NLTK analysis

Source: https://www.youtube.com/watch?v=-UqDljMymMg

To start with the analysis of our sentence generator, we should first answer the question “why?” Why is it important to determine the efficiency and precision of the sentence generator when we already have a sentence generated, which even though might not be exactly as expected, will definitely involve most of the keywords of objects present in the image?

To start with, it is essentially the first thing to keep in mind while writing a sentence generator algorithm that one must understand the end user of the algorithm. Means basically to know who is going to read these sentences. There can be four categories to this

1. Will directly be read by a human
2. Will get edited first by a human and then get published for humans to read
3. Will be read completely by another machine.
4. Will be used to increase the size of database and will be read by data mining Softwares.

These four outputs are arranged in the descending order of requirement of high efficiency and precision.

Second question that arises is “How do we do it?”

There are multiple way to analyze the algorithm.

The most generalized being to compare it with the sentences generated by humans. Other ways include judging on the basis of grammatical errors, linguistic correctness, word ordering etc.

But since the best sentence generator would be the one which creates sentences best suited for humans to read, comparing machine generated sentences with the human generated sentences is the best way to judge the algorithm and is the same we have used to compare ours.

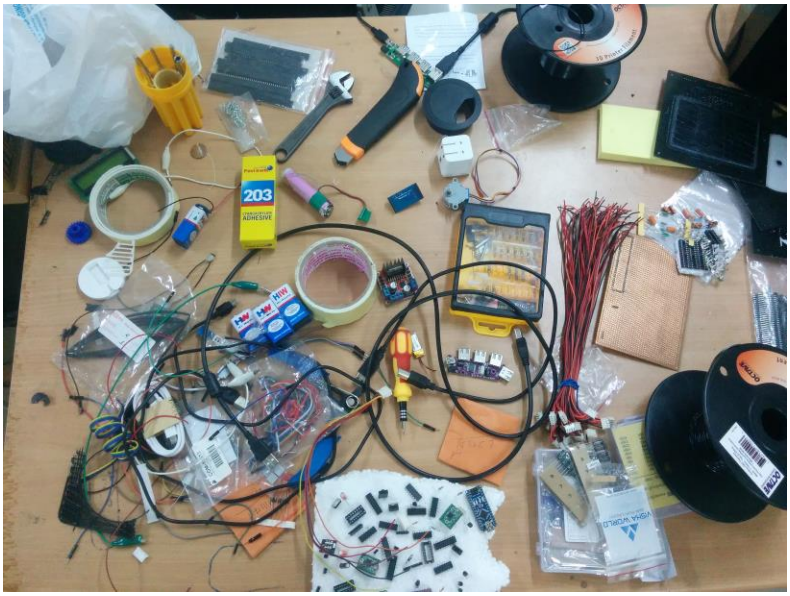
The scale that we shall be using to study the efficiency of our generator shall be the Bilingual Evaluation Understudy (BLEU) score.

Source: http://en.wikipedia.org/wiki/BLEU

The bleu is based on the principle that it gives the output as a number between 0 and 1 referring to the closeness of the machine generated sentence to the human generated reference sentence which is further multiplied by 100 to get the score in percent.

Source: http://www.aclweb.org/anthology/P02-1040.pdf

Consider the image:



One of the many sentences generated by the algorithm is “A USB multiplexer is placed on a table.”

Some ideal reference sentences for this scene were

1. This is a cluttered scene consisting of ICs, batteries, USB multiplexer and bunch of wires lying on a wooden desk.
2. This is a cluttered wooden desk with ICs, batteries to the left of bunch of wires.
3. There are a lot of objects in this scene, some of which are batteries, ICs, bunch of wires and a USB multiplexer.

The unigram precision for the algorithm generated sentence is given by

$$P = \frac{m}{w_t}$$

Where m is the total number of words from the algorithm generated sentence found in the reference sentence database. We generally choose m to be m_{max} which is the maximum number of matches of the word into consideration from algorithm generated sentence with any of the sentence from reference database.

w_t is the total number of words in the database.

Thus the unigram score for the generated sentence is

$$P = \frac{2}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{0}{8} + \frac{1}{8} + \frac{0}{8}$$

$$P_{unigram} = \frac{3}{4} = 0.75 = 75\%$$