

Dr. Vered Shwartz (Institute for Artificial Intelligence - AI2 and Paul G. Allen School of Computer Science & Engineering, University of Washington.)

Fast breaking and slow building of textual inference models.

Abstract

With the availability of massive training data for the task, natural language inference (NLI) has become the go-to task for testing natural language understanding. But a recent line of work showed that the current benchmarks over-estimate the actual performance on the task, and that models often perform well even when shown incomplete data, due to overfitting to benchmark artifacts. I will first demonstrate how state-of-the-art NLI models, pre- pre-trained LMs, failed on simple examples that required lexical inferences. I will then show that despite the contribution of pre-trained LMs to NLI models, there yet remain linguistic phenomena they fail to generalize to even when given direct signal for them.

In the second part of the talk, I will discuss a related task with a more realistic setup: cross-document event coreference resolution. Rather than strict entailment, in real world data, it may be more beneficial to be able to tell when multiple texts discuss the same event, even if in different specificity levels. I will present an ever-growing resource of predicate-level coreferences that we collect from news headlines in Twitter, and a neural model for joint cross-document entity and event coreference resolution.