Summary 1: Climbing Towards NLU

Sam Showalter
University of California, Irvine (showalte)
showalte@uci.edu

1 Content Summary

This is a review of a position paper by (Bender and Koller, 2020) arguing that no language model can learn meaning if it has been trained solely on linguistic form alone. The authors define meaning as an encoding of intent (separate from language) using an expression, while form is "any observable realization of language." The authors feel this contribution is important because they believe these two terms and their relationship to **understanding** are conflated; some researchers declare state-of-the-art (SOTA) language models (LM) *understand* language, something the authors feel is an overstatement that clouds true progress towards Natural Language Understanding (NLU).

The core of the author's contribution is the notion that language provides a medium of expression by which one can communicate intent. This intent is not separate from language; the listener must infer intent from the meaning they extract from messages by grounding it in their contextual situation and world-view. According to the authors, LMs that haven't received perceptual information can't tie meaning to patterns they find in linguistic form, no matter how sophisticated. In some fields, this is known as the "Clever Hans" effect. The authors spin their own version of this with a hypothetical Octopus that knows no meaning but communicates purely using language patterns it observes.They follow this with truncated examples of LM objectives, including teaching a model to code in Java by exposing it to all of Github. They explain this is fundamentally impossible, as even human children do not learn language passively but by interacting with the world. Empirically they also cite studies where language form is perturbed in LM learning, causing performance for SOTA models like BERT to drop substantially.

The authors close by noting that some applications, like reading comprehension, may learn meaning, though not necessarily. They also preemptively refute counterarguments, by stating that even if some meaning is learned by these models, it is incomplete for full understanding. They feel their paper ensures that the field of NLP is "climbing the right hill" towards the goal of NLU.

2 Analysis

This paper was well written from a language-theoretic perspective. The authors, provocative in their position, were scrupulous in defending their claims. The flow of the paper was smooth, transitioning well through sections and culminating well in their final thoughts. One possible improvement would be to reduce some thought experiment content and replace it with additional concrete studies that support their claims. That would have made this more convincing for me.

Their topic, NLU, is certainly an important and fundamental goal in NLP. I think their claim that many researchers are "bottom-up" in their thinking about progress is correct; a position paper such as theirs can have high impact for that reason. While their contributions are theoretical and, to some extent, speculative, I believe it is certainly a novel stance. Regardless of their ultimate veracity, more researchers should approach NLU with a top-down mindset and LM capability is over-hyped.

Because this is a position paper, there is no evaluation of their methods beyond their own justification of their work. While compelling, I think the authors left a logical gray area in their supporting evidence and refutation. They cede that some applications *may* learn "something about meaning," that is incomplete in their closing arguments, which contradicts their original statement that there is "no way" this could happen. Their ideas, while stimulating, are perhaps not rigorous enough to stand up to all applications of language modeling - I think a discussion of edge cases of their claims would be beneficial.

In addition, a few ablations empirically aligned with the Octopus thought experiment would have led to a stronger paper. If LMs are fundamentally devoid of meaning, it should be simple to explicitly display examples where a lack of grounding causes a SOTA model to fail.

References

Emily M Bender and Alexander Koller. 2020. Climbing towards nlu: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5185–5198.