Homework 3

CS 685, Spring 2021

Due: May 12

The class presentation writeups are required. The rest of the questions are optional, extra credit work on syntax and semantics, that count toward your homework grade.

1 Required: Presentation writeup

Please write summaries and comments on 3 of the class presentations you saw. You should take notes during the presentations in order to help you write a good commentary. Each summary/commentary can be fairly short, such as a few sentences, since it is difficult to go into much detail from a short presentation. However, it's definitely possible to talk about related or future work.

2 Extra credit: Grammatical agreement

(Pencil-and-paper)

- INLP chapter 10, question 1 (pg 253)
- INLP chapter 10, question 2 (pg 253)

3 Extra credit: Dependency statistics

(Computational experiment)

- INLP chapter 11, question 8 (pg 280)
- INLP chapter 11, question 9 (pg 280)

4 Extra credit: Semantics

(Pencil-and-paper)

- INLP chapter 12, question 2 (pg 303)
- INLP chapter 13, question 1 (pg 322)

5 Extra credit: Natural Language Inference and Semantic Phenomena

Go to the SuperGLUE "Diagnostics" tasks. They're described here: https://super.gluebenchmark.com/diagnostics. And the data are available under "Tasks."

Look at the outline under 'Linguistic Categories'. The fine-grained list of categories I'll call "syntax/semantic phenomena," which they tag the the "Broad Coverage" NLI examples with. For each NLI example in the JSONL file, they're listed under various JSON keys for the high-level categories. (Try pretty-printing the JSONL file to make it easier to see.[1]) The idea is supposed to be, that an NLI model can get the example correct only if it really understands how to deal with the tagged syntax/semantic phenomenon.

Choose **three** of the Broad Coverage Diagnostics phenomena, one each from 'Lexical Semantics,' 'Predicate-Argument Structure,' and 'Logic' high-level categories.[2] For each:

1. Explain what the phenomenon is, in your own words. See Wikipedia and other references for more details if you don't understand it.
2. Choose **two** of the NLI examples in the dataset that are tagged with this phenomenon. For each example (consisting of a premise-hypothesis pair),
   (a) Show the sentences.
   (b) Explain how the phenomenon is critical to determining the entailment category.
   (c) Change one of the sentences in a small way—for example, changing/inserting/deleting a small number of words—so that the entailment status changes, and that the semantic phenomenon is still critical to determining the entailment status. Explain what you did and how the new example works. (You have created a *minimal pair* of NLI pairs!)

6 Extra Credit: Gender and Coreference

Create (at least) **five** of your own Winogender-style templated examples (following Rudinger et al.), where a template has 3 instantiations for (at least) three different pronouns of varying genders (two for the two binary genders, and then another that's non-binary or gender-unspecified), but the coreference resolution should be invariant to the choice of pronoun. Try them on a publicly available coreference system (the CoreNLP web demo may be easy to use: https://corenlp.run/). Report the results, including accuracy, and some way of summarizing how sensitive the predictions are to pronoun gender.

[1] `cat AX-b | jq .` or `for line in sys.stdin: print(json.dumps(json.loads(line), indent=4)` or https://github.com/brendano/cmdutils/blob/master/jsonpp etc.

[2] Do not use the 'Knowledge and Common Sense' category. We're excluding them because they're too infinitely powerful/broad, and thus too easy for this question—precisely because they're so important!