Can neural networks identify speaker commitment?

Marie-Catherine de Marneffe

When we communicate, we infer a lot beyond the literal meaning of the words we hear or read. In particular, our understanding of an utterance depends on assessing the extent to which speakers are committed to the events they describe. An unadorned declarative like "The cancer has spread" conveys firm speaker commitment of the cancer having spread, whereas "There are some indicators that the cancer has spread" imbues the claim with uncertainty. When I say, "I don't think you should go", you take me to believe that you should not go. In this talk, I will investigate how well BERT, a current neural language model, performs on predicting speaker commitment of embedded events in English. I will show that, although BERT achieves very good results, it does so by exploiting surface patterns that correlate with certain speaker commitment labels in the training data, but it fails on items that necessitate pragmatic knowledge. These results highlight directions for improvement to build robust natural language understanding systems.