Review of The Winograd Schema Challenge

The Winograd Schema Challenge proposes an alternative to the Turing Test with some advantages. The paper first introduces a few problems with the Turing Test in that it revolves around the central role of deception. Rather than trying to hold a conversation with the interrogator, the machine actually tries to trick them into believing they are a real person. In essence, the author states that the machine should be able to show us that it is thinking without having to pretend to be somebody. The example using ELIZA, a program included in some operating systems which fooled some people into thinking they were talking to a psychiatrist, brings an important supporting idea to the equation; that the deception works because we are quite forgiving in what we accept as legitimate conversation. Next, we run into the issue of subjectivity in the Turing Test, that if multiple interrogators looked at the same transcript they may come to different conclusions. Importantly, a test for machine intelligence should be objective with its results, and that is one way the Winograd Schema Challenge claims to be better than the Turing Test.

The author then introduces the Winograd Schema and its four features. The idea is to be given a sentence and then determine which word is being referred to. In addition, it should be a sentence in which one word, the special word, can be replaced to change the meaning of the sentence while still making perfect sense before asking which of the two is the correct word. Another way in which this test would be preferred over a Turing Test is the ability to fully automate the test. When actually testing, it is expected to randomly choose between the special word and its alternate. The reason this is considered a good quality for a test to have is that there is no subjective bias based on the examiner and anyone can test their own algorithm

without having to come up with many different people to act as an interrogator and the human in the Turing test. In addition to this, the Winograd Schema challenge also allows itself to be built upon with more and more questions to give a better, more accurate result of the participant.

There are however some limitations mentioned in the paper and ones that are not considered. One limitation is that some questions can have answers which are too obvious, i.e. "The women stopped taking the pills because they were <pregnant/carcinogenic>. Which individuals were <pregnant/carcinogenic>?" Obviously, the pills cannot be pregnant, and the women cannot be carcinogenic, but this is mentioned in the paper as something which should be avoided. On the other hand, it is also possible to have questions which are note obvious enough. Consider the example "Frank was pleased when Bill said he was the winner of the competition. Who was the winner?" The problem here is that it Frank could be pleased if his friend Bill won the competition, and likewise, he would be pleased if the found out through Bill that he had won the competition. This is another issue mentioned in the paper that should be avoided due to the ambiguity of the statements.

One issue that I have with this challenge is that I feel it does not fully pass as a test for human-level AI. Since this test can be seen as a task in and of itself, an agent being able to perform well on this task does not imply that it will perform well over a large set of environments. It would follow then that this challenge is only a good test for a good understanding of the English language and the ability to answer questions, much less be an indicator of high intelligence.