Teaching Probabilistic Logical Reasoning to Transformers

Aliakbar Nafar, Kristen Brent Venable, Parisa Kordjamshidi

Michigan State University, Florida Institute for Human and Machine Cognition

UncerteiNLP Workshop

EACL 2024 Findings

The challenge of Probabilistic Logical Reasoning

Logical Reasoning	Probabilistic Logical Reasoning
Dave is big. If someone is big, then they are green. If someone is green, then they are round.	Dave is big. Usually, If someone is big, then they are green. Normally, If someone is green, then they are round.
Conclusion: David is round.	Conclusion: David is round with a probability of 72%.

Learning Reasoning Heterogeneous

Motivation

- Importance of reasoning over uncertain text
  - Majority of rules in DBpedia are uncertain
  - Scientific content often utilizes hedges to express uncertainty

- Limitations of current language models in probabilistic logical reasoning
  - Coherent step by step reasoning
  - Arithmetic calculations

Problem Definition

- Context: Facts + Rules $((p_1, p_2, \ldots, p_n) \rightarrow q,\ Pr)$
- Question: what is the Probability of an inferable fact?
- Answer: A probability from 0 to 1

- Datasets: RuleBERT and our newly introduced RuleTaker-pro
  - Rules in RuleBERT
  - Rules in RuleTaker-pro → Context specific rules

Datasets

RuleBERT	RuleTaker-pro
(Fact 1) David is a cousin of Ann. (Fact 2) Mike is a child of Ann. (Rule 1, 0.90) If A is a spouse of B and C is a child of B, then C is a child of A. (Rule 2, 0.15) If A is a cousin of B, then A is a spouse of B.	(Fact 1) Dave is big. (Fact 2) Erin is sad. (Rule 1) Usually, If someone is big then they are green. (Rule 2) Normally, If someone is green then they are round. (Rule 3) Seldom, If someone is sad then they are round.
(Query) Mike is a child of David.	(Query) Dave is round.
Required Steps of Reasoning to Answer	
Fact 1 (1.00) & Rule 2 (0.15) $\Longrightarrow$ Fact 3: David is a spouse of Ann. (0.15) (Inferred) Fact 3 (0.15) & Fact 2 (1.00) & Rule 1 (0.90) $\Longrightarrow$ Fact 4: Mike is a child of David. (0.135) (Inferred) **Answer: 0.135**	Fact 1 (1.00) & Rule 1 (0.90) $\Longrightarrow$ Fact 3: Dave is green. (0.90) (Inferred) Fact 3 (0.90) & Rule 2 (0.80) $\Longrightarrow$ Fact 4: Dave is round. (0.72) (Inferred) **Answer: 0.72**

RuleBERT and RuleTaker-pro examples with their require steps of reasoning

Approach: Probabilistic Constraints Training (PCT)

- Motivation: following steps of reasoning
- Constraints:
  - Rule: $(p_1, p_2, \ldots, p_n) \rightarrow q$
  - Logical rules (previous work): $|1 - \min(1, \frac{P(q)}{P(p_1) * P(p_2) * \ldots * P(p_n)})| = 0$
  - Probabilistic logical rule constraint: $|P(q) - P(p_1) * P(p_2) * \ldots * P(p_n) * \boldsymbol{Pr}| = 0$

Required Steps of Reasoning to Answer	
Fact 1 (1.00) & Rule 2 (0.15) $\Longrightarrow$ Fact 3: David is a spouse of Ann. (0.15) (Inferred) Fact 3 (0.15) & Fact 2 (1.00) & Rule 1 (0.90) $\Longrightarrow$ Fact 4: Mike is a child of David. (0.135) (Inferred) **Answer: 0.135**	Fact 1 (1.00) & Rule 1 (0.90) $\Longrightarrow$ Fact 3: Dave is green. (0.90) (Inferred) Fact 3 (0.90) & Rule 2 (0.80) $\Longrightarrow$ Fact 4: Dave is round. (0.72) (Inferred) **Answer: 0.72**
Approach: Converting Probabilistic Reasoning Steps to Equality Constraints	
Constraint 1: P(Fact 1) * 0.15 = P(Fact 3) Constraint 2: P(Fact 3) * P(Fact 2) * 0.90 = P(Fact 4)	Constraint 1: P(Fact 1) * 0.90 = P(Fact 3) Constraint 2: P(Fact 3) * 0.80 = P(Fact 4)

Examples of constraint conversion

Approach: Probabilistic Constraints Training (PCT)

- Training

$$Loss = TaskLoss + \sum_{i=1}^{m} \lambda_j * C_i$$

Approach: Converting Probabilistic Reasoning Steps to Equality Constraints	
Constraint 1: P(Fact 1) * 0.15 = P(Fact 3) Constraint 2: P(Fact 3) * P(Fact 2) * 0.90 = P(Fact 4)	Constraint 1: P(Fact 1) * 0.90 = P(Fact 3) Constraint 2: P(Fact 3) * 0.80 = P(Fact 4)

Examples of constrains

- Inference

Learning Reasoning
Heterogeneous

Experiments

- Usefulness of textual rules in probabilistic reasoning
- To what extent does the baseline language model improvements from PCT
  - Probabilistic reasoning
  - Intermediate inferred facts
- Transferring the probabilistic reasoning capabilities
- Evaluation of Generative Large Language models

- Metrics: BA, CA1, CS1
- Models: RoBERTa Large, GPT3.5 and GPT4

Usefulness of Textual Rules in Probabilistic Reasoning

- Better results without the text of the rules
- Why?
- How about RuleTaker-pro?

	Roberta With Text of the Rules				
	M1	M2	M3	M4	M5
D1	76.9	79.8	79.9	70.7	64.9
D2	77.5	77.8	76.6	70.4	65.4
D3	78.4	76.9	76.2	78.8	71.6
D4	76.2	73.4	72.4	78.2	73.8
D5	77.1	73.0	69.6	77.5	78.1
	Roberta Without Text of the Rules				
D1	76.8	82.0	82.2	83.6	82.1
D2	75.4	78.8	78.2	80.0	78.5
D3	77.9	80.6	80.6	82.8	80.6
D4	75.0	76.2	77.2	79.6	77.0
D5	78.4	75.2	78.7	79.6	76.7

RuleBERT results with and without text of the rules with BA metric.

Improvements with PCT

- RuleBERT

	Roberta Without Text of the Rules				
	M1	M2	M3	M4	M5
D1	76.8	82.0	82.2	83.6	82.1
D2	75.4	78.8	78.2	80.0	78.5
D3	77.9	80.6	80.6	82.8	80.6
D4	75.0	76.2	77.2	79.6	77.0
D5	78.4	75.2	78.7	79.6	76.7
	Roberta + PCT				
D1	79.1	81.7	82.4	84.1	81.1
D2	78.5	79.7	77.3	80.9	77.7
D3	79.8	83.4	81.9	86.2	82.2
D4	77.4	81.4	80.2	85.1	81.3
D5	80.1	84.3	84.3	86.1	83.6

RuleBERT results with and without PCT with BA metric.

- RuleTaker-pro

	RoBERTa			
D/M	M1	M2	M3	Mmax
Total	38.2	38.3	20.4	33.8
D1	56.0	52.7	29.6	43.7
D2	36.4	38.2	20.3	32.8
D3	29.3	31.3	14.9	28.3
D4	27.4	28.5	14.0	27.1
D5	24.9	26.7	14.7	28.2
CS1	47.8	35.7	16.2	20.7
	RoBERTa + PCT			
Total	38.0	39.5	41.1	37.6
D1	53.3	50.8	50.5	46.9
D2	37.4	40.4	42.2	37.0
D3	26.4	32.9	36.0	32.4
D4	26.5	31.9	33.9	31.8
D5	23.3	30.4	33.4	31.4
CS1	44.9	42.6	34.5	35.2

RuleTaker-pro results with and without PCT with CA1 and CS1 metrics.

Learning Reasoning Heterogeneous

Transfer Learning

- Lower depth to higher depths questions
- Simple to complex examples
- Another domain

	CE			CE+PCT		
	M2	M3	Mmax	M2	M3	Mmax
S	39	20	34	41	40	37
C	34	18	32	36	38	36

RuleTaker-pro simple and complex examples evaluated separately with and without PCT with CA1 metric.

	Baseline RoBERTa		
	M2	M3	M5
D2	77.8	76.6	65.4
D3	76.9	76.2	71.6
D4	73.4	72.4	73.8
D5	73.0	69.6	78.1
	Augmented Data		
D2	76.8	80.6	83.4
D3	75.9	83.2	81.6
D4	70.4	76.4	74.8
D5	68.0	72.6	67.1
	Transfer Learning of PCT		
D2	84.8	84.6	72.4
D3	84.9	82.2	72.6
D4	84.4	77.4	73.8
D5	86.0	66.6	81.1

Transfer learning from RuleTaker-pro to RuleBERT with BA metric.

LLM Results

- RuleTaker-pro Results
- Additional results:
  - RuleBERT results
  - Chain of though attempt

	RoBERTa	GPT3.5	GPT3.5*	GPT4
D1	44	28	41	41
D2	33	20	26	27
D3	28	23	25	26
D4	27	18	20	17
D5	28	18	20	21

LLM results on RuleTaker-pro with CA1 metric.

Why LLM
failed here?

Learning Reasoning
Heterogeneous

Conclusion

- New Dataset: RuleTaker-pro
- Novel Probability constraints training (PCT)
- PCT helped Transfer Learning
- Generative Large Language Models failed

Future Work

- Dealing with uncertainty in Realistic domains
- Improving LLMs for probabilistic logical reasoning

Learning Reasoning
Heterogeneous

Questions?