COMP 6751: Project 3

Akash Ambashankar
40232517

November 9, 2022

1. **Justify your design (Why did you make the grammar do what it does? Why did you make it not do what it doesn't?)**

   The grammar consists of features to determine the sentiment of a given sentence. The sentiment of a sentence is most often found in adjectives, and certain nouns, but for this project, the grammar focuses on the sentiment held in adjectives, which eventually percolates upward towards the root of the tree through Noun Phrases, Verb Phrases, Adjective Phrases, etc.

   The grammar uses three classes of features, *pos, neg, neut*, for positive, negative and neutral sentiments respectively. The features are assigned at the lexicon level, and as the *Feature Earley Parser* completes lexicons based on appropriate rules, the sentiment feature moves up as an argument toward the root of the tree. The sentiment that reaches the top of the tree is assigned to the entire sentence.

   The major challenge here is creating appropriate rules for the unification of features when the sentiments between multiple phrases in a sentence are not in agreement. The approach followed in this grammar is detailed below.

   (a) For any sentence of the form *"...phrase1...* ***but*** *...phrase2..."*, the sentiment of the second phrase is taken as the sentiment of the entire sentence. This is purely a design choice, but it is visible quite often in English. For example, the sentence *"this does not have gut-wrenching impact but it's a compelling story"* is a positive sentence because of the positive adjective *compelling* in the second phrase.

   On the other hand, the Sentiment Analyser from CoreNLP does not agree with this approach for all sentences. Hence some sentences such as *"well-intentioned but manipulative movie making"* are marked as *neutral* by CoreNLP, although according to the design of the grammar in this project it would be a negative sentence (hence although I personally feel it is a negative sentence, this sentence has been put in the False.txt file in the submission).

   (b) For any sentence of the form *"...phrase1...* ***but*** *...phrase2..."*, if the sentiments of the two phrases are different, the whole sentence is marked as neutral (agreed, this may not be the best solution).

   (c) Other feature unifications that deal with phrases that have opposing sentiments include instances such as *NP[pos] and PP[neg]*, but the number of such occurrences was trivial and were handled case-by-case.

   After going through all of the rules, the parser may generate one or more trees for each sentence, and each tree can classify the sentence as positive, negative, or neutral. In order to aggregate all results, a voting module has been defined, that counts the number of trees for each sentiment category, and the sentiment for the sentence is determined to be the category with the most number of parse trees.

2. **Critique your design (What does your grammar not do that you think important?)**

   From the perspective of the sentiment analysis, many of the sentences included in the *False* file were there because they determined the sentence to be neutral instead of positive or negative. Many of the longer, more convoluted sentences, tend to come out as neutral, due to the presence of more inaccurate trees. This leads to the neutral trees outweighing the positive or negative trees in the voting process.

   The grammar in this assignment only includes features that relate to the sentiment of a word (typically adjectives and nouns). By creating more specific grammar rules, and/or introducing more features, we can reduce the number of incorrect trees, making the voting process more accurate, although this can also lead to much stricter grammar which may not parse some valid English sentences.

   Another major pitfall of the grammar is the lack of importance given to commas. As discussed in class, commas hold a lot of meaning in a sentence, but from a programming perspective, it's just much easier to ignore them, although leaving them in could reveal some useful emphases, increasing the positivity or negativity of a phrase.

3. **What kind of semantics you can do with your grammar (and how much additional work that would require)**

   The most obvious improvement to the grammar is to include more grammatical features, which would improve the accuracy of the sentence classification.

   Similar to the comment in 1(a), there may be more semantics in the English language that can be discovered by further developing this grammar.

   Further, if we were to develop a domain-specific sentiment analyzer, the sentiments assigned to lexicons can be tweaked to represent the domain with more accuracy than what can be done in a general case.