Meaning Representations: Recent Advances in Parsing and Generation

Juri Opitz

October 2020

Abstract

In this course we will visit recent research that – given a theory of symbolic meaning – aims at developing functions for generating natural language text from meaning representations (‘Generation’) and generating meaning representations from natural language text (‘Parsing’).

1 Content of this course

“What’s the meaning of text, signs, and symbols?” This question has troubled philosophers for ages. In this course, we aren’t likely to find an answer to this question (because we aren’t even trying!). Instead, we will pleasantly circumnavigate this question and sail in calmer waters, where we are already given a theory of meaning, and therefore we are absolved from all potential blame for the inevitable loopholes in our theories of meaning (because we will not develop them!). We will also not create data for this task, so that, here too, we are absolved from blame of everything that can go wrong in complex data annotation.

Instead, the task we consider is just to learn a mapping between (compositions of) symbols in one domain (that is phrases composed of words or characters of natural language text) and (compositions of) symbols from another domain, the domain of (abstract) meaning (representations). To learn our ‘meaning mappings’, we won’t design any rules based on our intuitions of how the world works, leaving all this dirty work to neural networks.

More precisely, we visit recent research on neural approaches that provide high performance on meaning representation benchmark data sets. Thereby, we will improve upon our technical knowledge about complex, high-performance machine learning systems. Moreover, by doing this, perhaps, we may, or may not, hope to extend our knowledge about the meaning of signs and symbols, obtaining not only a gain of technical knowledge but possibly also a gain in knowledge about the meaning of signs and symbols.

Does all this excite you (or slightly raises your attention)? Then please join this course, we will have a good time.
2 Requirements to pass the course

1. Participation in our weekly heiconf-meetings. A link will be sent to you.

2. Presentation of a research paper (max. 25 minutes + max. 10 minutes discussion).

3. either i) a term paper or ii) a small implementation project with a technical report of the experimental settings and the results (or iii), a second presentation, subject to availability).

3 Schedule

First meeting (9.11.2020) Welcome and small introduction.

3.1 Subsequent meetings

After the first session, we will try to roughly stick to the following agenda:

Laying the groundwork

- **16.11.2020, Basics I: DRS and AMR meaning representation theories (2 presenters).** We will discuss two prominent theories of text meaning. Discourse representation theory [8] and abstract meaning representation [2].

- **23.11.2020, Basics II. Evaluation and IAA metrics for parsing and generation evaluation (2 presenters).** We discuss the Smatch [5] algorithm that compares AMR graphs, and we visit Bleu [16] that is commonly used for evaluation of all kinds of text generations (including meaning representation to text). In addition, we ask ourselves: what do humans actually have to say about the quality of text generated from AMRs? [13].

- **30.11.2020, Neural sequence-to-sequence models (2 presenters).** We discuss the ‘classical’ piece of Sutskever et al. [19] and ‘the transformer’ [21]. This prepares us well for the next session(s).

Assessing recent advances in MR parsing and generation

- **7.12.2020, Neural sequence-to-sequence (seq2seq) for AMR parsing and AMR generation (2 presenters).** We will discuss Konstas et al. [9], who find out that neural seq2seq models work well for AMR parsing and AMR generation. Then we see some more ‘tricks of the trade’ and a character-based AMR parsing seq2seq approach by Van Noord and Bos [20].
• 14.12.2020, MT pre-training greatly boosts seq2seq AMR parsing models (1 presenter) We discuss the recent work of Xu et al. [23], who show that large-scale MT pre-training provides very useful inductive biases for AMR parsing with seq2seq.

• 11.1.2021, Graph encoders for generation from AMR (2 presenters). We discuss the work of Song et al. [18] who use a recurrent graph encoder for better sequence generation from AMR (without the need of silver data). And we discuss the language model-based approach by Mager et al. [12] who show that pre-trained LMs (GPT2 [17]) can be taught to understand AMR language for improved generation.

• 18.1.2021, AMR parsing as graph prediction I (2 presenters). We will discuss Zhang et al. [24] who encode the sentence with BERT [7] and generate AMR nodes, finally predicting relations between pairs of nodes. And we discuss the work of Cai and Lam [4] who use iterative graph decoding for refined AMR graph generation and a BERT sentence encoding.

• 25.1.2021, Transition-based AMR parsing (2 presenters) We discuss the work of Wang et al. [22], to get started on transition-based AMR parsing and then discuss the work of Lindemann et al. [10], who aim at greatly increased AMR parsing speed.

• 1.2.2021, Cross-lingual AMR parsing (2 presenters) We will talk about the ‘early’ work of Damonte and Cohen [6] and the recent work by Blloshmi et al. [3], who both target a setup that predicts (English) AMR graphs from sentences of all kinds of languages.

• 8.2.2021, Let’s move to discourse level! (2 presenters) We will see how i) we can use structural decoding for DRS parsing [11] and ii) find out which things we can learn from a recent shared task in DRS parsing [1].

• 15.2.2021 Papers of your choice (max. 3 presenters) If you have a paper that fits well in the framework of this seminar but you find it missing from the proposed schedule and wished it was part of it, this is your chance to present it!

• 22.2.2021 Wrap-up, project and term paper discussion.

4 Possible implementation projects

There are several possible implementation projects. Many of the above papers provide code that can be installed and experimented with and the institute possesses a license for the AMR data. Alternatively, you are free to propose something, and we can discuss about it, whether it will be feasible. Furthermore, if you are interested in the quality assessment of automatically generated AMRs
or text generation from AMRs, you could build on some code that the teacher of this seminar wrote [14, 15].

5 Contacting the teacher

You can ask questions directly after the sessions have ended, or by writing an email to opitz@cl.uni-heidelberg.de (Please put [MRPG] in the email subject).

References


