

EN 600.465

Natural Language Processing

Assignment 1

Xuan Zhang, Zhouyang Zhang, Chenxi Liu

September 13, 2016

1 Problem 2

(a) Why does your program generate so many long sentences? Specifically, what grammar rule is responsible and why? What is special about this rule?

Answer: It contains some rule of high probability (i.e. high weight) of repeating it self. For example, the rule $NP \rightarrow NP \ PP$ have $1/3$ probability of getting back to itself (So, there are at least $1/27$ chance $NP \rightarrow NP \ PP \ PP \ PP \dots$ or longer). So, when we add more recursive rule, we should be careful when specify the rule which has the chance of repeating itself (imagine the sentence: I like the fact that I like the fact that I like the fact ...)

(b) The grammar allows multiple adjective, as in *the fine perplexed pickle*. Why do your program's sentences do this so rarely?

Answer: The particular grammar rule that allow multiple adjective is $Noun \rightarrow Adj \ Noun$, however, the probability of activating this rule is $1/6$. So, the chance of generating more than two consecutive *Adj* is $1/36$, and this chancing decreasing exponentially.

(d) Which numbers must you modify to fix the problems in (a) and (b), making the sentences shorter and the adjectives more frequent? (Check your answer by running your generator!)

Answer: As discussed above, we can decrease the weight of the rule $NP \rightarrow NP \ PP$ to some smaller number, e.g. 0.3 to prevent situation (a) happening too frequently. Similarly, we can change the weight of the rule $Noun \rightarrow Adj \ Noun$ to some bigger number, e.g. 4, if we want to see more consecutive *Adjs*.

(e) What other numeric adjustments can you make to the grammar in order to favor more natural sets of sentences? Experiment. Hand in your grammar file in a file named grammar2, with comments that motivate your changes, together with 10 sentences generated by the grammar.

Answer: In really life, we somehow want see more interesting nouns, which appearing with adjective, something like sky v.s. beautifully sky. So, we can give $Noun \rightarrow Adj \ Noun$ more weight. Also, $NP \rightarrow Det \ Noun$ used frequently in really life like "the apple", "an apple" (compared to "apple from"). Some change can be made when generating from *ROOT* to *S*. Sentences ended with "." is obviously more than sentences ended either with "!" or "?" (unless you are reading some dialogical corpus). The following is the sample sentences produced, again, simple grammar rules can only generate very simple sentences, and the "goodness" of sentence also determined by the domain you are in (dialogical corpus have more "?").

every pickle rate every floor !
a floor kissed a pickle .
the chief of staff wanted a sandwich .
is it true that every pickled pickle kissed every chief of staff ?
a pickled pickled president pickled every pickled chief of staff .
the chief of staff rate every perplexed pickled perplexed fine president .
every president wanted the pickle !
the fine pickled floor kissed every floor !
is it true that every fine president understood the pickled pickled perplexed delicious president ?
a chief of staff under a fine fine president pickled a sandwich .