From Heuristic to Analytic: Cognitively Motivated Strategies for Coherent Physical Commonsense Reasoning

Zheyuan Zhang[1]   Shane Storks[1]   Fengyuan Hu[1]   Sungryull Sohn[2]

Moontae Lee[2,3]   Honglak Lee[1,2]   Joyce Chai[1]

[1]University of Michigan   [2]LG AI Research   [3]University of Illinois at Chicago

INTRO

Humans are thought to reason with dual processes [3]:

Heuristic processes give quick intuitive decisions for extracting relevant information.

Analytic processes further operate on relevant information to perform inference and rationalize.

Can the synergy of humans' dual processes enable more coherent reasoning in PLMs?

COHERENT PHYSICAL COMMONSENSE REASONING

TRIP [1]

Story A:
1. Mary went to the fridge.
2. Mary took out a bowl from the fridge.
3. The bowl had a cucumber and a donut in it.
4. Mary put the cucumber on the counter.
5. Mary ate the donut.

Story B:
1. Mary went to the fridge.
2. Mary took out a bowl from the fridge.
3. The bowl had a cucumber and a donut in it.
4. **Mary tossed the donut in the trash.**
5. **Mary ate the donut.**

Plausible story: **A**
Conflicting sentences: **(4, 5)**
States: inedible(**donut**) → edible(**donut**)

✓ *accurate*
✓ *consistent*
✓ *verifiable*

Tiered-ProPara [2]

Story A:
1. Air is brought in through the mouth.
2. Passes through the lungs.
3. And into the bronchial tissue.
4. The *carbon dioxide* is removed.
5. The lungs bring the oxygen to the rest of the body.

Story B:
1. *Carbon dioxide* enters the leaves through the stomates by diffusion.
2. Water is transported to the leaves in the xylem.
3. Energy harvested through light reaction is stored by forming ATP.
4. *Carbon dioxide* and energy from ATP are used to create sugar.
5. Oxygen exits the leaves through the stomata by diffusion. ...

Carbon dioxide conversion story: **B**
Carbon dioxide conversion sentence: **4**
Carbon dioxide conversion entity: **sugar**

HEURISTIC-ANALYTIC REASONING IN PLM IN-CONTEXT LEARNING (HAR-ICL)

HAR enables more coherent physical commonsense reasoning in in-context learning (ICL) with PLMs, outperforming traditional prompting approaches.

Language Model Prompt

Story A:
1. Mary went to the fridge.
2. Mary took out a bowl from the fridge.
3. The bowl had a cucumber and a donut in it.
4. Mary tossed the donut in the trash.
5. Mary ate the donut.

Story B:
1. Mary went to the fridge.
2. Mary took out a bowl from the fridge.
3. The bowl had a cucumber and a donut in it.
4. Mary put the cucumber on the counter.
5. Mary ate the donut.

Language Model Generation

"**Story B** is more plausible."

"In Story A, **sentences 4 and 5** conflict with each other."

"For sentence 4: After *Mary tossed the donut in the trash* ... the **donut** is now **inedible.**"

"For sentence 5: Before *Mary ate the donut* ... the **donut** was **edible.**"

Heuristic Decisions

Analytic Rationalization

InstructGPT

Approach	TRIP			Tiered-ProPara		
	Acc.	*Cons.*	*Ver.*	Acc.	*Cons.*	*Ver.*
ICL-U	70.9	40.7	7.1	54.9	17.4	5.2
ICL-CoT	75.0	40.7	10.8	50.7	19.2	7.5
ICL-HAR	72.6	**47.9**	**23.9**	54.9	**31.5**	**20.7**

LLaMA

Approach	TRIP			Tiered-ProPara		
	Acc.	*Cons.*	*Ver.*	Acc.	*Cons.*	*Ver.*
ICL-U	70.4	42.3	14.8	51.2	3.8	1.4
ICL-CoT	74.6	42.3	19.7	57.3	9.4	4.2
ICL-HAR	55.6	**44.4**	**35.2**	41.8	**17.8**	**13.1**

ATTENTION IN HAR-ICL

Increased coherence is due to more faithful attention to relevant language context while reasoning.

Sentence Selection Step

Approach	TRIP			Tiered-ProPara		
	Ratio	Prec.	Rec.	Ratio	Prec.	Rec.
ICL-U	0.96	42.6	39.6	0.90	14.8	30.6
ICL-HAR	**1.07**	**75.2**	**48.7**	**1.80**	**51.1**	**58.2**

Physical State Prediction Step

Approach	TRIP			Tiered-ProPara		
	Ratio	Prec.	Rec.	Ratio	Prec.	Rec.
ICL-U	1.23	43.0	35.4	1.21	14.6	25.9
ICL-HAR	**1.95**	**79.8**	**98.2**	**2.20**	**72.1**	**83.3**

ICL-U Attention

Story A: 41.0%
1. Coal is heated in the boiler.
2. The water tank over the boiler is heated.
3. Creates steam.
4. The steam is funneled to the piston.
5. Piston uses the steam as energy.
6. The piston causes the crankshaft to move.
Story B: 59.0%
1. Plates on the Earth's crust move slowly past each other.
2. As the plates move, they exert a great force.
3. When the force is large enough, the crust breaks.
4. The stress is released as energy.
5. The energy moves through the Earth in the form of waves.
6. We feel the earthquake.

Story A:
1. Tom found he is out of ice cream. 9.0%
2. Tom peeled a hard boiled egg. 5.5%
3. Tom sliced the egg with a knife. 4.6%
4. Tom washed the knife in the sink. 4.4%
5. Tom ate ice cream for dessert. 8.6%
Story B:
1. Tom poured a glass of milk. 10.4%
2. Tom peeled a hard boiled egg. 25.4%
3. Tom sliced the egg with a knife. 3.3%
4. Tom washed the knife in the sink. 16.2%
5. Tom ate ice cream for dessert. 12.5%

ICL-HAR Attention

Story A: 16.3%
1. Coal is heated in the boiler.
2. The water tank over the boiler is heated.
3. Creates steam.
4. The steam is funneled to the piston.
5. Piston uses the steam as energy.
6. The piston causes the crankshaft to move.
Story B: 83.7%
1. Plates on the Earth's crust move slowly past each other.
2. As the plates move, they exert a great force.
3. When the force is large enough, the crust breaks.
4. The stress is released as energy.
5. The energy moves through the Earth in the form of waves.
6. We feel the earthquake.

Story A:
1. Tom found he is out of ice cream. 21.3%
2. Tom peeled a hard boiled egg. 7.1%
3. Tom sliced the egg with a knife. 5.3%
4. Tom washed the knife in the sink. 4.4%
5. Tom ate ice cream for dessert. 15.4%
Story B:
1. Tom poured a glass of milk. 7.2%
2. Tom peeled a hard boiled egg. 4.0%
3. Tom sliced the egg with a knife. 2.4%
4. Tom washed the knife in the sink. 20.8%
5. Tom ate ice cream for dessert. 7.9%

HEURISTIC-ANALYTIC REASONING IN PLM FINE-TUNING

HAR enables significantly more coherent reasoning in fine-tuned PLMs and SOTA verifiability on TRIP.

Story A:
1. Air is brought...
4. The carbon...
5. The lungs...

Story B:
1. Carbon dioxide...
4. Carbon dioxide...
5. Oxygen exits...

PLM → Story Classifier → *Story:* **B**

Story B:
1. Carbon dioxide...
4. Carbon dioxide...
5. Oxygen exits...

PLM → Sentence Classifier → *Conversion Sentence:* **4**

PLM → Entity Classifier → *Converted Entity:* **sugar**
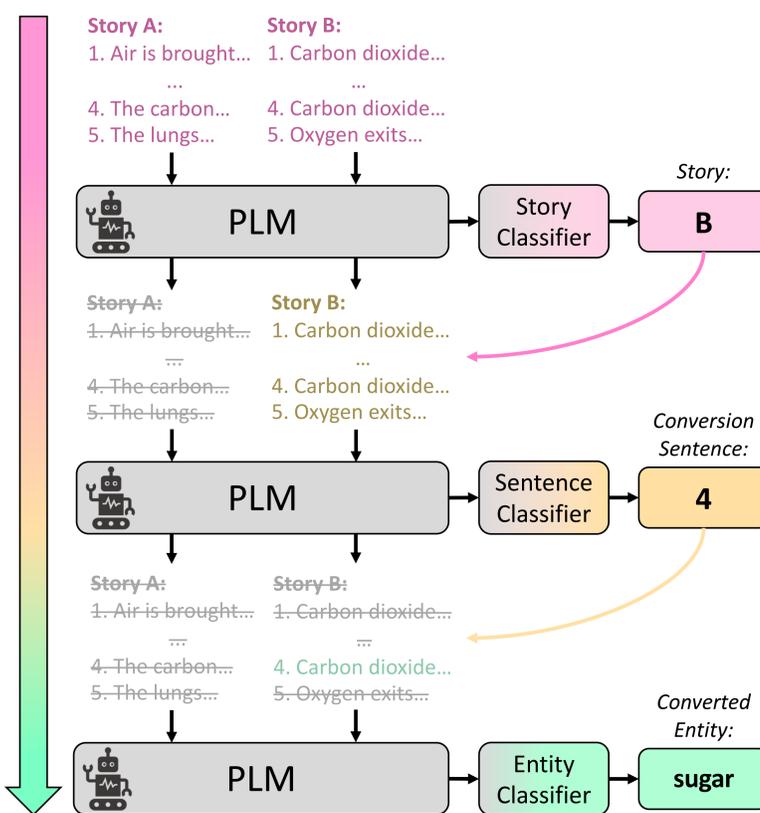
TRIP

Approach	Accuracy	*Consist.*	*Verif.*
RoBERTa	72.9	19.1	9.1
CGLI [4]	94.1	**77.3**	28.0
Breakpoint [5]	80.6	53.8	32.4
FCGLI	93.7	66.2	33.8
FCGLI-HAR	94.3	75.4	**41.1**

Tiered-ProPara

Approach	Accuracy	*Consist.*	*Verif.*
FCGLI	94.5	56.7	36.2
FCGLI-HAR	95.1	**83.6**	**57.4**
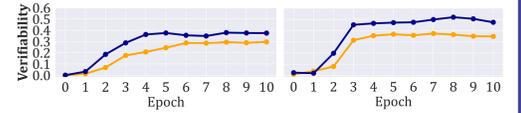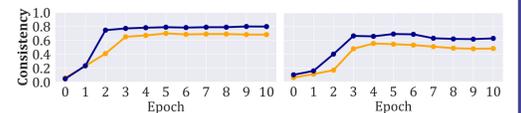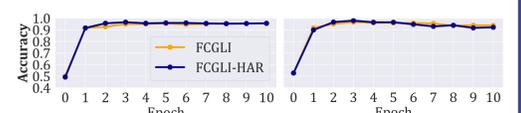
REFERENCES

[1] S. Storks, Q. Gao, Y. Zhang, and J. Chai. Tiered Reasoning for Intuitive Physics: Toward Verifiable Commonsense Language Understanding. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, 2021.

[2] B. Dalvi, L. Huang, N. Tandon, W.-T. Yih, and P. Clark. Tracking State Changes in Procedural Text: a Challenge Dataset and Models for Process Paragraph Comprehension. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2018.

[3] Jonathan St B. T. Evans. Heuristic and analytic processes in reasoning*. *British Journal of Psychology*, 75(4), 1984.

[4] K. Ma, F. Ilievski, J. Francis, E. Nyberg, and A. Oltramari. Coalescing Global and Local Information for Procedural Text Understanding. In *COLING*, Gyeongju, Republic of Korea, 2022. International Committee on Computational Linguistics.

[5] K. Richardson, R. Tamari, O. Sultan, D. Shahaf, R. Tsarfaty, and A. Sabharwal. Breakpoint Transformers for Modeling and Tracking Intermediate Beliefs. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 2022.

LINKS

PAPER   CODE