

Sentiment Analysis

Supplement for Module 4 Programming Assignment

Note: this is not the assignment, just more depth and discussion about sentiment analysis as used in the assignment.

Background

Sentiment analysis is a task from the field of computational linguistics that seeks to determine the general attitude of a given piece of text. For instance, we would like to have a program that could look at the text “This assignment was joyful and a pleasure” and realize that it was a positive statement while “It made me want to pull out my hair” is negative.

Scoring Statements:

When reading the statements above, we know words such as “joyful” are very positive and suggest the statement is positive. More precisely, to calculate the sentiment score of a statement we compute the mean (arithmetic average) of the word scores. Words used multiple times contribute for each occurrence, “happy happy happy days” should be more positive than “happy days” (assuming “happy” has a higher sentiment score than “days”). Unknown words start with a neutral sentiment, and therefore when evaluating a statement.

Given input word sentiments:

happy	1
sad	-1
positive	2
negative	-2

The sentiments of the following statements would be:

Happy positive	1.5
Happy happy days	0.66666...

Positive sad	0.5
Positively sad	-0.5 (positively is unknown)
\$!\$ #\$\$@ 2 1 3	0 (no valid words)

Scoring Words

The sentiment of a word is the mean of the contextual scores of all occurrences in statements that have been analyzed to date. The contextual score of a word's occurrence in the statement is simply the score of the statement. Consider the statement "It was a joyful happy day". If we aren't familiar with the word "joy", finding it in the same sentence as "happy" suggests it is a positive word. Repeated use of a word in a single statement does count as separate occurrences.

In the assignment you will be given a corpus of pre-scored statements that you will use to score words. For instance, our corpus may look like this:

```

0 This was not as much fun as I thought it would be .
1 I had a lot of fun on this assignment and learned a lot .
-1 It would be more fun if we had more time to work on it .
2 I didn't think programming in Java could be so much fun !
-2 I would have preferred an easier assignment .
2 I can't think of anything more fun than learning Java !
1 Is it fun or not fun to try to find funny ways to have FUN ?

```

Each statement is labeled with a score from -2 to 2 (the number at the start of each line) as follows:

- 2: very negative
- 1: somewhat negative
- 0: neutral
- 1: somewhat positive
- 2: very positive

To determine the overall sentiment of the word "fun," we take the average of the occurrence scores. In this case, it would be $(0 + 1 + -1 + 2 + 2 + 1 + 1 + 1) / 8 = 0.875$.