COPA-SSE: Semi-Structured Explanations for Commonsense Reasoning

Ana Brassard¹,²  Benjamin Heinzerling¹,²  Pride Kavumba²,¹  Kentaro Inui²,¹

¹Riken AIP  ²Tohoku NLP Lab

Making explanation data is hard.
Free-form crowdsourcing?  —> Quality-control issues
Extracting from KGs?  —> Coverage/relevance issues

Our solution: semi-structured crowdsourcing!

I cut my foot on the floor. What was the cause of this?

a) The tile was cracked. (correct)
b) The tile was wet.

We then asked (different) crowdworkers to rate the explanations on a scale of 1 to 5 stars.

Finally, the explanations can be aggregated into graphs, serving as examples of relevant generated and/or extracted subgraphs.

Usage example | Automatic explanation scoring

- COPA-SSE has explanations w/ various quality ratings; can be used to create automatic scoring systems.
- Bigger T5 models⁶ had higher correlation with human ratings, but still much room for improvement!

Quick stats

- 9,747 explanations for 1,500 Balanced COPA⁵ questions
- 61% one-statement, 39% multi-statement
- 44% with 3.5 stars or more
- 98% of questions have at least one 3.5+ star explanation