

Dataset Collection

We scrape political claims from PolitiFact. Our annotators reverse engineer the justifications by annotating questions that they think lead to the justification.

Claim Nathan Deal stated on September 9, 2016 in an education conference: "We have almost 68,000 Georgia students who are required by law to attend a chronically failing school."

Justification We ran the numbers -- and he's right. There are about 68,000 students in these failing schools. But his administration sets the standard, its numbers are out of date and there are other options. Opponents also claim some of these "chronically failing" schools have made gains in recent years.

Looks like whoever wrote the PolitiFact article was thinking: "Are Nathan Deal's numbers up to date?"

Annotator

Statistics: claims have one implied question on average, which has very low lexical overlap with the claim.

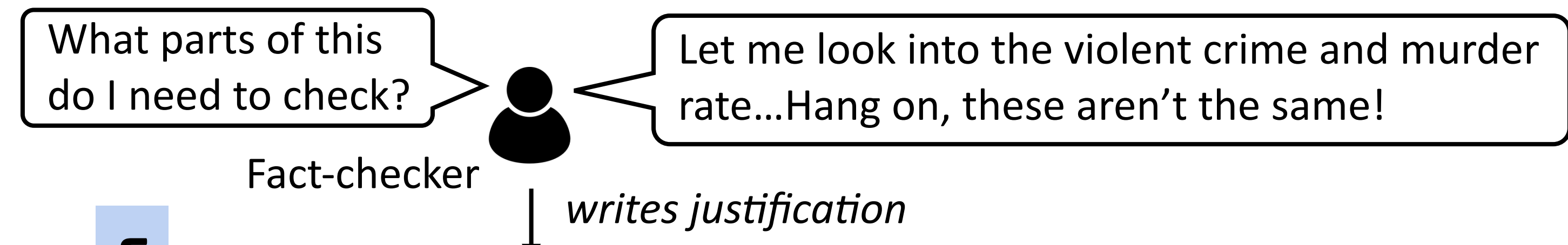
Type	# questions/claim	Rouge1-P	Rouge2-P	RougeL-P
Literal	2.2	0.56	0.30	0.47
Implied	1.0	0.28	0.09	0.22

Types of implied questions:

Domain knowledge (38.8%)	<p>Claim: "When President Obama was elected, the market crashed ... Trump was up 9%, President Obama was down 14.8% and President Bush was down almost 4%. There is an instant reaction on Wall Street."</p> <p>Implied Question: Did Obama cause the stock market crash when he was elected? (Domain knowledge that the speaker is implying that the stock market is correlated with the election.)</p>
Context (37.6%)	<p>Claim: With voting by mail, "you get thousands and thousands of people ... signing ballots all over the place."</p> <p>Implied Question: Is there a greater risk of voting fraud with mail-in ballots? (Need to know the background that the claim is about the potential risks of mail-in ballots.)</p>
Implicit meaning (16.5%)	<p>Claim: Nancy Pelosi bought \$1.25 million in Tesla stock the day before Joe Biden signed an order "for all federal vehicles" to be electric.</p> <p>Implied Question: Were the stock purchases improper insider trading? (The claim implies this purchase is insider trading.)</p>

What happens when a fact-checker checks a complex political claim?

Claim Joe Biden stated on August 31, 2020 in a speech: "When I was vice president, violent crime fell 15% in this country. ... The murder rate now is up 26% across the nation this year under Donald Trump."



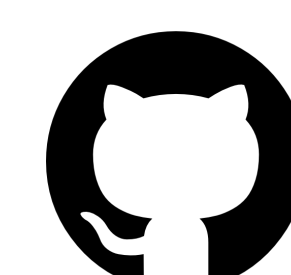
Justification The statistics themselves are correct. However, Biden didn't compare crime rates from the same time interval. The violent crime rate and murder rate are not directly comparable. Therefore, this is only half-true!

Checking the claim requires checking both explicit and implicit facets! We decompose a claim into a comprehensive set of yes-no questions:

Claim	Type	Description	Result
Can be directly derived from the claim	Literal	Did the murder rate in 2020 increase by 26% from 2019?	✓
		Did the crime rate fall by 15% during Joe Biden's vice presidency?	✓
	Implied	Is Biden comparing crime rates from the same time interval?	✗
		Is violent crime rate and murder rate directly comparable?	✗

Our contributions:

- An annotated dataset of decomposed claims into yes-no questions, including implicit aspects of the claims
- Automatic methods for claim decomposition
- Evaluation of automatic decomposition, claim veracity prediction, and evidence retrieval using decomposed questions



Dataset and code available at:

<https://github.com/jifan-chen/subquestions-for-fact-checking>

Automatic Claim Decomposition

We train T5-3b based question generators to produce a set of N questions given a claim c :

$$c, N \xrightarrow{\text{T5}} q_1 [S] q_2 [S] \dots q_N$$

Human Eval	Recall-all	Recall-Literal	Recall-implied
	0.58	0.74	0.18

- Most of the literal questions can be generated while only a few of the implied questions can be recovered. Still a hard problem!

Applications

Evidence retrieval: does using decomposed questions help us find evidence?

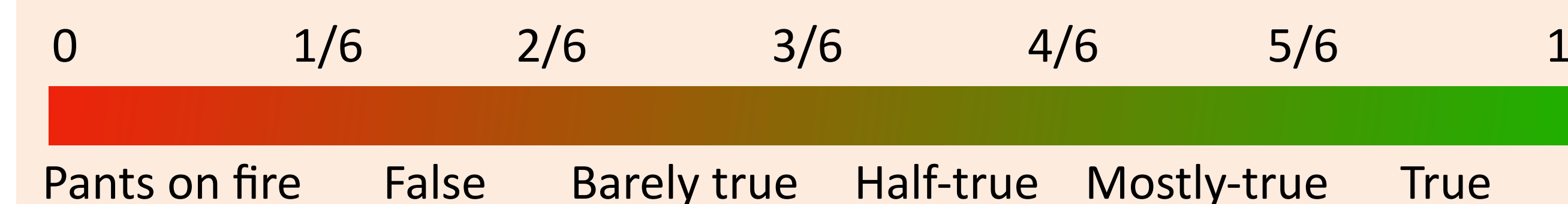
Simple test setting: can we retrieve the relevant evidence paragraphs from the PolitiFact article itself? We use NLI models to find sentences entailing statement version of subquestions.

Model	Decomposed predicted	Decomposed gold	Original claim
MNLI	41.0	48.8	35.2
NQ-NLI	38.8	34.5	40.9
DocNLI	44.7	59.6	36.9
BM25	36.2	47.5	39.2

- Decomposed questions are effective to retrieve evidence for both a BM25 model and NLI-based models.
- See paper for a comparison to Fan et al. (2020); they use wh-questions with a human-in-the-loop approach, whereas our questions are typically more specific.

Veracity: do answers to decomposed questions tell us the claim's veracity?

We compute a simple aggregated veracity score as the fraction of subquestions with yes answers.



Model	Macro-F1	Micro-F1	MAE
Question aggregation	0.46	0.45	0.73
Random	0.16	0.18	1.68
Most frequent	0.06	0.23	1.31

- Simple question aggregation outperforms simple baselines.
- Shows the potential of building explainable fact-checking models using the decompositions.