EXPLAINABLE FACT-CHECKING THROUGH QUESTION ANSWERING

Jing Yang[1], Didier Vega-Oliveros[1], Tais Seibt[2], Anderson Rocha[1]
[1] Artificial Intelligence Lab. (Recod.ai), Institute of Computing, University of Campinas, Brazil
[2] Department of Journalism, Universidade do Vale do Rio dos Sinos, Brazil

Paper ID: 3152

Problem

To mitigate the impact of mis/disinformation, many researchers have proposed automated fact-checking methods. However, most fact-checking methods cannot explain the reasoning behind their decisions, failing to build trust between machines and humans [1].

Contributions

Inspired by the QA works in checking factual consistency of documents and their summaries [2], we address fact-checking explainability through question answering. By breaking down automated fact-checking into several steps, our method allows for more detailed analysis of their decision-making processes. We compare the proposed method with several baselines, achieving state-of-the-art results in addition to adding explainability to the fact-checking process. In summary, our contributions are:

- A novel pipeline for using question answering as a proxy for explainable fact-checking;
- An answer comparison model with an attention mechanism on questions to learn their importance on the claims.

Methodology

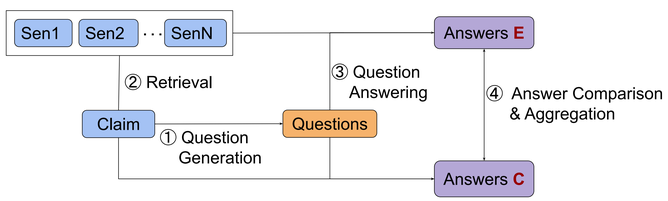

Fig. 1. Fact-checking pipeline integrating question generation and answering. The pipeline is composed of four main steps: (1) given a claim, generate multiple questions; (2) retrieve and re-rank evidence based on the claim; (3) for each question generated, obtain answer from claim and evidence respectively; (4) compare the answer pairs and transform the result into a label.

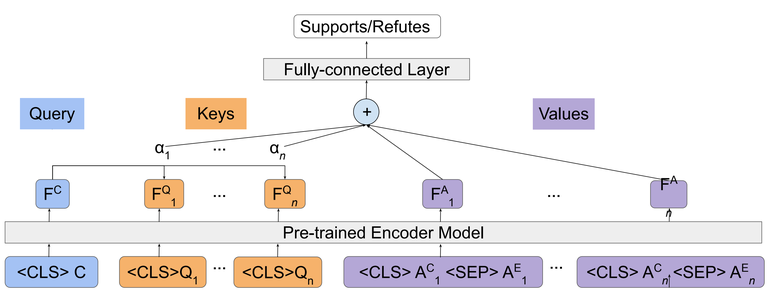

Fig. 2. Answer comparison and aggregation model with attention. C represents a given claim, Q_i represents i_{th} question, and (A^C_i , A^E_i) represents i_{th} answer pairs for claim and evidence. *n* denotes the number of questions and answer pairs.

Results and Analysis

Table 1. Label accuracy of different methods. `X-AI' denotes Explainability capabilities.

Methods	**Dev Acc**	**Test Acc**
Blackbox (**No X-AI**)	**76.17±1.23**	**74.58±1.66**
QUALS	56.12	56.01
BERTscore	58.68	62.32
Cosine similarity	61.16	62.75
F1-score	64.07	63.77
Attention C-Q-AA (**Ours, X-AI**)	75.44±0.52	73.43±0.83

Table 2. Ablation study of the model without attention

Inputs	**Dev Acc**	**Test Acc**
C	59.15±1.22	61.57±1.67
Q	56.22±1.37	56.90±0.90
AA	74.15±1.33	72.61±1.04
Q-AA	74.46±0.80	72.62±1.59
CQ-AA	74.88±0.81	72.89±1.14
Attention C-Q-AA	**75.44±0.52**	**72.43±0.83**

Claim: *Lee MacPhail passed away at his home at the age of 92.*
Evidence: *MacPhail lived in Delray Beach, Florida, where he died November 8, 2012, at his home. He was 95. At time of his death he was the oldest living Hall of Famer.*

Question	**Answer for claim**	**Answer for evidence**	**Attention weight (%)**
How old was MacPhail when he died?	*92*	*95*	12.17
Where did MacPhail die?	*at his home*	*Delray Beach, Florida*	4.29
Who died at the age of 92?	*MacPhail*	*no_answer*	1.66
When did MacPhail die?	*at the age of 92*	*November 8, 2012*	3.23
What was MacPhail's age?	*92*	*95*	13.19
At what age did MacPhail die?	*92.*	*95*	18.87
Where did MacPhail die?	*his home*	*Delray Beach, Florida*	4.29
What age was MacPhail when he died?	***92.***	***95***	**20.67**
At what age did MacPhail die?	*92*	*95*	18.87
Who died at age 92?	*MacPhail*	*no_answer*	2.75

Predicted label: REFUTES **Gold label**: REFUTES

Fig. 3. An example of our model generated questions, answer pairs, and attention weights. The question with the highest weight is in bold, and the second highest underlined. Although all answer pairs contain different words, the model is able to give more weight to the discrepancy between the claim and evidence, which is the age of the person when he died (92 vs 95).

Limitations

- Generating diverse and relevant questions aiming at the factuality of a claim is challenging.
- Answering correctly giving the context is a non-trivial and crucial step in the pipeline.
- A failing example:

> **Text**: Weber was born in Eutin, Bishopric of Lübeck, the eldest of the three children of Franz Anton von Weber and his second wife, Genovefa Weber, a Viennese singer.
> **Question**: How many siblings did Albert Weber have?
> **Answer**: three.
> **Correct answer**: two.

In the example, the model is not able to give the correct answer, because it is an extractive QA model, which is a limitation of this type of model.

Reasoning over text is a very challenging task; other ways of transforming the claim into a format like tabular data [3] may also help simplify the reasoning and thus improve performance.

Conclusions & Future Work

Conclusions:

- Our ablation study showed that the model can achieve near state-of-the-art performance with only information from answer pairs.
- Using QA, we can encourage the model to learn from more precise evidence; this can aid fact-checkers in better understanding models' decisions.

Future work:

- Add the retrieval step to the pipeline instead of using gold evidence.
- Answer questions directly from a more extensive set of document evidence.
- Work on more datasets to address the generalization capabilities of the method.
- have human evaluations on the questions and answers

Acknowledgment

Research funded by the São Paulo Research Foundation (FAPESP) under the Grants DéjàVu **#2017/12646-3**, **#2019/04053-8** and **#2019/26283-5**.

References

[1] Preslav Nakov et al., “Automated fact-checking for assisting human fact-checkers,” in International Joint Conference on Artificial Intelligence (IJCAI), 2021.
[2] Alex Wang, Kyunghyun Cho, and Mike Lewis, “Asking and answering questions to evaluate the factual consistency of summaries,” in Annual Meeting of the Association for Computational Linguistics (ACL), 2020.
[3] Vivek Gupta, et al., “Is my model using the right evidence? systematic probes for examining evidence-based tabular reasoning,” arXiv preprint, 2021.