Towards unsupervised language models for QUD prediction

Matthijs Westera, Universitat Pompeu Fabra

Why QUD prediction?

- **Question Under Discussion (QUD):** a set of relevant pieces of information that are jointly pursued [1,2].
- QUD is a very useful theoretical notion...
- ... but in practice QUD-based theories often require **explicit** questions to yield testable predictions.
- **Problem:** QUDs are almost always **implicit**.

Related work

- Applications of QUD-based theories:
  - Exhaustivity / scalar implicatures [6]
  - Negation [7]
  - Intonation [2,8,9,10]
  - Discourse coherence [2,10], cf. rhetorical relations [12]

Question prediction (among many):

- Visual question prediction [13]
- LearningQ (from online forums) [14]

QUD annotation:

- Some exploratory work [15]

Current approach

Data

- **LAMBADA** raw training data [3]:
  - 2.4K unpublished novels
  - 15M sentences (233M tokens)
  - around 1% (150K) ends with "?"
- Prefix sentences with tags `<say>`, `<ask>` based on punctuation (? vs ./). [2,10]

Results (for now...)

For what it's worth (some hyperparameter optim.)

- **Test perplexity per word** overall: 140.25
  - Questions only: 112.49
  - (i.e., model chooses right word as often as a 112-sided die.)
  - This isolated number doesn't mean much...
  - Except perhaps that questions are more predictable than statements.

Model (for now...)

Standard neural language model [4].

- Vocabulary: 50K×150 embeddings
- LSTM [5]: 2×500 units
- 30 epochs; backpropagate 130 tokens.

Example output

**Prompt:**

"I carefully opened the box and looked inside. **<ask>**"?

**Generated:** (most likely 3-5 word questions from random sample):

<table>
<thead>
<tr>
<th>How did you know?</th>
<th>How did you know that?</th>
</tr>
</thead>
<tbody>
<tr>
<td>you don't know?</td>
<td>how did you know that?</td>
</tr>
<tr>
<td>you don't know?</td>
<td>when did you?</td>
</tr>
<tr>
<td>you don't know?</td>
<td>what's this?</td>
</tr>
<tr>
<td>you don't know?</td>
<td>what's that?</td>
</tr>
<tr>
<td>you don't know?</td>
<td>are you sure?</td>
</tr>
<tr>
<td>you don't know?</td>
<td>can you see what?</td>
</tr>
<tr>
<td>... many generic questions, only a few &quot;correct&quot; ones.</td>
<td></td>
</tr>
</tbody>
</table>

Some open issues:

- **Explicit questions vs QUDs**
  - Are implicit and explicit questions sufficiently similar? **Suggestion:** Yes, but explicit questions are more difficult to predict.
  - Explicit questions may explicate only part of a QUD.
  - Not all 'questions' end with a '?'.

What sort of data to train on?

- Enough questions, sufficiently large, and sufficiently natural (so: switchboard, wiki, news)
- Movie subtitles? Not self-contained...
- In fiction virtually all questions are in reported speech...

What sort of data to evaluate on?

- More natural (crowdsourcable) task: [work in progress]
  - "which questions does this story evoke?"
  - Secondary effects (e.g., intonation, exhaustivity).

References


Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 715154). This paper reflects the authors’ view only, and the EU is not responsible for any use that may be made of the information it contains.

I also gratefully acknowledge the support of NVIDIA Corporation with the donation of GPUs used for this research.