



School of Computing



Web Information Retrieval
Natural Language Processing Group



Zero-shot Fact Verification by Claim Generation

Liangming Pan, Wenhui 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

 DONALD TRUMP



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:**

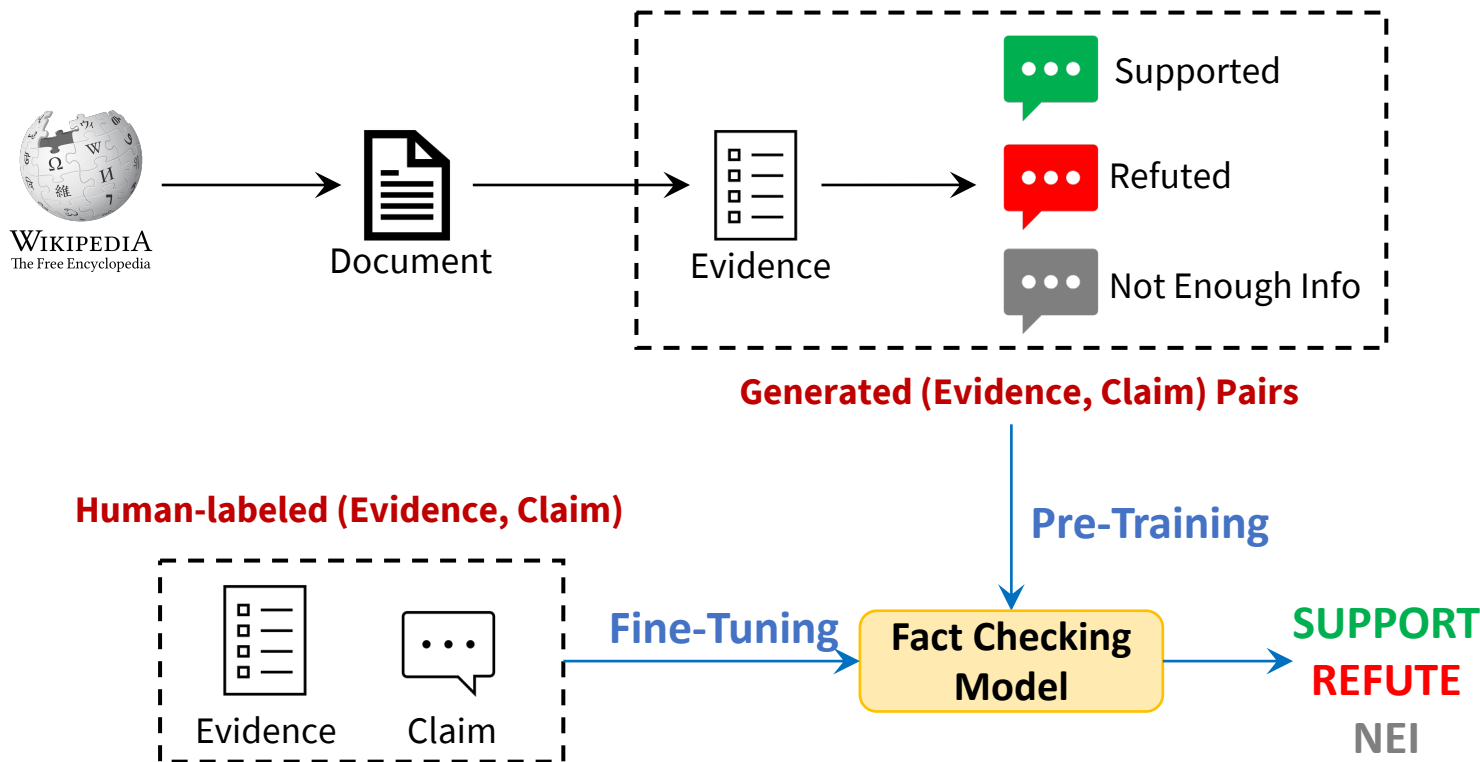
Supports, Refutes, Not Enough Info (NEI)

Motivation

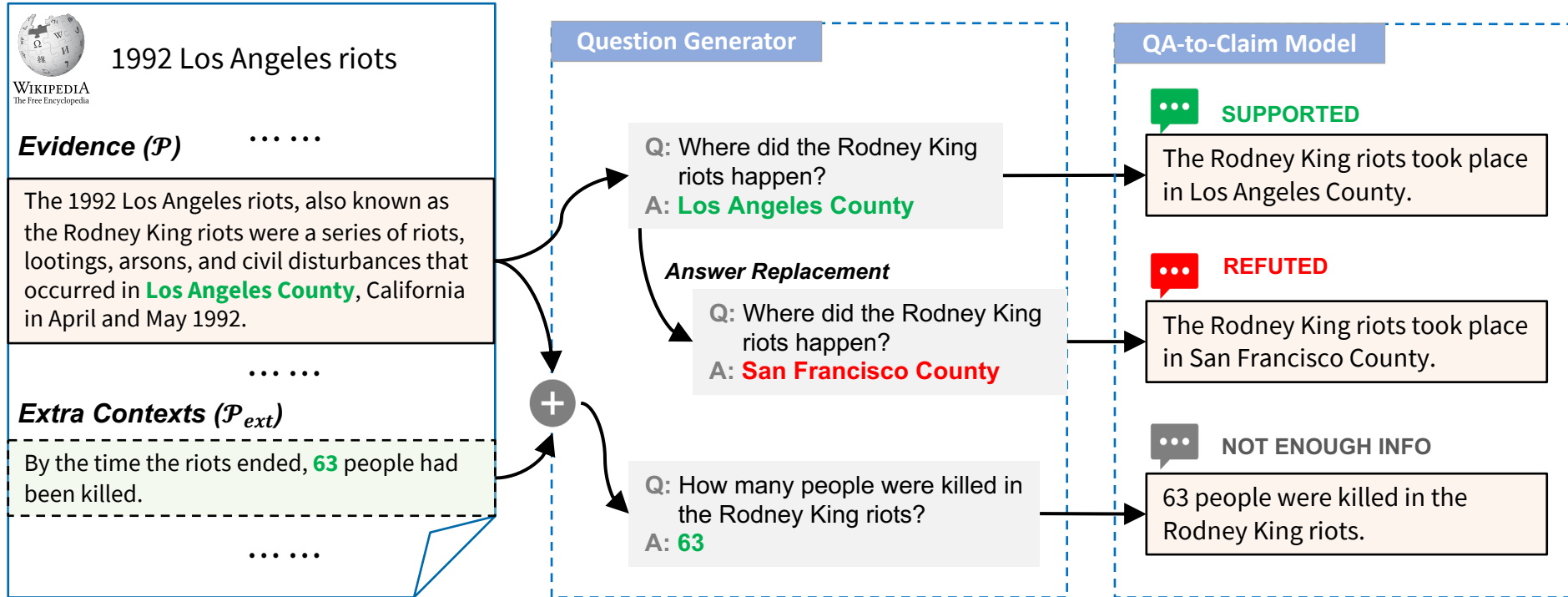
1) Training fact checking models **require large scale human-annotated (Evidence, Claim, Label) data**, such as FEVER.

- 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?

Zero/Few-shot Fact Verification



Claim Generation with QG



Examples of Generated Claims

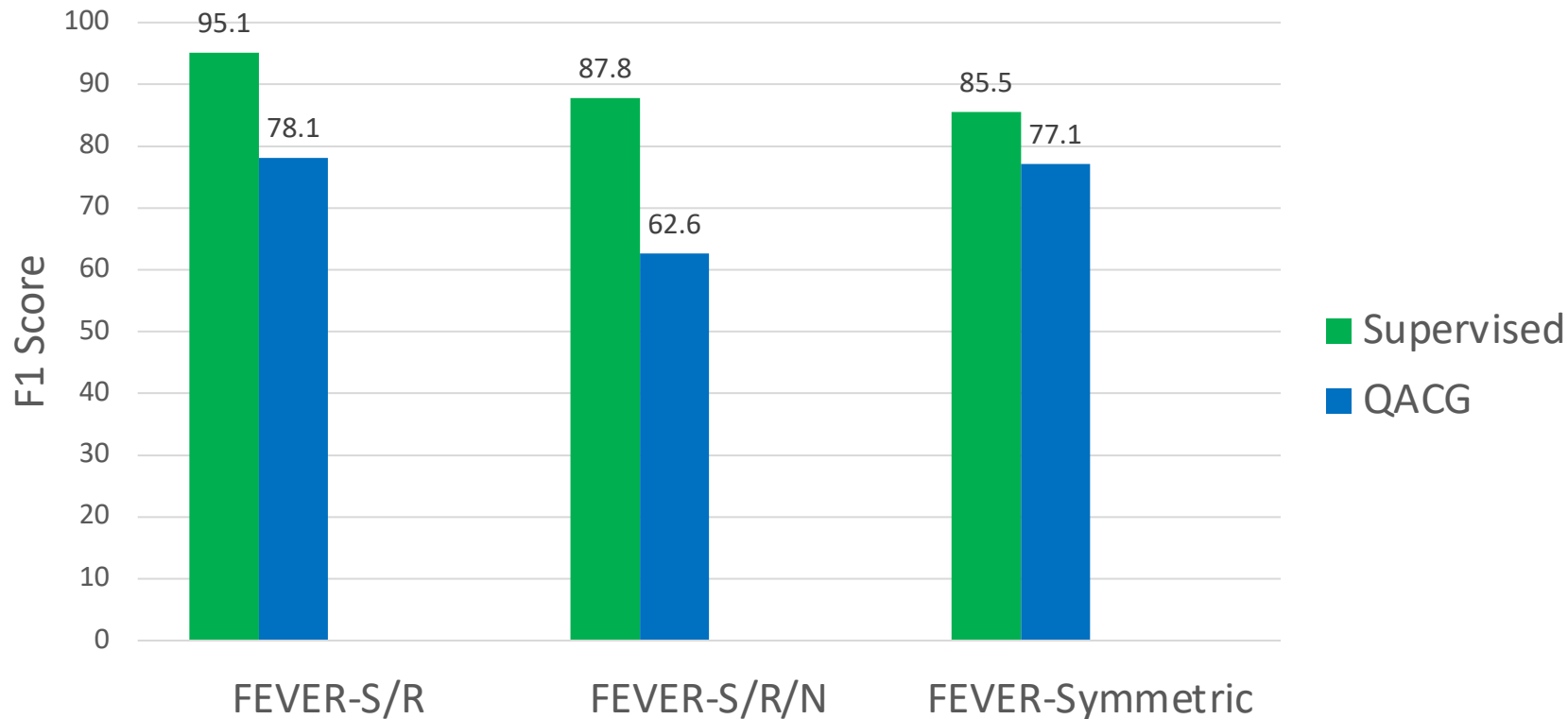
| Evidence | Generated Claim |
|---|--|
| <p>Budapest is cited as one of the most beautiful cities in Europe, ranked as the most liveable Central and Eastern European city on EIU's quality of life index, ranked as “the world’s second best city” by Conde Nast Traveler, and “Europe’s 7th most idyllic place to live” by Forbes.</p> | <p>SUPPORTED claims</p> <p>Budapest is ranked as the most liveable city in central Europe. Budapest ranks 7th in terms of idyllic places to live in Europe.</p> <p>REFUTED claims</p> <p>Budapest ranks in 11th in terms of idyllic places to live in Europe. Budapest is ranked the most liveable city in Asia.</p> <p>NEI claims</p> <p>Budapest is one of the largest cities in the European Union. Budapest is the capital of Hungary.</p> |
| <p>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.</p> | <p>SUPPORTED claims</p> <p>Bhatt won the Filmfare Award for Best Actress in Udta Punjab. Bhatt received the Filmfare Critics Award for her role in Highway.</p> <p>REFUTED claims</p> <p>Alia Bhatt won the Best Original Screenplay award in Highway. 2 States (2014) won Alia Bhatt the Filmfare Award for Best Actress.</p> <p>NEI claims</p> <p>Alia Bhatt made her acting debut in the 1999 thriller Sangharsh. Bhatt played her first leading role in Karan Johar’s romantic drama.</p> |

Zero-shot Fact Verification

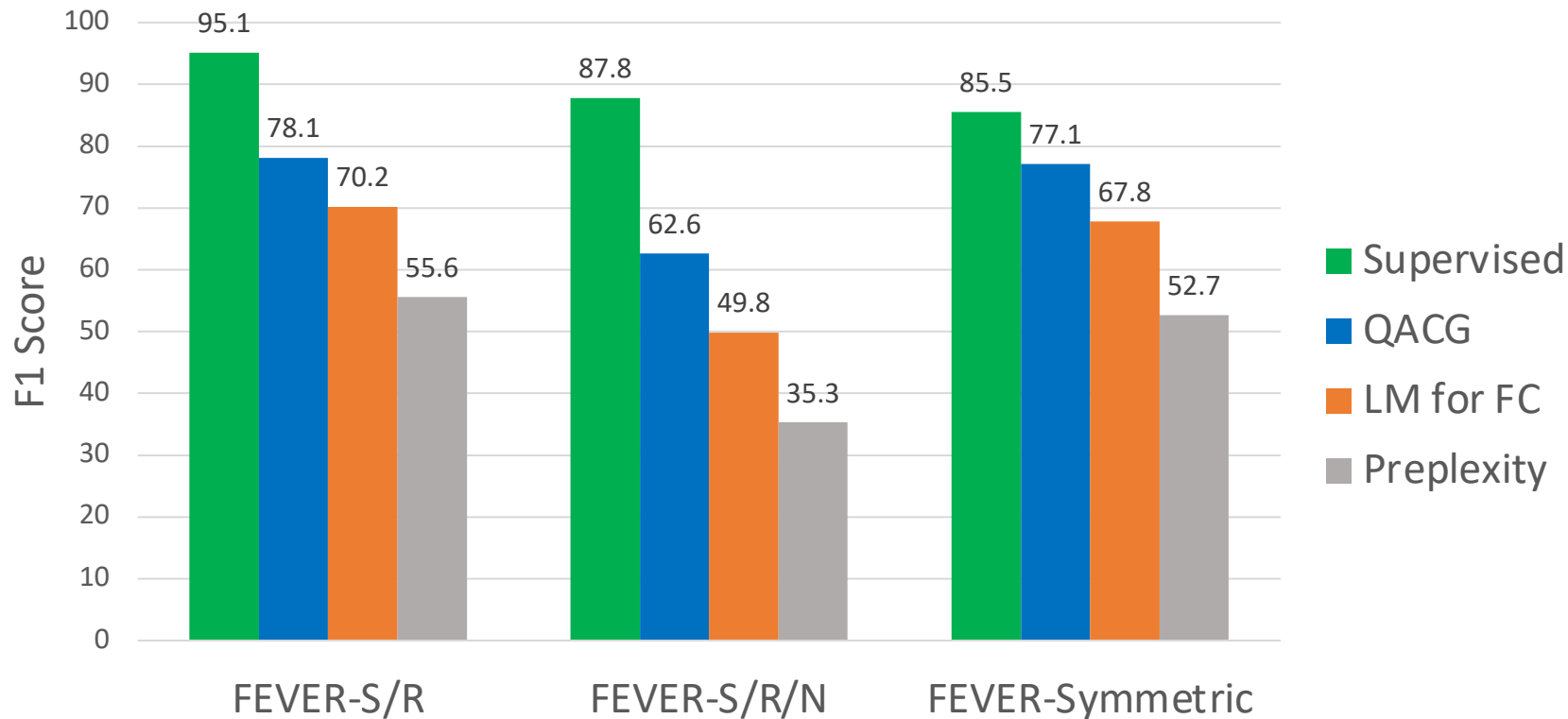
| Model | | FEVER -Symmetric | FEVER-S/R | FEVER-S/R/N |
|-------------------|--|---------------------------|---------------------------|---------------------------|
| | | $P / R / F_1$ | $P / R / F_1$ | $P / R / F_1$ |
| <i>Supervised</i> | S1. BERT-base (Devlin et al., 2019) | 81.5 / 81.3 / 81.2 | 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 |
| <i>Zero-shot</i> | 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.

Zero-shot Fact Verification

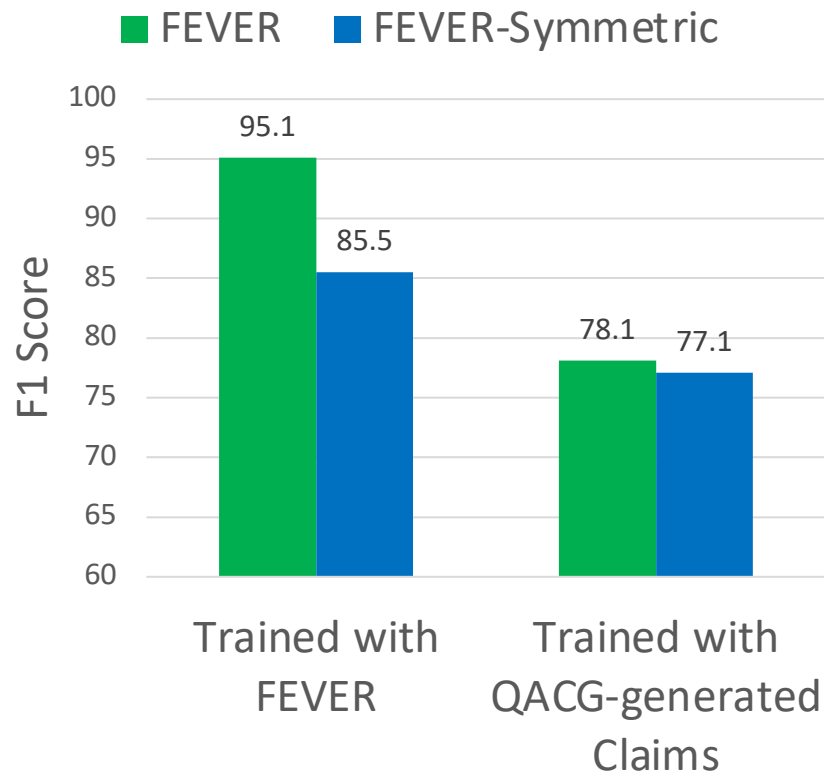


Zero-shot Fact Verification

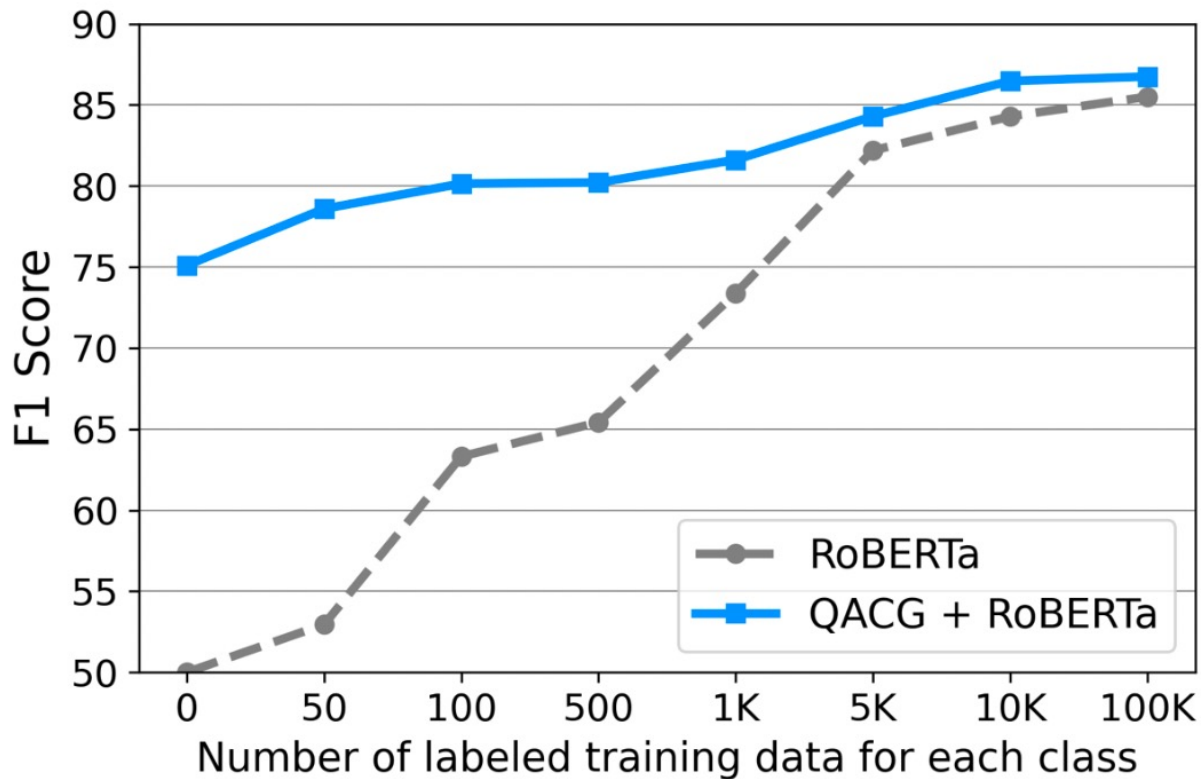


Zero-shot Fact Verification

- ❖ 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)



Few-shot Fact Verification



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

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

Thanks!

Any questions?

Liangming Pan

Email: liangmingpan@u.nus.edu