Question Answering In Context

Eunsol Choi*, He He*, Mohit Iyyer*, Mark Yatskar*, Scott Yih, Yejin Choi, Percy Liang, Luke Zettlemoyer*

*: the authors contributed equally

Overview

Question answering is a multi-turn task where you ask further questions based on what you have learned. We present a large-scale, multi-turn question answering dataset, simulating an information seeking dialogue. Each interaction is between a student and a teacher on a text. A student poses questions to learn as much as possible about a hidden Wikipedia text, and a teacher who answers the questions by providing short excerpts from the text.

Task Setup

<table>
<thead>
<tr>
<th>Given</th>
<th>Student</th>
</tr>
</thead>
<tbody>
<tr>
<td>Article / Section Title / Article Summary</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Teacher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section Text</td>
</tr>
</tbody>
</table>

Provide a feedback to the question!

Follow Up | Yes | No | Not a Yes / No

Question Analysis

- Manual analysis showed about half the questions are not factoid. (Asking more open ended question such as “was the show success?”)
- Also 80% of the questions are contextual, on the conversation history (40%) and to the document (60%)
- About 20% questions are unanswerable
- 40% questions have multiple possible answer spans in the text

Treemap visualization of the eight most frequent “Wh” word in the dataset, where the box area is proportional to number of occurrences

- There is a large gap (>20%) between human’s and the best model’s performance
- High human scores demonstrates high inter-annotator agreement.

Performance Analysis

- Better modeling of dialog history will help.