Assignment-3: Parsing

Anish M M, 203050066
Pooja Verma, 203050072,
Shubham Nemani, 203050011,

Indian Institute of Technology Bombay

November 17, 2020

Conversion from CP to DP

- Input is list of tokens - (,) , tags , words for CP to DP conversion.
- We are using a shift reduce parser to parse CP string into tree using stack.
- It involves two steps:
  - Rule[1] based approach is used to find head of phrase using transforming dependencies into phrase structures [2]
  - Assigning dependency labels between head and modifier [3]
- We are using 2 validation metrics
  - head finding accuracy
  - dependency relation label accuracy

[1] https://www.aclweb.org/anthology/W07-2416.pdf
[2] http://people.seas.harvard.edu/ srush/naacl15.pdf
[3] https://www.researchgate.net/publication/324940566_Guidelines_for_the_CLEAR_Style_Co

```
( ROOT
        ( S
                ( NP
                        ( DET The )
                        ( NOUN detective ) )
                ( VP
                        ( VERB listend )
                        ( PP
                                ( ADP with )
                                ( NP
                                        ( DET a )
                                        ( ADJ wooden )
                                        ( NOUN face ) ) ) ) ) )
```

Figure: Input:Constituency Parse Tree generated by Berkeley Parser

```
det(detective-2 , The-1)
nsubj(listend-3 , detective-2)
root( ROOT-0 , listend-3)
prep(listend-3 , with-4)
det(face-7 , a-5)
amod(face-7 , wooden-6)
pobj(with-4 , face-7)
```

Figure: Dependency parse generated by our code (output)

Constituency Parse Tree for a sentence is generated using **Berkeley Neural Parser** in parallel with **Spacy Library**.

CP to DP Error Analysis

1. In some sentences adjectives are tagged as NN by parser in original CP output , so wrong head chosen by rules leading to incorrect labeling.
   Sentence- Students played street football.
   CP Tree-(S (NP (NNS Students)) (VP (VBD Played) (NP (NN street) (NN Football))))
   Here , street is tagged as NN , so for phrase NP , street is selected as Head.
2. Sentences containing punctuations give very low accuracy , need to handle them.
3. Dependency relation labeling accuracy is very low for CASE , CCOMP , XCOMP and COMPOUND.

Conversion from DP to CP

- Rule-based approach.
- Input dependency parse created from sentence using StanfordNLP parser.
- Uses Universal POS tags for CP.
- Groups words in a bottom up manner: First, link adjectives to their nouns. Then, numeral modifier followed by determiner. Then, link object and subject to verb. Finally we end up with DP for whole sentence.
- Our tool handles sentences with a verb, subject, object, one or more adjectives, one or more adverbs, numeral modifier for nouns, determiners.

Consider the input sentence:
The five small white cats very quickly ate the 4 large black mice .

```
-> % python3 dp_to_cp.py
Write sentence:
The five small white cats very quickly ate the 4 large black mice .

Dependency parse (generated by StanfordNLP api.):
('The', '5', 'det')
('five', '5', 'nummod')
('small', '5', 'amod')
('white', '5', 'amod')
('cats', '8', 'nsubj')
('very', '7', 'advmod')
('quickly', '8', 'advmod')
('ate', '0', 'root')
('the', '13', 'det')
('4', '13', 'nummod')
('large', '13', 'amod')
('black', '13', 'amod')
('mice', '8', 'obj')
('.', '8', 'punct')
```

Figure: Dependency parse generated by StanfordNLP parser (input)

```
Corresponding Constituency Parse:

( ROOT
        ( S
                ( NP
                        ( DET The )
                        ( NUM five )
                        ( ADJ small )
                        ( ADJ white )
                        ( NOUN cats ) )
                ( ADVP
                        ( ADV very )
                        ( ADV quickly ) )
                ( VP
                        ( VERB ate )
                        ( NP
                                ( DET the )
                                ( NUM 4 )
                                ( ADJ large )
                                ( ADJ black )
                                ( NOUN mice ) ) )
                ( . . ) ) )
```

Figure: Constituency parse generated by our code (output)

Analysis for DP to CP

- We only support a subset of sentence types as indicated in previous slides. The tool gives an error message for unsupported types.
- We are unable to identify Proper Nouns (PROPN) since the dependency parse does not provide this distinction.
- The DP also does not give information on whether subject/object is a noun or pronoun. However, since pronouns are a closed set, we use a lookup to check if pronoun or not.
- The StanfordNLP parser fails to produce parse for some sentences. In those cases, we report the same.
- Sometimes, the StanfordNLP parser keeps running without returning anything, and we have to manually kill the process. In that case too, our tool will fail.

The END