[Work in progress]

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.**

Related work

Applications of QUD-based theories:

- Exhaustivity / scalar implicatures [6]
- Negation [7]
- Intonation [2,8,9,10].
- 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]

Idea: learn about *implicit questions* by observing *explicit questions.*

Models explored so far:

Model 1. Recurrent neural network

Standard **neural 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.)
- Questions more predictable than statements?

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?	*are you sure?*	
you don't know?	*how did you know that?*	
you're not sure?	*where are you?*	
you don't know what?	*what's it?*	
what are you doing?	*that's what?*	(more likely)
what did you do?	*I don't know?*	
where did you get?	*is there anything else?*	
you want to go?	*does it matter?*	
how did you know that?	*is that what you think?*	
so , what was it?	*can you see what?*	

... many generic questions, only a few 'correct' ones.

Model 2. Transformer neural network

- Stacked *attention* layers combining & transforming token repres..
- Pre-trained BERT [3]: state-of-the-art on many tasks.
  'Bidirectional Encoder Representations from Transformers'
- As *classifier*: distinguish actual question from 19 random questions.

Preliminary results

- Predict *next* question: ~40% accurate (chance level = 5%)
- Predict *preceding* question: ~50%

(• Predicting assertions: ~45%)

Predicting preceding questions harder than next questions?

Train and evaluation data:

Training data

- Only *dialogue data* contains sufficient questions.
  - Task-oriented dialogue? Restricted domain. ☹
  - Movie subtitles? Not self-contained.
- **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?*"

[Work in progress]

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

[1] Roberts (1996). Information structure in discourse [...]. OSU WPL.
[2] Van Tiel et al. (2016). Scalar diversity. JoS 33.
[3] Devlin et al. (2019). BERT: Pre-training [...]. Proc. of NAACL.
[4] Pytorch examples. https://github.com/pytorch/examples .
[5] Hochreiter & Schmidhuber (1997). Long short-term memory. NC9.
[6] Westera (2016). An attention-based explanation [...]. *SuB21*.
[7] Kruszewski et al. (2016). [...] Conversational Negation [...]. CL42.
[8] Rooth (1992). A theory of focus interpretation. NLS1.
[9] Westera (2018). Rise-fall-rise [...]. In Gutzmann et al.
[10] Büring (2003). On D-trees, Beans, and Accents. L&P26.
[11] Westera & Brasoveanu (2014). Ignorance in context. *SALT24*.
[12] Hunter & Abrusan (2017). Rhetorical relations and QUDs. *LENLS*.
[13] Mostafazadeh et al. (2016). Generating nat. questions [...]. ACL54.
[14] Chen et al. (2018). LearningQ: A large-scale dataset [...]. AAAI.
[15] Riester et al. (2018). Annotation Guidelines [...]. In Adamou et al.

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.