004, Slovenia entered NATO and the	
	European Union.	
Claim:	Slovenia uses the euro.	
Evidence:	He has traveled to Chad and Uganda to raise	
	awareness about conflicts in the regions.	
Claim:	Ryan Gosling has been to a country in Africa.	
Claim:	Ryan Gosling has been to a country in Africa	

Table 5: Examples of claims in FEVER that require commonsense or world knowledge (underlined).

Thanks!

Any questions?

Liangming Pan

Email: liangmingpan@u.nus.edu