Decontextualization: Making Sentences Stand Alone

Eunsol Choi, Jennimaria Palomaki, Matthew Lamm, Tom Kwiatkowski, Dipanjan Das, Michael Collins

Overview

We isolate and define the problem of sentence decontextualization: taking a sentence together with its context and rewriting it to be interpretable out of context, while preserving its meaning.

This can support models for question answering, dialogue agents, and summarization often interpret the meaning of a sentence in a rich context and use that meaning in a new context.

Decontextualization Examples

**Document Title:** Croatia at the FIFA World Cup

**Paragraph:** Croatia national football team have appeared in the FIFA world cup on five occasions, since gaining independence in 1991. Before that, from 1930 to 1990 Croatia was part of Yugoslavia. Their best result thus far was reaching the 2018 final, where they lost 4-2 to France.

**Original**

The Croatia national football team’s best result in FIFA world cup thus far was reaching the 2018 final, where they lost 4-2 to France.

**Decontextualized**

Their best result thus far was reaching the 2018 final, where they lost 4-2 to France.

**Document Title:** the FIFA World Cup

**Paragraph:** Croatia national football team have appeared in the FIFA world cup on five occasions, since gaining independence in 1991. Before that, from 1930 to 1990 Croatia was part of Yugoslavia. Their best result thus far was reaching the 2018 final, where they lost 4-2 to France.

**Original**

The Croatia national football team’s best result in FIFA world cup thus far was reaching the 2018 final, where they lost 4-2 to France.

**Decontextualized**

Their best result thus far was reaching the 2018 final, where they lost 4-2 to France.

Decontextualization Tasks

**Task:** Given a Wikipedia page and a sentence inside it, rewrite the sentence such that it can stand alone.

Sometimes it is not feasible (e.g., a sentence in the middle of a narrative).

Sometimes it is unnecessary (e.g., first sentence in the Wikipedia page).

**Decontextualization Examples**

**Document Title:** Croatia at the FIFA World Cup

**Paragraph:** Croatia national football team have appeared in the FIFA world cup on five occasions, since gaining independence in 1991. Before that, from 1930 to 1990 Croatia was part of Yugoslavia. Their best result thus far was reaching the 2018 final, where they lost 4-2 to France.

**Original**

The Croatia national football team’s best result in FIFA world cup thus far was reaching the 2018 final, where they lost 4-2 to France.

**Decontextualized**

Their best result thus far was reaching the 2018 final, where they lost 4-2 to France.

**Dataset**

Sentences from English Wikipedia, specifically answer sentence from Natural Questions dataset and another sentence sampled from the same document. Train dataset 1-way annotated, evaluation dataset 5-way annotated.

<table>
<thead>
<tr>
<th># examples</th>
<th>Train</th>
<th>Dev</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>11,290</td>
<td>1,945</td>
<td>1,945</td>
</tr>
</tbody>
</table>

Decontextualization Measures

**Intrinsic Evaluation**

**Evaluation Measures:**
- Exact Match (on one of four references, feasible examples only)
- SARI-add (added token overlap F1)
- SARI-del (deleted token overlap F1)

**Models**

- Coref
- spanBERT model, Untrained baseline
- T5-base
- T5-1B

**Upper Bounds**

- Human
- Annotator

**Results**

- Well trained seq2seq model performs surprisingly well.
- Vanilla evaluation metric is challenging to tell apart good vs. bad decontextualization.

**Application 1: Decontextualization as preprocessing**

Use decontextualization to generate retrieval corpus for open domain QA. Using decontextualized sentences as a retrieval corpus provides a better retrieval performance than using the original sentence.

**Application 2: Decontextualization As Is**

For question answering, instead of showing answer highlighted in the original answer sentence, we present an answer highlighted in a decontextualized answer sentence.

**Q:** what is the most viewed video on YouTube in 24 hours?

**Original**

The most viewed music video in this time period is Taylor Swift’s Look What You Made Me Do.

**Decontextualized**

The most viewed music video within 24 hours of its release is Taylor Swift’s Look What You Made Me Do.

**Users preferred decontextualized answer sentence, for providing succinct yet informative answer to the question.**