Report – Project 4

SSAP and Project 3 feature grammar performance evaluation on sentiment analysis:

SSAP is been provided with AFINN text, where words with sentiments have been provided a certain degree of numbers between -5 to 5. Taking all the words from a sentence present in the AFINN.txt and get a sum of them to find a certain degree of the sentiment of the sentence, whether it is pos (positive), neg (negative) or neut (neutral).

Comparison and Analysis:

When we evaluate sentences on SSAP there are certain conditions where sentences were tagged correctly and incorrectly, I analyzed the performance of feature grammar from Project-3 keeping SSAP as baseline.

Example:

SSAP =, Feature Grammar Project 3 = neg

Sentences	SSAP Performance	Feature Grammar Performance
a) This cd was horrible.	-1.34	neg
b) The controls are lousy.	0.0	neg
c) This is not fun.	1.79	neg

As we can see 3 different scenarios where baseline has similar values or varies from the feature grammar output. Sentence a) is an actual negative sentence and it is tagged as neg by feature grammar and SSAP as well, but the sentences are actually tagged correctly by feature grammar as they are incorrectly tagged in baseline.

Design better than SSAP (afinn) [in some aspects]:

Let's talk about how the design of feature grammar is better than SSAP?
Taking some examples such as "**This is not fun.**" It is basically a negative sentence and it is handled in the feature grammar as the negated ("not") positive word ("fun") makes it negative, but whereas the SSAP gives is a mark of 1.79 which makes it positive and it is not the correct sentiment inference for the mentioned sentence.
Another example where "**The controls are lousy.**" It is basically a negative sentence and it is handled in the feature grammar as the negated ("not") positive word ("fun") makes it negative, but whereas the SSAP gives is a mark of 1.79 which makes it positive and it is not the correct sentiment inference for the mentioned sentence.

Precision, Recall and F1-score:

Given 29 test sentences, here are the results that were tagged correctly and incorrectly by system and baseline.

TP = correctly predicted having correct sentiment, FN = incorrectly predicted having correct sentiment, FP = correctly predicted having incorrect tagged sentiment.

	Predicted: Real Sentiment	Predicted: Tagged Sentiment
Actual: Real Sentiment	0	5
Actual: Tagged Sentiment	9	15

Containing 29 sentences in test data, we have,

TP = 15, FP = 9, FN = 5

Since we know that,

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Therefore,

Precision = 15 / (15+9) = 15/24 = 0.625

And, we also know that,

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Therefore,

Precision = 15 / (15+5) = 15/20 = 0.75

F1-score:

$$f_1\text{-score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

F1-score = 2*(0.625*0.75 / (0.625+0.75)) = 0.68