Understanding Causality through Conceptual Explanation Generation

Athiya Deviyani, Mehak Malik, Prasoon Varshney
11-711 Advanced Natural Language Processing

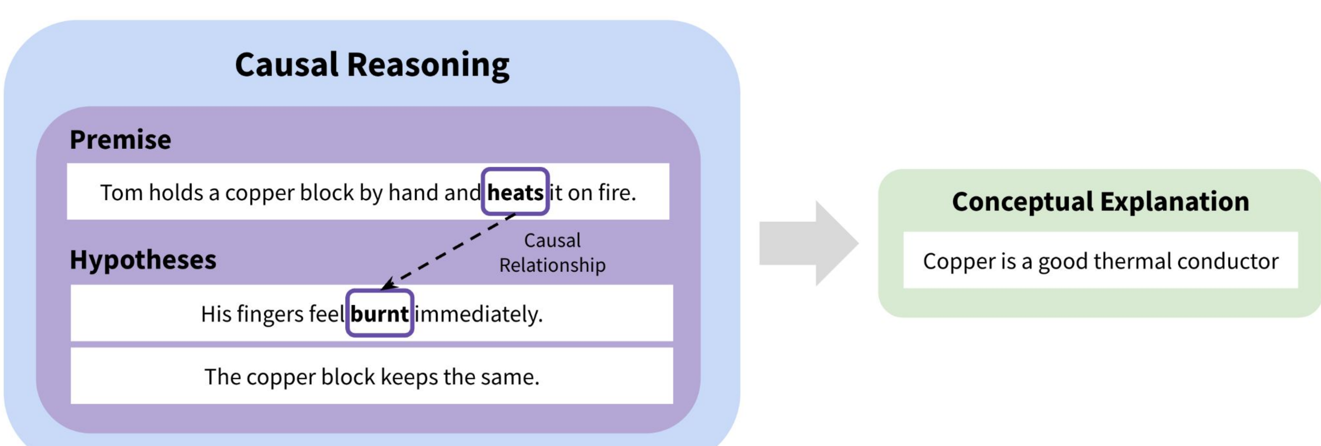
Motivation

Understanding causality has the potential to improve **robustness**, **fairness**, and **interpretability** of Natural Language Processing models.

Existing causal reasoning models lag behind humans, as **humans naturally have a deep conceptual understanding of causality** and can explain observed causal facts based on world knowledge.

Tasks

1. **Causal Reasoning (CR):** Given a premise and two hypotheses, choose the hypothesis that is related to the premise through a *cause* or *effect* relationship.
2. **Conceptual Explanation Generation (EQ):** Given a premise and the causally-related hypothesis, generate a conceptual explanation the sentence pairs are causally linked to each other.



Challenges

1. **Causal Reasoning (CR):**
   The second hypothesis need not be on another subject, it might be on the same subject and be related to the premise through some non-causal relationship.
2. **Conceptual Explanation Generation (EQ):**
   Generating the reason two entities are linked often requires external knowledge. For example, the explanation "Copper is a good thermal conductor" talks about a property of copper.

Dataset

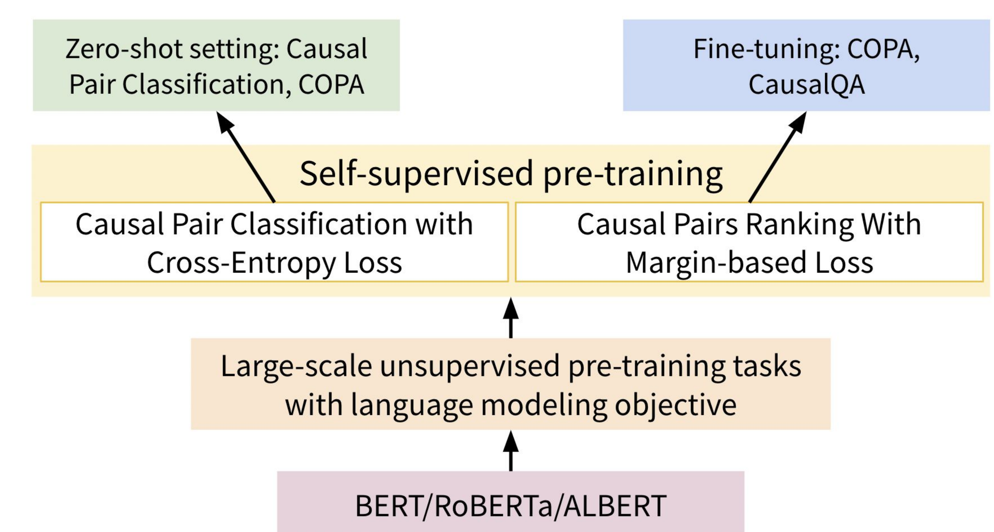
explainable CAusal REasoning (e-CARE) dataset [1]

- Contains over **21,000 causal reasoning questions** and over **13,000 unique conceptual explanations**
- Each instance consists of **two components**:
  (1) a multiple-choice causal reasoning question which contains a premise and two hypotheses
  (2) free-text-formed conceptual explanations to explain why the causation exists

Methodology

Causal Reasoning

CausalBERT [2] is a three-stage sequential transfer learning framework, solving two main issues in causal reasoning datasets: small scale and low quality



Conceptual Explanation Generation

Prompting

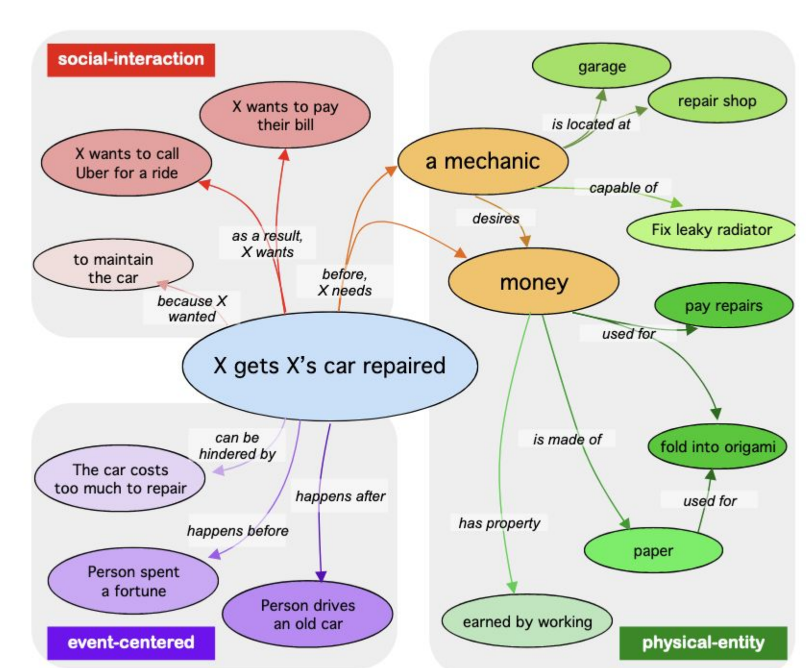
No prompt
"{cause} {effect} <generation>"

Words
"For cause {cause}, and effect {effect}, the explanation is that <generation>"

Special tokens
"<|cause|> {cause} <|effect|> {effect} <|explanation|> <generation>"

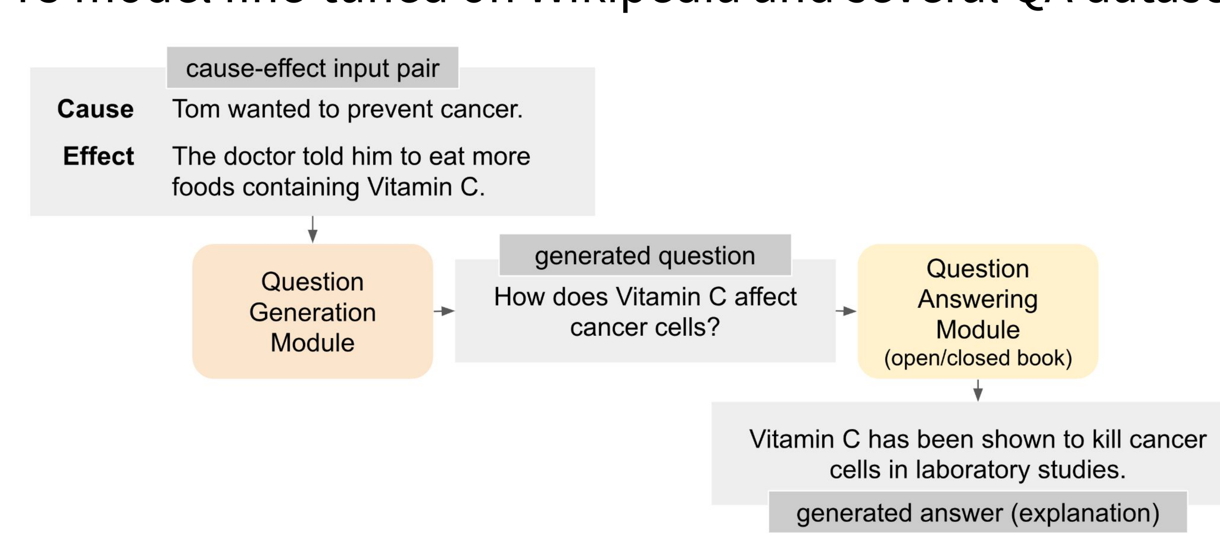
(COMET-)ATOMIC BART [3]

BART trained on ATOMIC 20-20 Which is a new Common Sense Knowledge Graph containing knowledge not commonly available in pre-trained models. The Common Sense transformer, or COMET trains transformers on example tuples from knowledge graphs.



Question Generation and Answering

Generate cause-effect questions using pre-trained ProphetNet for question generation [4] and perform closed-book question answering using a T5 model fine-tuned on Wikipedia and several QA datasets [5].



Results

Causal Reasoning

Approach	Accuracy
BERT (base, uncased)	76.78%
BERT (base, cased)	75.66%
CausalBERT	73.45%

We observe that **CausalBERT** did not improve the causal reasoning ability, even though that it is trained to make distinctions between causes, effects, and confounders, while BERT is not.

Conceptual Explanation Generation

Approach	Accuracy	AVG-BLEU	AVG-ROUGE	Perplexity
GPT2 [6]	-	0.26	0.40	10.90
+ P(W)	-	0.26	0.37	11.69
+ P(ST)	-	0.26	0.38	10.81
+ MT	71.54%	0.36	0.29	6.50
+ MT + P(W)	71.30%	0.37	0.30	6.42
+ MT + P(ST)	72.01%	0.36	0.35	6.62
BART	-	0.47	0.34	3.92
+ COMET	-	**0.52**	**0.41**	**8.41**

We found that **COMET-BART** which uses the Common Sense transformer performs best. We hypothesize that this is because the **causal pairs in e-CARE requires world knowledge** to deduce causal relationships between premises and hypotheses, and would **benefit from models which are trained on external knowledge bases.**

Premise	The worker fermented sugarcane.
Hypothesis	They got some rum.
Relationship	Premise-*Cause* Hypothesis-*Effect*
GPT2	Rum is a type of sugarcane made from sugarcane.
COMET-BART	Rum is made from sugarcane.

We can see that **GPT2 generates circular explanations**, while the explanations generated by **COMET-BART utilizes information outside of what is explicitly mentioned** in the premise and the hypothesis.

Future work

Reranking generative model top-k beam outputs based on an external model with a causal strength/Causal Explanation Quality (CEQ) objective.

[1] Du, Li, et al. "e-CARE: a New Dataset for Exploring Explainable Causal Reasoning." (2022)
[2] Li, Zhongyang, et al. "CausalBERT: Injecting Causal Knowledge Into Pre-trained Models with Minimal Supervision" (2021)
[3] Hwang, Jena D., et al. "(Comet-) Atomic 2020: On Symbolic and Neural Commonsense Knowledge Graphs" (2021)
[4] Qi, Weizhen, et al. "ProphetNet: Predicting Future N-gram for Sequence-to-Sequence Pre-training" (2020)
[5] Raffel, Colin, et al. "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer" (2020)
[6] Radford, Alec, et al. "Language Models are Unsupervised Multitask Learners" (2019)