Exploring Transitivity in Neural NLI Models through Veridicality

Hitomi Yanaka[1,2] Koji Mineshima[3] Kentaro Inui[4,1]
[1]RIKEN, [2]University of Tokyo, [3]Keio University, [4]Tohoku University

Abstract

Question: To what extent neural models can learn the systematicity in language from training instances?
- We analyze the systematic generalization ability of neural models in the domain of Natural Language Inference (NLI) on **transitivity inference** with synthetic and naturalistic datasets involving **veridicality**

Transitivity

Key concept for Systematicity [Fodor and Pylyshyn, 1988] in NLI
- If a model learns $A \rightarrow B$ and $B \rightarrow C$, it should compose the two and draw a new inference $A \rightarrow C$

Otherwise, a model must memorize every possible combination independently!

Dataset creation

- **Veridical inference** can easily compose transitivity inference at scale by embedding various inferences into clause-embedding verbs
- We select 30 veridical (e.g. *s knows P* entails *P*) / non-veridical (e.g. *s thinks P does not entail P*) verbs f from previous veridicality datasets [White+2018;Ross and Pavlick,2019] and embed synthetic/naturalistic inference s1→s2 into f

Synthetic datasets

veridical + synthesized Boolean inference (*and*, *or*, *not*)

f(s1): *Someone knew that* [*Bob found Tom, Jim and Fred*]
s1: *Bob found Tom, Jim and Fred*
s2: *Bob found Jim*

Naturalistic datasets

veridical + lexical/structural inference (SICK [Marelli+,2014])

f(s1): *Someone thinks that* [*a person is brushing a cat*]
s1: *A person is brushing a cat*
s2: *A person is combing the fur of a cat*

yes : Entailment
unk : Non-entailment

$f(s_1) \rightarrow s_1$	$s_1 \rightarrow s_2$	$f(s_1) \rightarrow s_2$
yes	*yes*	***yes***
yes	*unk*	***unk***
unk	*yes*	***unk***
unk	*unk*	***unk***

Evaluation protocol

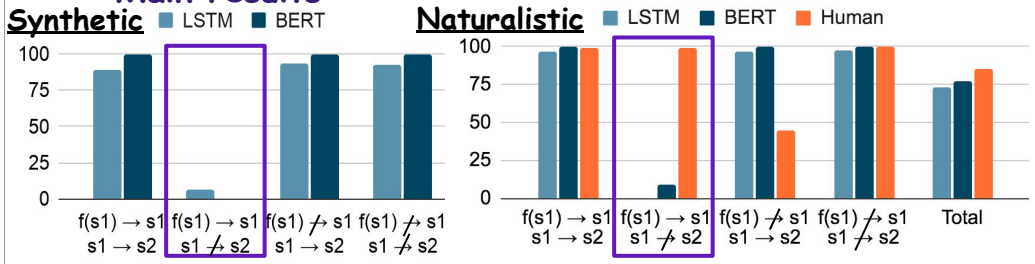
Training

Basic 1. veridical inference: f(s1) → s1
Premise: *Jo {**knew**/**hoped**} that Bob and Ann left.* [f(s1)]
Hypothesis: *Bob and Ann left.* [s1] (Entailment/Non-entailment)

Basic 2. various inference patterns: s1 → s2
Premise: *Bob and Ann left.* [s1]
Hypothesis: *Ann left.* [s2] (Entailment)

Test transitivity inference involving veridicality: f(s1) → s2
Premise: *Jo {**knew**/**hoped**} that Bob and Ann left.* [f(s1)]
Hypothesis: *Ann left.* [s2] (Entailment/Non-entailment)

Main results



- The models do not perform well on the cases where the verb f is veridical but embedded sentence s1 does not entail s2.
  ⇒ only look at the veridicality of f.
- Humans tend to fail on the cases where f is non-veridical and s1 entails s2.
  ⇒ only look at the relation between s1 and s2, neglecting the veridical verb. (Veridicality bias [Ross and Pavlick, 2019])

Conclusion

Current NLI models fail to consistently perform transitivity inference, lacking the generalization capacity for drawing composite inferences from provided training examples. There remains room for improving the systematic generalization capacities of models with respect to combining basic inferential abilities.

Contact: hitomi.yanaka@riken.jp
https://github.com/verypluming/transitivity