NLP 681

Rutvik Pansare – rp2832@rit.edu

Problem Set 3

Q1.

Sentence1: “He hath eaten me out of house and”

Sentence2: “tis not long after but I will wear my heart upon my sleeve”

Sentence3: “You know that smoodle pinkered and that I want to get him”

Sentence4: “My door sat through the lamp in the”

|  |  |  |
| --- | --- | --- |
| Sentence | Library1  (Stanford CoreNLP) | Library3  TextBlob |
| Sentence 1 | 1.Parse.  Text  Description automatically generated  2.Sentiment  Text  Description automatically generated  3.Coref- | Sentiment(polarity=0.025, subjectivity=0.4) |
| Sentence 2 | 1.Parse.  Text  Description automatically generated  2.Sentiment  Text  Description automatically generated  3. Coref-  Text  Description automatically generated | 2.Sentiment  2.Sentiment  Sentiment(polarity=0.0, subjectivity=0.0) |
| Sentence 3 | 1.Parse.  Text  Description automatically generated  2.Sentiment  Text  Description automatically generated | 2.Sentiment  Sentiment(polarity=0.0, subjectivity=0.0) |
| Sentence 4 | 1.Parse.  Text  Description automatically generated  2.Sentiment  Text  Description automatically generated | 2.Sentiment  Sentiment(polarity=0.4, subjectivity=0.7) |

|  |  |
| --- | --- |
| Sentence | Library – Berkeley Neural Parser |
| Sentence1 | Graphical user interface, application  Description automatically generated |
| Sentence2 | Graphical user interface, application  Description automatically generated |
| Sentence3 | Graphical user interface, application  Description automatically generated |
| Sentence4 | Graphical user interface  Description automatically generated with medium confidence |

Comparison:

1. The constituency parsing done by CoreNLP is very accurate and detailed while the Berkeley neural parser fails to recognize small details.

2. As the sentences are complex in nature, coreference resolution is a tough task to perform and only CoreNLP was able to achieve this task. It found the reference between “my” and “I” In the sentence 2.

3. The CoreNLP parser provides a CNF style parsing format while the other libraries don’t always follow CNF style parsing.

4. In the task of sentiment analysis CoreNLP performs better as it gives detailed explanation of every word in the sentence. It shows the sentiment of every word, while other libraries only show the overall polarities of the sentence.

5. When CoreNLP was not used for parsing many of the parse tree had incorrect tags for the words as the sentence was broken with incorrect grammar in the first place.

6. As the sentences had incorrect grammar, CoreNLP could not find the coreference resolution for 3 sentences.

Q2.

Sentence = A small-sized self-driving car stopped at the light.

Grammar rules:

S -> NP VP  
NP -> DET NP  
NP -> DET Noun

NP -> IN NP  
VP -> V PP  
VP -> P NP  
NP -> ADJ Noun  
ADJ -> ADJ ADJ  
DET -> the A  
Noun -> car light  
V -> stopped  
P -> at  
ADJ -> small-sized self-driving

Parse tree for CFG.

Graphical user interface

Description automatically generated

(1)

CNF grammar rules:

S -> NP VP  
NP -> DET NP  
NP -> DET Noun  
VP -> V PP  
PP -> P NP  
NP -> ADJp Noun  
ADJp -> ADJ ADJ  
DET -> the A  
Noun -> car light  
V -> stopped  
P -> at  
ADJ -> small-sized self-driving

Parse Tree for CNF:

Graphical user interface, diagram

Description automatically generated

(2)

Constituency parse:

Text

Description automatically generated

Diagram

Description automatically generated

(3)

Dependency Parse:

Text

Description automatically generated

Graphical user interface, application

Description automatically generated (4)

Similarities between the graphs:

1. The depth level 1 of all the graphs is always a combination of Noun Phrase and verb Phrase. This was not surprising as the English language follows the Subject-Verb-Object model.

2. The Verb Phrase side of the parse tree always spans out in a binary fashion most of the times. That is two nodes splitting from one single node.

3. Nouns are usually followed by verbs and prepositions.

Dissimilarities in the graphs:

1. The graphs 3 and 4 are more detailed. Graphs 3 and 4 even Identify the hyphen in the sentence while graph 1 and 2.

2. Graphs 1 and 2 do not identify VBD and VBG as parts of speech which is loss of information.

3. The dependency parse graph (4) goes a level ahead and shows the inter-related information about different parts of speech in the sentence. For example, It identifies that the word “A” is used as a determinant for the noun “car”.

4.The CNF (2) graph and CFG (1) Graph have a different form of splitting structure.

5.The CNF (2) graph have a binary splitting structure.

Q3.

Sentence 1 = “He hath eaten me out of house and”

Text

Description automatically generated

As the sentence contains unusual English words, the parser is having difficulty understanding the POS tag of the word. Even though the word “hath was correctly classified as a verb in the CKY parser.

As the CKY algorithm only works with a CNF style grammar rules, the tree has a binary splitting nature.

The word “and” is a conjunction here.

Sentence 2 = “tis not long after but I will wear my heart upon my sleeve”

Text

Description automatically generated with medium confidenceText

Description automatically generated

In this sentence there were many prepositions used in an un-usual manner for which the grammar rules

had to be modified to create the parse. The sentence also had a double adjective phrase “not long” which had “ADJp” or Adjective phrase as its parent.

Sentence 3 = “You know that smoodle pinkered and that I want to get him”

Text

Description automatically generatedText

Description automatically generatedText

Description automatically generated

As this sentence is spoken by a person suffering from the condition of Wernicke’s aphasia, the sentence is quite unclear. The word smoodle was termed as an adjective but it could be a verb as well because the true meaning of the sentence is not known in this case.

Sentence 4 = “My door sat through the lamp in the”

Text

Description automatically generated with medium confidence

Again, as this sentence is spoken by a person suffering from the condition of Wernicke’s aphasia, we cannot decipher the true meaning of the sentence just by reading it. The output neatly follows the CNF style i.e., two non-terminals or a single terminal from the parent node.

Sentence 5 = “A self-driving small-sized car stopped at the light.”

Text

Description automatically generated

This sentence also has a double adjective phrase which is parsed as an ADJp tag. The parser picks up the words “self-driving” and “small-sized” both as adjectives.

The following were the problems faced during creating the parser and parsing the sentence.

1. The CKY algorithm uses dynamic programming approach to find the solution to sub-strings which can get challenging to deal with.

2. The solution is stored in a multi-array table called OPT but retrieving the solution from the table is a tough part. A separate table was used to store the pointers to the solution which were used to back track the solution. A method “printRecursivlely” was created to visit all these pointers from the new table in a recursive manner.

3. As the sentences not always formed sense, the grammar rules of the parser had to be modified for each sentence so that the CKY parser could parse it.