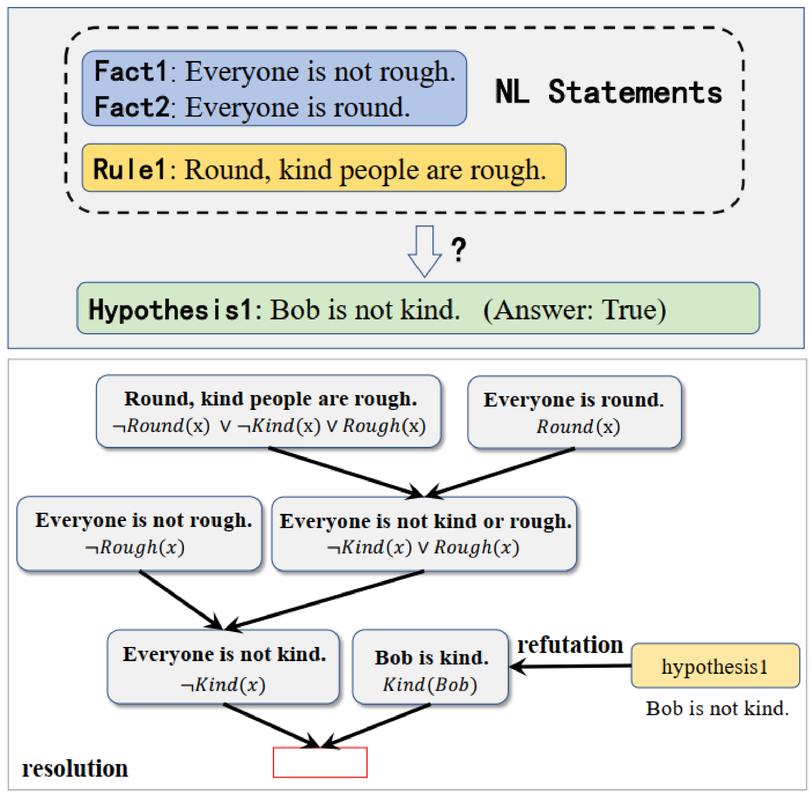
Towards Generalizable and Faithful Logic Reasoning over Natural Language via Resolution Refutation

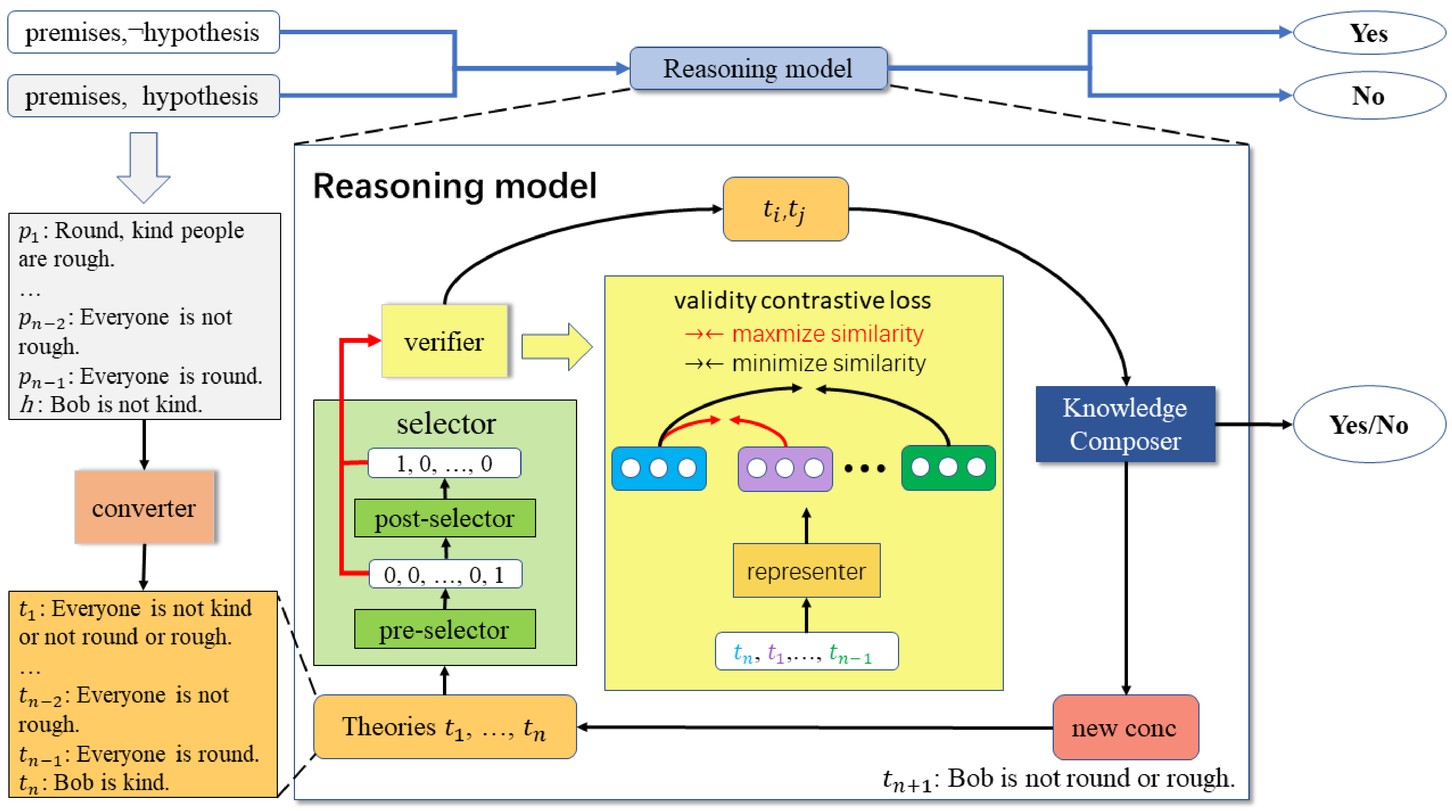
Zhouhao Sun[1], Xiao Ding[1], Li Du[2], Bibo Cai[1], Jinglong Gao[1], Ting Liu[1], Bing Qin[1]
[1]Harbin Institute of Technology, China [2]Beijing Academy of Artificial Intelligence, China

✿ Introduction

- **Previous works: forward or backward chaining based stepwise inference methods**
  - cannot generalize to complex reasoning scenarios
  - (because of) not complete
- **Resolution Refutation** is a **complete** reasoning method under first-order logic
  - all the hypotheses with determined labels can be inferred
- **GFaiR: Generalizable and Faithful Logic Reasoning over Natural Language via Resolution Refutation**
  - neuralize the symbolic reasoning process of resolution refutation to improve completeness
  - validity contrastive loss-based verifier to provide guarantees for resolution and improve faithfulness by reducing hallucinations



✿ Method



- Converter
  - augments the given natural language statements with the negated hypothesis for refutation
  - transforms the representations of the statements for the following reasoning steps.
- Pre-selector and post-selector *select two statements* for generating a new conclusion in the current iterative steps.
- The selection process is under the *guidance* of the validity contrastive loss-based verifier.
- Knowledge composer apply the resolution rule to the selected statements at the natural language level to generate a novel conclusion.

✿ Experiments

- **Generalization to Complex Logical Reasoning Scenarios**

Model	Ruletaker-3ext		Hard RT		Hard RT*	
	EA	FA	EA	FA	EA	FA
T5	97.7	—	57.3	—	57.5	—
Roberta	98.9	—	59.6	—	59.7	—
ChatGPT	56.5	42.8	57.0	2.7	38.9	6.9
IBR	98.9	98.1	59.6	12.1	59.7	29.6
FaiRR	99.0	98.4	14.1	12.2	41.1	39.8
NLProofs	**99.3**	**99.2**	14.3	13.8	41.8	41.4
GFaiR	98.1	98.0	**68.5**	**67.5**	**73.9**	**71.7**

- **Generalization to Complex Logical Reasoning Scenarios**

depth	FaiRR		NLProofs		GFaiR	
	EA	FA	EA	FA	EA	FA
N/A	99.4	99.4	99.4	99.4	96.2	96.2
0	100	100	100	100	99.9	99.9
1	99.5	99.2	99.9	99.9	99.5	99.5
2	98.4	96.0	99.0	99.0	98.2	97.9
3	93.1	84.8	94.1	93.4	95.8	95.1
4	88.8	77.3	79.5	77.2	94.2	92.5
5	78.7	67.8	69.6	57.3	94.2	91.9

- **In-domain Performance on Complex Reasoning Scenarios**

Model	Hard RuleTaker**	
	EA	FA
IBR	89.3	39.2
FaiRR	40.4	34.0
NLProofs	40.7	39.4
GFaiR	**92.2**	**92.2**

- **Ablation Study**

Model	Ruletaker-3ext		Hard Ruletaker*	
	EA	FA	EA	FA
FaiRR	99.0	98.4	41.1	39.8
FaiRR+	98.4	98.3	41.5	41.4
GFaiR-	97.5	97.2	72.4	68.6
GFaiR	98.1	98.0	73.9	71.7

Email: zhsun@ir.hit.edu.cn