Concise Answers to Complex Questions: Summarization of Long-form Answers

Abhilash Potluri*  
acpotluri@utexas.edu

Fangyuan Xu*  
fangyuan@utexas.edu

Eunsol Choi  
eunsol@utexas.edu

Motivation for Concise Answers

- Long-form question answering (LFQA) opens door to generate comprehensive answers to complex questions. But users prefer concise answers. Can we summarize long-form answers while still addressing questions?
- Recent study on discourse structure of long-form answers [1] found that nearly 40% of answer sentences contain non-essential information (e.g., providing examples, providing background information).

New summarization method: Extract-and-decontextualize

- Extractive summary often suffers from fluency issues.
- Decontextualization task rewrites a sentence in a document such that the sentence is intelligible when presented alone.
- We use an off-the-shelf decontextualization model [2] to rewrite the first sentence in extractive summaries (GOLD) to improve extractive summary (Decontext GOLD)

Human Evaluation: Results

- Decontextualization improves fluency of the gold extractive summaries (by almost 10% on sentences where decontextualization changed the sentence).
- GPT3 shows the strongest performance, followed by Decontext GOLD.
- Over 90% of questions had a functional (fluent + adequate + faithful) summary from at least one summarization system.
- Some long form answers are challenging to summarize (one example on the right, more in paper)

New Extractive Summarization Dataset for LFQA

- Annotators select a subset of answer sentences as "summary" for long-form answers.
- Question/answer pairs sourced from three datasets: ELI5 [3], NQ [4], WebGPT [5]
  - Total of 1134 examples (three-way annotated), with average long answer length of 118 tokens and average summary length of 35 tokens.
- We evaluate extractive and abstractive summarization models (Presumm, T5, Pegasus, GPT-3) on this new benchmark. Having question as an input always helps!

Human Evaluation: Setting

- Evaluated 175 question/long-form answers with Amazon Mechanical Turk.
- Compared four summarization systems: Pegasus, human-annotated extractive summary (GOLD), decontextualized GOLD summary (Decontext GOLD), GPT-3 (text-davinci-002).
- Out of 175 examples, decontextualization edited 63 sentences.
- Crowdsolver rated:
  a. Summary Fluency: is the summary grammatically correct with no incomplete references? Yes/No
  b. Summary Faithfulness: does the summary maintain the same main idea as the original long answer? Yes/No
  c. Summary Adequacy: does the summary provide a sufficient answer to the question? Yes / Partially / No

Human Evaluation: Results

- Decontextualization improves fluency of the gold extractive summaries (by almost 10% on sentences where decontextualization changed the sentence).
- GPT3 shows the strongest performance, followed by Decontext GOLD.
- Over 90% of questions had a functional (fluent + adequate + faithful) summary from at least one summarization system.
- Some long form answers are challenging to summarize (one example on the right, more in paper)

Human Evaluation: Setting

- Evaluated 175 question/long-form answers with Amazon Mechanical Turk.
- Compared four summarization systems: Pegasus, human-annotated extractive summary (GOLD), decontextualized GOLD summary (Decontext GOLD), GPT-3 (text-davinci-002).
- Out of 175 examples, decontextualization edited 63 sentences.
- Crowdsolver rated:
  a. Summary Fluency: is the summary grammatically correct with no incomplete references? Yes/No
  b. Summary Faithfulness: does the summary maintain the same main idea as the original long answer? Yes/No
  c. Summary Adequacy: does the summary provide a sufficient answer to the question? Yes / Partially / No

Question: What is the weather usually like in Australia?
Long Answer: Australia's climate is governed largely by its size and by the hot, sinking air of the subtropical high pressure belt. This moves north and south with Antarctica. But it is variable, with frequent droughts lasting several seasons - thought to be caused in part by the El Nino - Southern Oscillation. The climate varies widely due to its large geographical size, but by for the largest part of Australia is desert or semi-arid. Only the south - east and south - west corners have a temperate climate and moderately fertile soil. The north part of the country has a tropical climate, varied between tropical rainforests, grasslands, and part desert, grasslands, and part desert.

Example question, long form answer, and extractive summary of the answer.

References


Check out our code and data at: github.com/acpotluri/lfqa_summary