Climbing towards NLU:

On Meaning, Form, and Understanding in the Age of Data

Emily M. Bender, Alexander Koller

The paper addresses the gap between the claims made by popular papers in NLP and the interpretation of the same by the people around. From the perspective of the authors, the claims of models being able to “understand” or “comprehend” comes from the lack of understanding of the relation between linguistic form and meaning. The major issue with this is that it feeds into the AI hype. When reporting findings on a field that is watched keenly by the media and the world, it becomes necessary to showcase the results accurately.

Some of the information that LMs actually learn include English subject-word agreement, constituent types, dependency labels, NER, etc. A field as such has been created to understand the functioning and interpretations made by transformers like BERT and this field is often called BERTology. An analysis of the BERTology papers gave the conclusion that the LMs often learn aspects of the linguistic formal structure and do not actually have the ability to “reason”. To study these, datasets are often prepared in a way to frustrate the heuristics of BERT or other such transformers.

People speak to achieve some communicative intent and the meaning (M) can be defined as a cross product of the set of language expressions (E) and the corresponding communicative intent (I). Conventional meaning on the other hand is different from communicative intent in the sense that conventional meaning is something that is constant across all contexts of use of the expression. Conventional meaning represents the communicative potential (C) of an expression. When someone speaks, we could say they are trying to communicate the intent by using an expression that has a conventional meaning that would best suit to communicate his intent. The Chinese Room experiment can be used to understand to an extent what is happening in NLP systems. In the Chinese Room experiment, a person who does not speak Chinese is made to answer questions in Chinese by consulting a library of books according to predefined rules. So although to an outsider, it looks like learning and understanding, in actual just a manipulation of form is being done to suit the situation.

When we communicate in the real world, it is not sufficient to observe patterns and understand relations between M and C, but rather it is important to understand the inputs from the surroundings and not to forget reasoning. Say, take the case of a system where we train the language model with all the code written in Java in Github without pairing it with any sort of sample input or output. Now we ask this model to execute a sample Java program. Or we take the case where we train a language model on English text and provides it with a large collection

of unlabelled images with no connection between the photo and the text and we ask the model a bunch of questions based on the images. It is quite obvious that it is unfair to expect the model to give the right answers as they weren't trained for specific tasks. This is exactly what we do when we expect a model trained on text to understand the text given that no information was given to it on the communicative intent of the speaker or about any external entities.

One of the arguments put forth is that of human babies acquiring language meanings by just listening. However, it has been proved that toddlers pick up language not by observing language forms but rather by interacting with its care-givers. The authors accept that the arguments made by them do not negate the claims of certain works that use large language models or pre-trained embeddings like BERT for semantic parsing or comprehension tests. This is because certain tasks like comprehensions contain information beyond form.

Research in any field must be looked at with skepticism and a thorough analysis of success and failure is essential for the field to continue growing. A "top-down" questioning is necessary to see exactly where we are going rather than "bottom-up" where we are satisfied with the small successes seen in specific tasks. It is also important to be aware of the limitations of tasks. More appreciation must be given to tasks that are carefully created such as the DROP reading comprehension benchmark that uses more stringent tests of understanding. Also, when we see a model can understand or learn the meaning, we must understand that if that were the case then it must be true irrespective of the task. So the model must do well on multiple tasks. The final task is to develop machines that have Natural Language Understanding similar to that of humans and not to just improve baselines in a particular task.

There could possibly be several counter-arguments to the arguments made in the paper. The authors have addressed some of the thoughts or discussions they have had while writing the paper.

- First of all, "meaning" is not simply a relationship between form and syntax. But there is something else in it that helps the speaker communicate.
- When the form is augmented with some grounding data, then meaning can be learned to an extent that communicative intent is in the data itself.
- No matter how much data we have, there has to be a correlation to the real world. Because people can keep on coming up with new ways and forms of communicating and all that the model would do would be to memorize those.
- Neural representations do not have standing meanings nor do they have any communicative intent. If an unsupervised machine translation could reach accuracy levels that of supervised models, then we could say that the claims in this paper that meaning can't be learned out of form would be contradicted or we may reach the conclusion that machine translation does not require the model to understand the meaning.
- BERT has had significant improvement in meaning related tasks. It would have learned something about meaning, but the learning incomplete and the conclusions have been explored in papers related to BERTology.