Commonsense Knowledge Reasoning and Generation with Pre-trained Language Models: A Survey

Prajjwal Bhargava   Vincent Ng

Human Language Technology Research Institute
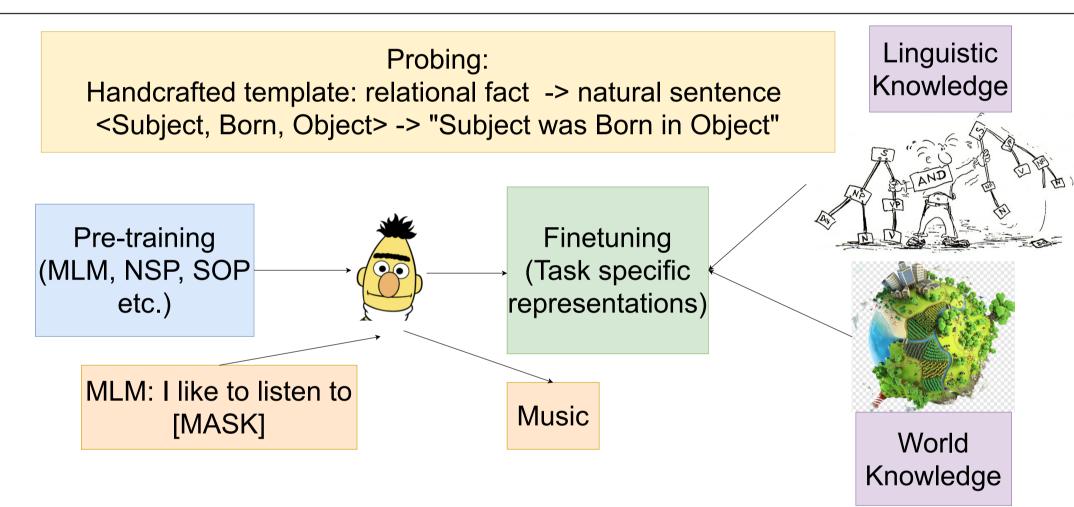The University of Texas at Dallas

What is Commonsense?

1. basic level of **practical knowledge** and **reasoning**
2. concerning **everyday life** and events
3. agreed upon by **majority** of people

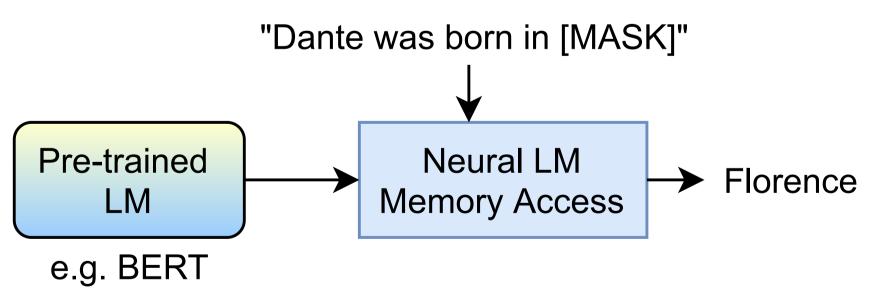## Research on Commonsense Knowledge Acquisition and Reasoning

▪ traditionally conducted in the knowledge representation and reasoning community
▪ recent surge of interest in the natural language processing (NLP) community

Goal: present a survey of (1) the kind of commonsense knowledge pre-trained language models (LMs) possess and (2) the extent to which this knowledge can be exploited for reasoning and generation

Pre-trained Language Models

Probing:
Handcrafted template: relational fact -> natural sentence
<Subject, Born, Object> -> "Subject was Born in Object"

Pre-training (MLM, NSP, SOP etc.)
MLM: I like to listen to [MASK]
Music
Finetuning (Task specific representations)
Linguistic Knowledge
World Knowledge

Capturing Commonsense Knowledge

"Dante was born in [MASK]"

Pre-trained LM — Neural LM Memory Access — Florence
e.g. BERT

▪ Recall 1:1 relations better than N-M.
▪ low correlation of predictions with that of humans
▪ posses ontological information "mango isA fruit"

Prompt	Model Predictions
A has fur, is big, has claws, has teeth, is an animal, eats, is brown, and lives in woods.	**bear**, wolf, cat, ...

Masked token predictions about **stereotypical assumptions** get refined as more properties are appended.

BIG **SMALL**

Can perform **Physical comparisons**

Lack of generalization on unseen entities ⟶ Memorization

Reasoning with Commonsense knowledge

Linguistic Reasoning:

▪ Concerned with understanding text
▪ Benchmark: WinoGrande

```
Birds can not [MASK] = Birds can [MASK]
Birds have [MASK] legs ⟶ BERT ⟶ 4
Performance ↓, complexity ↑
```

To separate egg whites from the yolk using a water bottle, you should...

a. **Squeeze** the water bottle and press it against the yolk. **Release,** which creates suction and lifts the yolk.

b. **Place** the water bottle against the yolk. **Keep pushing,** which creates suction and lifts the yolk.

a!   ???

Reasoning about Physical World:

▪ concepts based on the physical properties of objects
▪ Benchmarks: PIQA
▪ Learn explicit property associations (Water < Spoon)
▪ Struggle in understanding fundamental relations
▪ Dexterous at capturing affordances ("boats can be driven")

Abductive Reasoning:

▪ Determine most plausible explanation from incomplete observations
▪ Benchmarks: CosmosQA, Hellaswag
▪ Struggle in cross sentence representation and reasoning
▪ Not consistent with human commonsense
▪ Struggle in determining most plausible endings

Social Reasoning:

▪ Modeling mental states and likely actions
▪ Benchmarks: SocialIQA
▪ what/why > how
▪ Emotional > spatial

Why is [person4] pointing at [person1]?
a) He is telling [person3] that [person1] ordered the pancakes.
b) He just told a joke.
c) He is feeling accusatory towards [person1].
d) He is giving [person1] directions.

I chose a) because...
a) [person1] has the pancakes in front of him.
b) [person4] is taking everyone's order and asked for clarification.
c) [person3] is looking at the pancakes and both she and [person2] are smiling slightly.
d) [person3] is delivering food to the table, and she might not know whose order is whose.

Multimodal Reasoning:

▪ Textual representations have limitations
▪ Benchmarks: VCR, Visual Commonsense Graphs
▪ Higher quality commonsense inferences
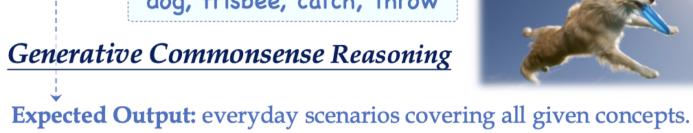
Temporal Reasoning:

▪ *attribute prediction*: understanding temporal dimensions
▪ *reason prediction*: how an event is temporally related
▪ Benchmarks: MCTACO, TimeDIAL
▪ Lack of KBs that incorporate notion of time
▪ LMs learn shallow cues about temporal patterns

Generating Commonsense Knowledge

▪ Knowledge Base Completion
▪ Constrained Commonsense Text Generation
▪ Benchmarks: Commongen, αNLG, COS-E

Concept-Set: a collection of objects/actions.
dog, frisbee, catch, throw
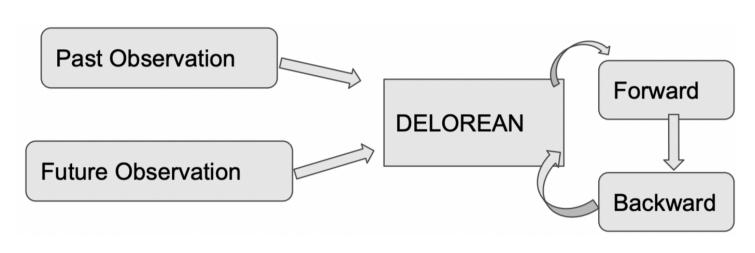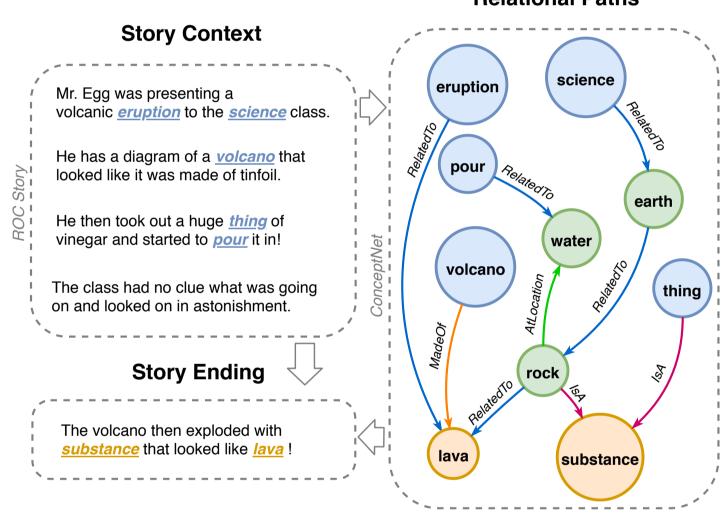*Generative Commonsense Reasoning*
Expected Output: everyday scenarios covering all given concepts.

X perceived as smart
Before, X needed to be a teacher
PersonX gives a tutorial

Challenges
▪ Poor coherency
▪ Insufficient concept coverage
▪ Limited reasoning capability

Addressing the Challenges
▪ Using prototypes
▪ Reasoning over multi-hop relational paths in knowledge graphs (KGs)
▪ Using iterative refinements

Past Observation — DELOREAN — Forward / Backward
Future Observation

Story Context
Mr. Egg was presenting a volcanic *eruption* to the *science* class.
He has a diagram of a *volcano* that looked like it was made of tinfoil.
He then took out a huge *thing* of vinegar and started to *pour* it in!
The class had no clue what was going on and looked on in astonishment.

Story Ending
The volcano then exploded with *substance* that looked like *lava*!

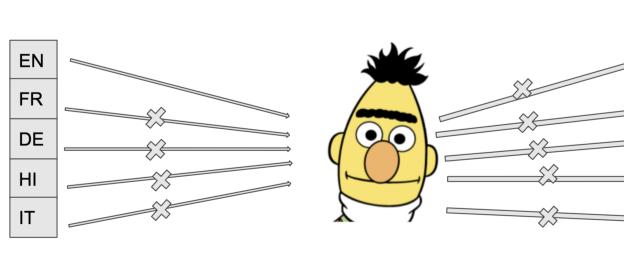**Relational Paths**

Future Directions

1. Improving benchmarks: Re-think what we are measuring in benchmarks. Important to highlight **"form"** and **"meaning"**.

Dataset	Model	Human	Dataset	Model	Human
hellaswag	93.85	95.6	winogrande	86.64	94.0
comosqa	91.79	94.0	socialiqa	83.15	88.1
piqa	90.13	94.9	vcr	63.15	85.0

2. Exploring new components of commonsense knowledge
3. Addressing reporting bias
4. Improving existing KGs (sparsity, non-contextualization)
5. Harnessing different modalities
6. Towards multilinguality

7. Reducing biases in benchmark datasets

AFLite
Adversarial Filterning
Reduce Lexical overlap
More complex context
More adversarial …..

EN NL
FR RU
DE UR
HI PT
IT VI