Andrew Godipelly

Professor Karen Mazidi

Human Language Technologies

13 November 2022

Reading ACL Papers Portfolio Assignment

For this portfolio assignment, I decided to read a paper written by Akari Asai from the University of Washington and Eunsol Choi from The University of Texas at Austin called "Challenges in Information-Seeking QA: Unanswerable Questions and Paragraph Retrieval". In this paper, Asai and Choi seek to address the issue of identifying the answerability of questions and what are some causes for unanswerability. For context: one of the most common things anyone does on the internet is ask questions in the hopes of finding an answer. In general, this is quite difficult since we need to search the entire web so instead, we will be looking at a far smaller problem of asking questions to a single document. In earlier work, we have seen that pretrained language models can outperform humans when it comes to reading comprehension, where the questions are written with an answer in mind. One of the more famous examples of such a model is Stanford's SQuAD. But when asked more general information-seeking questions, they really struggle. The authors noted that one way to view this problem is to break it into two smaller sub-problems: after given the question, the model must first identity whether any paragraph in the document is relevant to the question and then second, the model must determine if any one of those paragraphs actually has the answer to the question. But before even starting to search for an answer, the model should try and see if the question itself is answerable in the first place. In this context, unanswerable means that the document doesn't contain the answer or the information necessary to find the answer is spread across multiple paragraphs.

Asai and Choi wanted to find which sorts of questions and question formats are naturally more unanswerable compared to others. Some of their findings weren't too surprising such as the fact that if the question wasn't actually a question("bye and bye going to see the king by blind willie johnson"), that would clearly be unanswerable. They also found that certain keywords in a question would likely cause it to be unanswerable such as "which of the following would result in an snp?". Other times, they found that questions which require some level of gathering data from throughout the document like "top 10 best apple pie recipes" would also be unanswerable. All in all, the authors came up with 6 different categories that an unanswerable question could land in: factoid, non-factoid, multi-evidence, invalid questions, false premise, and invalid answers. Factoid questions are unanswerable because the model wasn't able to retrieve articles with answers available on the web("WHen is 'this is us' season 2 released on dvd"). A non-factoid question is unanswerable since the query is somewhat complex and whose answers are often longer than a single sentence and no one single paragraph has all the information to answer the question("What is the difference between a bernese mountain dog and a swiss mountain dog"). Multi-evidence questions require reasoning over a number of facts that are across 2 or more paragraphs or even documents("how many statues in India have at least one international border"). Invalid questions are questions that are simply too vague to answer("Where is apple production highest in our country"). False premise are questions that are based on false assumptions("What Harry Potter movie was released in 2008"[There was none. It was actually released in 2009]) And finally, invalid answers are just straight up errors. It is when the annotator missed an answer existing in the provided evidence document.

Some of the unique contributions of Asai and Choi are that they provide in-depth analysis on information-seeking question answering datasets such as Natural Questions and TyDi QA to

identify the remaining headrooms. Furthermore, they also show that answerability prediction paragraph retrieval still persists as a challenge even when the latest and greatest models are thrown at the problem. They also manually annotated reasons for unanswerability for 800 examples across 6 languages and the suggest some potential improvements for dataset collections and task design. When seeing how their models worked given varying levels of context, they used the metrics of precious, recall, and F1 to measure their work.

According to Google Scholar, Akari Asai has been cited around 1106 times and Eunsol Choi has been cited around 4390 times. Choi has the most citations in the list. I think one of the biggest parts of their work is the fact that they manually annotated 800 examples of unanswerability. Regardless of whether or not this paper is clever or helpful or anything else, the sheer fact that they were willing to put so much human effort into this paper is a real testament. And furthermore, I think this paper helps to further systematize and define what it really means for a question to be answerable or unanswerable

Works Cited

Asai, Akari, and Eunsol Choi. “Challenges in Information-Seeking Qa: Unanswerable Questions and Paragraph Retrieval.” Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021, https://doi.org/10.18653/v1/2021.acl-long.118.