

Question Answering

Review

Karan Praharaj

March 17th

Paper Summary

This paper proposes a new paradigm that helps in making data collection scalable. Their approach involves data collection by asking the annotator to write down what question a particular sentence answers and which sentence it emanated from. It introduces a corpus, DCQA (Discourse Comprehension by Question Answering), which captures both discourse and semantic links between sentences in the form of free-form, open-ended questions.

Summary of Strengths

- The corpus provided has great promise to provide high quality fodder for discourse comprehension training.
- The tables did a good job of distilling the text. I specially appreciated tables 3 and 5.
- The steps involved in the creation of the corpus, from start to finish, have been delineated very clearly.

Summary of Weaknesses

- The paper uses a recurring phrase namely "high cognitive load". I would have liked this to be explained more explicitly.
- The analyses provided by the paper are insightful. However, a little more analysis on the specific types of questions from the corpus that the models fails on, or even does well on, would make for an even stronger analysis. How does a model trained on the corpus do when evaluated on finer-grain divisions of evaluation sets by question type (e.g. polar, loaded, open-ended, binary etc.). There is already a useful table which shows performance stratified by distance between the question anchor and answer, but a further nested breakdown for the types of questions, would have helped even more.
- The paper has a fair few formatting errors, ranging from line spacing to some spelling mistakes, as has been acknowledged by the author in his presentation.

Comments/Suggestions/Typos

The corpus put forward in this paper could be extremely useful to the community, not only in giving more impetus to multiple sub-branches of text comprehension research, but also in aiding the development of applications that rely on deep understanding of the content, context and need complex reasoning abilities for good performance.

Overall Assessment

4 = This paper represents solid work, and is of significant interest for the (broad or narrow) sub-communities that might build on it

Reviewer Confidence

4 = Quite sure. I tried to check the important points carefully. It's unlikely, though conceivable, that I missed something that should affect my ratings.

Best Paper

Could this be a best paper in a top-tier *ACL venue?

I do not think it would be the best paper in a *ACL venue, but it certainly promises to be useful to the community for text comprehension in various applications.

Limitations and Societal Impact

Not mentioned.

Ethical Concerns

N/A

Needs Ethics Review

Not required.

Reproducibility

Because this project involves human annotation, reproducibility is almost impossible.

1 = They would not be able to reproduce the results here no matter how hard they tried.

Datasets

4 = Useful: I would recommend the new datasets to other researchers or developers for their ongoing work.

Software

1 = No usable software released.