Towards a quantitative model of 'Questions Under Discussion'
Matthijs Westera, Universitat Pompeu Fabra

Why a quantitative model of QUDs?

- **Question Under Discussion (QUD)** ([1]) is a useful explanatory notion.
- E.g. (exp. data from [2]):
  1. It is warm. This implies *it is not hot* for 75% of participants
  2. It is old. This implies *it is not ancient* for 17% of participants
- **Explanation:** 'is it warm or hot?' is a more natural QUD than 'is it old or ancient?', at least out of context.
- **Challenge:** QUD-based theories often require explicit questions to yield testable predictions. **But QUDs are almost always implicit.**

Idea: learn about implicit questions by observing explicit questions.

Models explored so far:

**Model 1. Recurrent neural network**
Standard natural network language model ([4]).
- Vocabulary: 50K+150 embeddings.
- Long Short-Term Memory ([5]): 2×500 units.
- 30 epochs; backpropagate 130 tokens.
Trained on data (right), with sentences ending in "?" prefixed by <ask>.

**Preliminary results**
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.)
**Example output**

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

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

- how did you know?
  - you don’t know?
  - you don’t think?
  - what are you doing?
  - what did you do?
  - where did you get it?
  - if you knew that?
  - so, what was it?
- are you sure?
  - how did you know that?
  - where are you?
  - what’s it?
  - what’s that?
  - isn’t it?
  - is there anything else?
- questions more predictable than statements?

Trained on data (right), with sentences ending
- **30 epochs; backpropagate 130 tokens.**

**Related work**

- Applications of QUD-based theories:
  - Exhaustivity / scalar implicatures ([6])
  - Negation ([7])
  - Intonation ([8,9,10]).
  - Interpreting experimental results ([11])
  - Discourse coherence ([2,10])
    (cf. rhetorical relations ([12])
**Question prediction** (among many):
- Visual question prediction ([13])
- LearningQ (from online forums) ([14])

Train and evaluation data:

**Training data**
- Only dialogue data contains sufficient questions.
  - Task-oriented dialogue? Restricted domain.
- **Current approach:** Extract dialogue from BookCorpus:
  - 75M sentences (1B tokens).
  - ~1% of sentences ends with "?"; all in dialogue.
  - Result: 850K dialogues (5+ turns): 140M words

**Evaluation data**
- QUD annotation is costly (e.g., [15]).
- Experimental data like (1)/(2): scarce and artificial.
- **Indirect but crowdsourcable method:**
  "which questions does this text evoke?"

Some open issues

- Are implicit and explicit questions sufficiently similar? 
  **Suspicion:** Yes, but explicit questions are more difficult to predict.
- Explicit questions often explicate only part of a QUD.
- Not all ‘questions’ end with a ‘?’.

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 715950). 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 Tobii Technology Corporation with the donation of eye-tracker used for this research.

I also gratefully acknowledge the support of Tobii Technology Corporation with the donation of eye-tracker used for this research.