CS 562 - Homework 5

Satheeshkumar Karuppusamy

karuppus@usc.edu

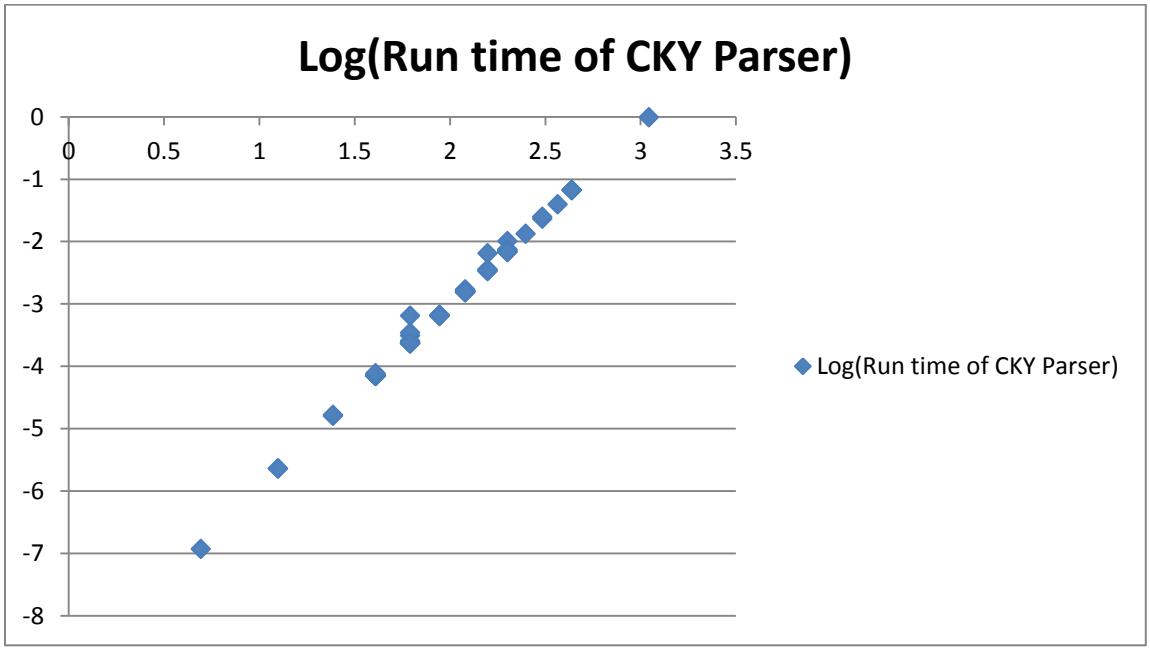
1. **Maximum Branching Factor: 8**
   (TOP (FRAG (NP (NN Flight) (NN number) (NNP C) (NNP O) (CD one) (CD seven) (CD one) (CD nine))) (PUNC .))
   Longest Branch: (NN Flight) (NN number) (NNP C) (NNP O) (CD one) (CD seven) (CD one) (CD nine)

2. **Number of words in dev not in train : 9**
   M,does,connecting,provides,seventh,Restriction,anywhere,breakfast,eighty

3. **Number of Rules in the Grammar : 752**

4. **Scatter Plot – log(# of Words in Input sentence) vs log(CKY Parse time) – dev.strings**

   K = (-6.931960206+0.008388077)/(-0.693147181+3.044522438)
   k value using least-square fit = 2.351781



Handling words which are not in train and Handling Unparseable senences :

1. Unknown words are replaced with <unk> and then parsed
2. Dummy trees are produced for unparseable strings. Dummy trees with following structure is produced for all the nonparseable strings
   (TOP (NP (NP word1) (NP (NP word2)…….)))

5. **Output of evalb.py script**

Input: dev.strings file : F1 score = 0.868085106383

dev.parses.post 466 brackets
dev.trees 474 brackets
matching 408 brackets
precision 0.875536480687
recall 0.860759493671
F1 0.868085106383

Input : test.strings file : F1 score = 0.859259259259
test.parses.post 474 brackets
test.trees 471 brackets
matching 406 brackets
precision 0.856540084388
recall 0.861995753715
F1 0.859259259259

6. **Improving the Parser :**

Head Lexicalization method is tried and applied. Head of the VP is choosen as the verb, DT as that particular determiner and so on. Since most of the sentences in our training examples with PP attachment are already aligned properly depending on the context it dint improve the performance of the parser. Rather it degraded the performance for new input. F score was reduced around 1% for the test and the dev strings. F1 score for dev sets are about. Then I had just tried to apply the head words lexicalization for particular words like prepositions which wasn't increasing the performance that much great.

7. **Input : test.strings file**

test.parses.post 474 brackets
test.trees 471 brackets
matching 406 brackets
precision 0.856540084388
recall 0.861995753715
F1 0.859259259259