Natural Language Understanding - E1246
Assignment 3
PCFGs and Constituency Parsing

Suman Gupta
Computer Science and Automation
Sr No. - 15788
Indian Institute of Science
sumangupta@iisc.ac.in

1 Introduction

In this assignment, we have implemented unlexicalized PCFG parser using Penn TreeBank dataset provided in the NLTK library.

2 Solution Sketch

I have implemented PCFG using the induce_pcfg method provided by the nltk library

2.1 Training

1. Data Loading done from Penn Treebank and 80-20 train-test split.

2. Productions extraction

3. Addition of new productions of the form N–'UNK', corresponding to every production of the form N–t, where N is a non terminal and t is a terminal.

4. Inducing grammar.

5. Generation of grammar pickle file.

2.2 Testing

1. Loading the saved grammar pickle file.

2. Generating the parse tree for the input sentences.

3. Substituting the 'UNK' tokens in the parse obtained with the original words that are not covered by the grammar.

4. Comparing the parse trees generated with the true parse tree labels.

5. Precision, recall, F1 score computation.

3 Results

3.1 Task 1

Precision =0.61
Recall=0.62
F1-score=0.61

3.2 Task 2

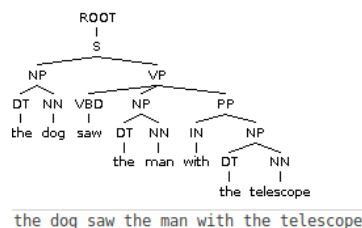**Comparision of results with standard PCFG parsers**

3.3 Short sentence

Parse generated by parser implemented

Parse generated by standard parser
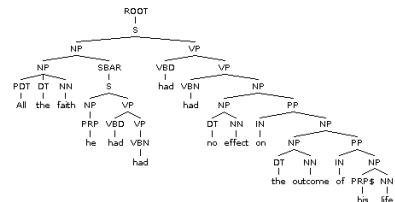

The parser is online.

3.4 Long sentence

Parse generated by parser implemented

```
(base) suman@suman-pc:~/NLU/Newfolder$ ./run.sh "All the faith he had had had had no effect on the outcome of his life"
(S
  (NP-TMP (DT All) (NP-TMP|<JJ-NN> (JJ the) (NN faith)))
  (S|<NP-SBJ-3-VP>
    (NP-SBJ-3 (PRP he))
    (VP
      (VBD had)
      (PP-LOC-PRD
        (ADVP
          (VBN had)
          (PP
            (VBN had)
            (PP
              (VBN had)
              (PP
                (NP-ADV (DT no) (NN effect))
                (PP|<IN-NP>
                  (IN on)
                  (NP (DT the) (RBS outcome)))))))))
      (PP-LOC-PRD|<IN-NP>
        (IN of)
        (NP (PRP$ his) (NN life))))))) (p=1.36878e-62)
```

Parse generated by standard parser

```
(ROOT
  (S
    (NP
      (NP (PDT All) (DT the) (NN faith))
      (SBAR
        (S
          (NP (PRP he))
          (VP (VBD had)
            (VP (VBN had))))))
    (VP (VBD had)
      (VP (VBN had)
        (NP
          (NP (DT no) (NN effect))
          (PP (IN on)
            (NP
              (NP (DT the) (NN outcome))
              (PP (IN of)
                (NP (PRP$ his) (NN life))))))))))
```



```
All the faith he had had had had no effect on the outcome of his life
```

Parse!

The parser is online.

4 Path to the shell script file

/home/e1-246-25/a3/run.sh