Title

Do Transformers Show Human-Like Sensitivity to Syntactic Embedding Depth?

Author

Hao, Hailin

Do Transformers Show Human-Like Sensitivity to Syntactic Embedding Depth?

Hailin Hao

University of Southern California, Los Angeles, California, United States

Abstract

Various probing studies have investigated the ability of neural language models trained only on word prediction over large corpora to process hierarchical structures. For instance, it's been shown that LSTM can capture long-distance subject-verb agreement patterns and attraction effects found in human sentence processing. However, although it's found in human experiments that syntactically closer attractors elicit greater attraction effects than linearly closer ones, LSTM shows the opposite, suggesting that it lacks knowledge regarding syntactic distance and is more sensitive to local information. The current article investigates whether state-of-the-art Generative Pre-trained Transformers (GPTs) can capture the prominence of syntactic distance. We experimented with various versions of GPT-2 and GPT-3 and found that all of them succeeded at the task. We conclude that GPT may better model human linguistic cognition than LSTM and that further investigating what mechanisms enable GPTs to do so can inform research on human syntactic processing.