REPORT

PROJECT 4

PREPARED FOR

Dr. Sabine Bergler

PREPARED BY

Karandeep

40104845

TABLE OF CONTENTS

Introduction: **2**
Purpose: 2
Getting Started: 3
Design Justification: 3
Design Critique: 4
Comparing with SSAP: 5
Conclusion 5
References: 5

Introduction:

About NLTK - NLTK aka the Natural Language Toolkit, is a suite of open source Python modules, data sets, and tutorials supporting research and development in Natural Language Processing. It contains text processing libraries for tokenization, parsing, classification, stemming, tagging and semantic reasoning. It also includes graphical demonstrations and sample data sets as well as accompanied by a cook book and a book which explains the principles behind the underlying language processing tasks that NLTK supports.

Purpose:

Hidden information often lies deep within the boundaries of what we can perceive with our eyes and our ears. Some look to data for that purpose, and most of the time, data can tell us more than we thought was imaginable. But sometimes data might not be clear cut enough to perform any sort of analytics. Language, tone, and sentence structure can explain a lot about how people are feeling, and can even be used to predict how people might feel about similar topics using a combination of the Natural Language Toolkit, a Python library used for analyzing text, and machine learning.

For this project I was to develop a Feature grammar to parse given sentences and judge the sentiment behind the sentence and compare with AFFIN sentiment analyzer. The console has options to run **Sentiment Analyzer** either on a good inbuilt data or bad inbuilt data sentences to parse. Analysis of data has been done in a way that is easily comprehensible. Overall goal of the project was to create a parser and a FCFG which filters data by using NLTK access commands and creates sentence trees and visualize them in a more informative way. NLTK is not perfect and lacks in some areas which is being explored below.

Getting Started:

1. Open CMD (for Windows), Terminal (for MacOS).
2. Goto directory of the script.
3. Type in command “python3 sentiment.py”

Design Justification:

How Grammar work?

The grammar is a feature grammar built on the top of NLTK's feature grammar feat0.fcfg.

- Grammar tags the negative and positive adjectives and the parent nodes of the sentences.
- Grammar has the ability to handle negations while parsing bottom up, to judge the correct sentiment.
- Grammar has the ability to handle conjunctions and parse declarative sentences.
- Score of the sentence is based on tags in parse trees and the individual score given to words/adjectives and nouns.
- Opinion Lexicon is used to assign the weight to identified words, from their tags.
- A cumulative logic is applied to gather all weights and give a final score.
- Score calculating algorithm decides the final polarity by sum logic.

What Grammar does?

- In addition to the already existing grammar, new Non-terminals are added:

ADJ[SNT=POS]	ADJ[SNT=NEG]	DT	PRP	MD	Comp	CD

- ADJ non terminals are populated by using the words in the opinion lexicon:
    1. **Positive opinion lexicon** words are used to populate the ADJ with feature *SNT=POS*
    2. **Negative opinion lexicon** words are used to populate the ADJ with feature *SNT=NEG*
- Negation is handled by adding non terminals to support adjectives.
- Production rules justifies the structure of the sentence to judge correct sentiment.

AND

If the conjunction is 'and' then the sentiment is calculated on the basis of total POS or NEG tags and the cumulative score.

BUT

If the conjunction is 'but' then higher weight is assigned to the tags after 'but'.

NEGATION

If the adjective is preceded by a negating word like 'not', 'rarely', or 'never' then the polarity of the adjective is reversed and transferred up to the parent.

- The grammar parses sentences that contain **oxymorons** like 'pretty dreadful' or 'wonderfully bad'.
- Negative sentiment is assigned to the parent node, this is only done for negative sentences as the grammar would become significantly large taking into account all sentences. The tree generated by the grammar needs to be traversed to the lowest level to check all the tags.

- Each adjective is assigned a weight which is then used to calculate the final score for the sentiment. If the score is positive, then the sentence is positive, if not then it is either negative and if zero it is neutral.

Design Critique:

The grammar does not include the following points that should be considered as important:

- **Negations:** Position of negating words is important if to calculate the real sentiment of the sentence. If the negating word is very far from the adjective then it is likely that the sentence will not be correctly identified.
- **Oxymorons:** The grammar does not identify all oxymorons but can identify some.

Complex sentences with conjunctions on adjectives:

The calculation for conjunction can sometimes work for bad than for good.

- **Conjunctions:** Complex sentences containing conjunction of adjectives are handled based on the conjunction.
- **AND:** On occurrence of AND, both the adjectives on the either side are given equal priority.
- **BUT:** If ‘but’ is the used conjunction then the adjective after the conjunction is Assigned more weight.

Complex sentences with conjunctions in sentences:

- Complex sentences containing conjunction of sentences are handled based on the conjunction.
- **AND:** If ‘and’ is the used conjunction then both the sentences are given equal priority.
- **BUT:** If ‘but ’ is the used conjunction then the sentence after the conjunction is given more weight and the calculation is done based on that.

Comparing with SSAP:

SENTENCE	SSAP OUTPUT	PROJECT 4 OUTPUT	PROJECT 3 OUTPUT
this is one of the best book by Crichton. the characters of Karen Ross , Peter Elliot , Munro and Amy are beautifully developed and their interactions are exciting , that get lost in the film. this may be the absolute worst disparity in quality between novel and the screen adaptation. the book is really good. the movie is just dreadful.	NEGATIVE	POSITIVE	NEUTRAL
there is no movie I have been more prepared to dislike than this one. How dare some Aussie come over here and tell us about the meaning of one of the great works of American literature. Especially this Aussie, Baz Luhrmann , who is known to overload.	POSITIVE	NEGATIVE	NEUTRAL
this may not have the dramatic gut-wrenching impact of other holocaust films, but it is a compelling story mainly because of the way it is told by the people who were there	NEUTRAL	POSITIVE	NEGATIVE

Conclusion

It can be concluded that for the stated examples, SSAP failed to identify the polarity of the sentences whereas my grammar was able to annotate it correctly.

- Conclusion can be made from here that negation words are not considered by the SSAP sentiment analysis. Also, the vocabulary contains only limited words which are given weights and the analysis depends on these words alone. If a word that is not in the list is used to express a sentiment, it will not be identified.

References:

- https://github.com/nltk/nltk_teach/blob/master/examples/grammars/book_grammars/feat0.fcfg2.
- https://www.nltk.org/book/ch09.html