A Quick Summary:
CommonsenseQA: A Question Answering Challenge Targeting Commonsense Knowledge

Original Paper: https://arxiv.org/pdf/1811.00937.pdf

24th Feb 2019

1 Ideas:

The authors describe a method to generate questions that require contextual knowledge and logical reasoning to answer. This does not seem much at first, but there are some very important reasons why this is quite crucial.

(a) The questions must not be able to be easily answerable simply by doing a web search.

(b) The answer choices are single words. This enables annotation artifacts to be removed. Annotation artifacts are linguistic characteristics of sentences that can make them seem "weird" to the machine (and thus gives them away as wrong answers).

2 Explanations:

(a) The questions ought to require background knowledge that is trivial to humans but is only seldom **explicitly** reported on the web. This is to ensure that an understanding/model of the environment is learned, and to prevent the model from just finding a naive way (e.g. just counting frequencies of answer words of web queries) to be learned.

(b) Annotation artifacts can be thought of as "weird" statistical features in sentences that seem unnatural to the machine (or humans) even if they are syntactically (grammatically) correct. **The presence of annotation artifacts enables the model to be trained to distinguish the correct answers even if the question is not seen.**

3 Results:

(a) Best performance was by OpenAI's GPT at 54.8%, way below human performance at 95.3%.

4 Notes:

(a) Generating questions in this way (crowdsourcing) is way too expensive for poor graduate students. Might it be possible to generate questions using a GAN instead? (Although traditionally it might be a little inadvisable to generate textual data).