Task3

Tao He

12/7/2021

Sentence-level Analysis

tnum package

Using a single word as a marker for sentiment analysis can be less than ideal. This is because nearby words add context - in particular, negation makes analysis tricky. For example, "magic" is considered a positive word, while "dark magic" is considered a negative meaning. To further analyze **Heidi**, I decided to look at the entire sentence where these emotional words are found.

Sentiment by sentences

Then I use sentiment function to get sentiment score ,and then, group by these scores with section to get the average result. I plot the average sentiment score though sentence by sentence.
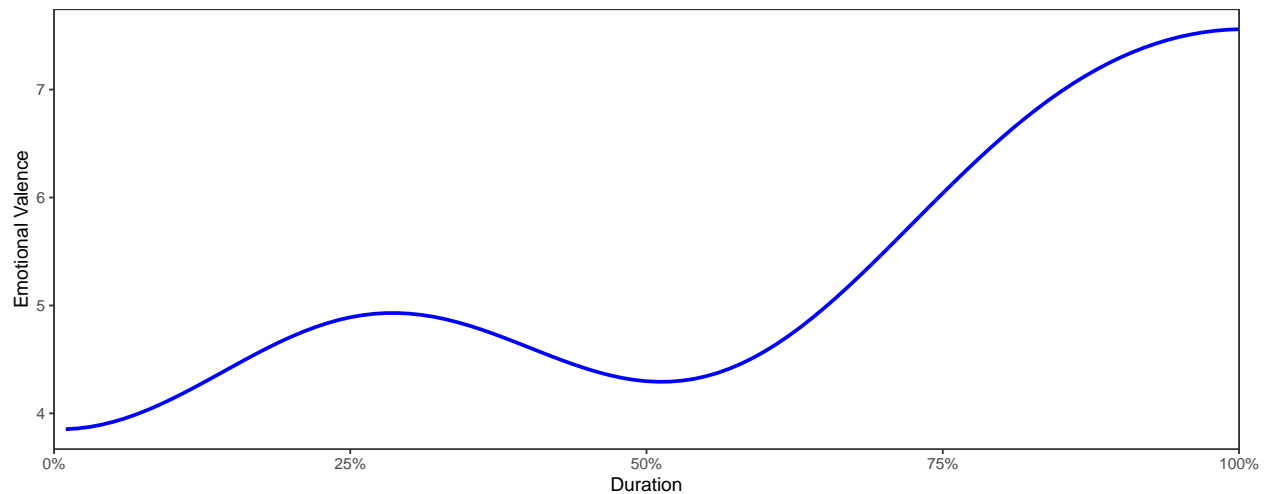

Figure 1: The average sentiment score from high to low.

Figure 2 shows the distribution of sentiment level of the whole book measured by each sentence.
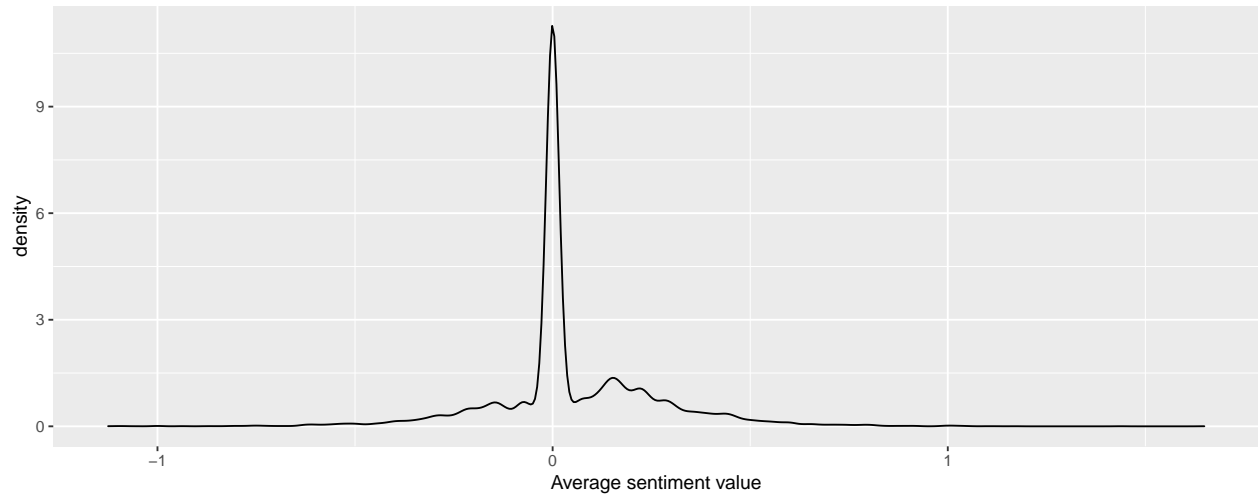
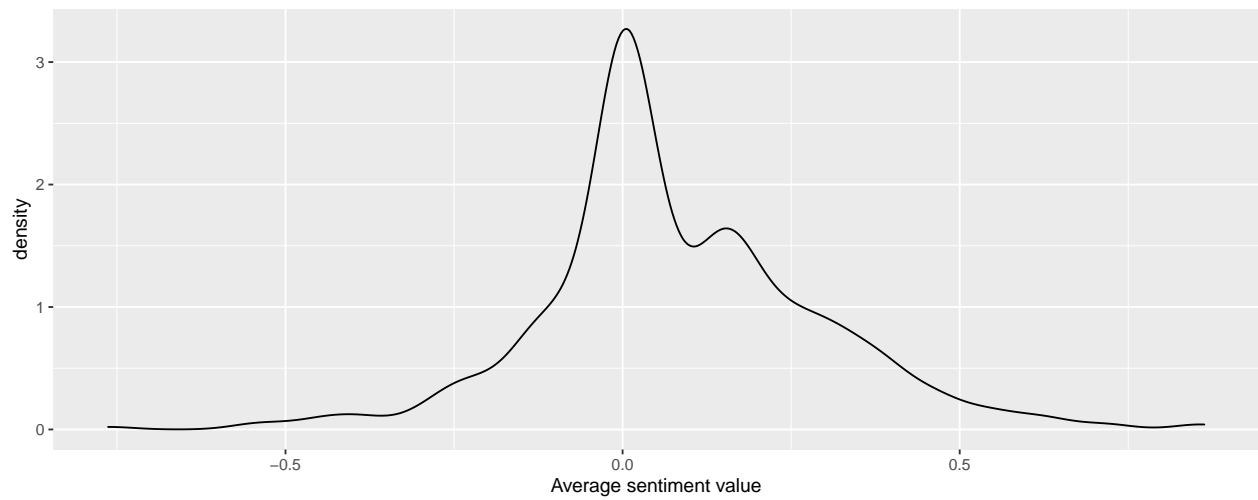Figure 2: Distribution of average sentiment level of the whole book

Figure 3 shows the distribution of sentiment level of each sentence of the first 500 sentence of the whole book.
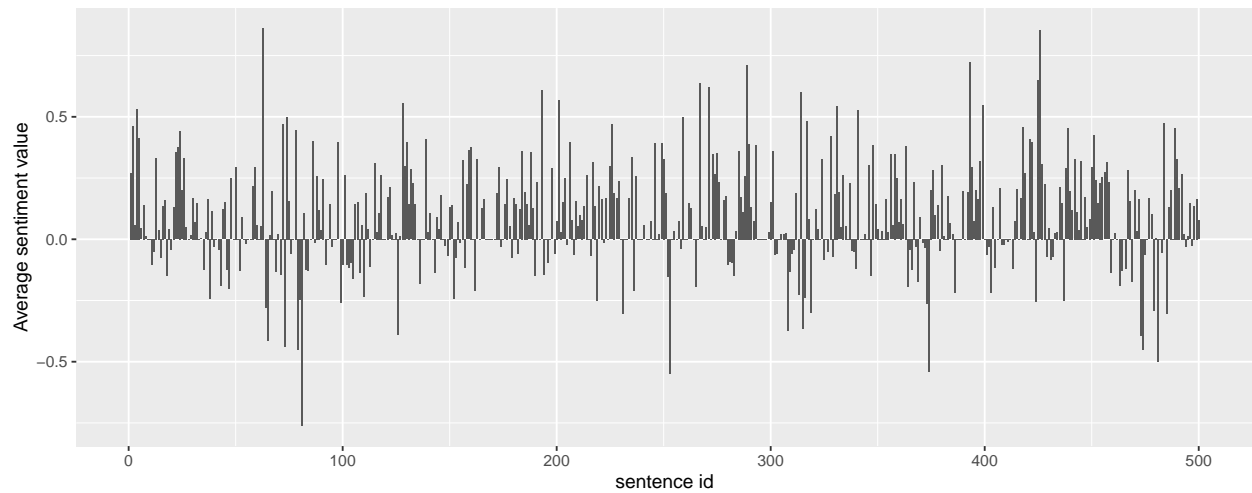

Figure 4 shows the sentiment score for each sentence of first 500 sentence of whole book.

Compare this analysis with the analysis I did in Task TWO

Since the sentiment function and the sentiment lexicons have different scales for measuring sentiment. I used the sentiment lexicons in task2 to count the positive and negative words in every 80 lines, however, in task3, I used the sentiment function is to rate the sentiment of each sentence.

Therefore, I was supposed to use the sentiment dictionary to count the positive and negative words in each sentence to see the overall sentiment trend.

Then, I extracted all the words in each sentence and used the sentiment lexicons I used in task2 inner_join().
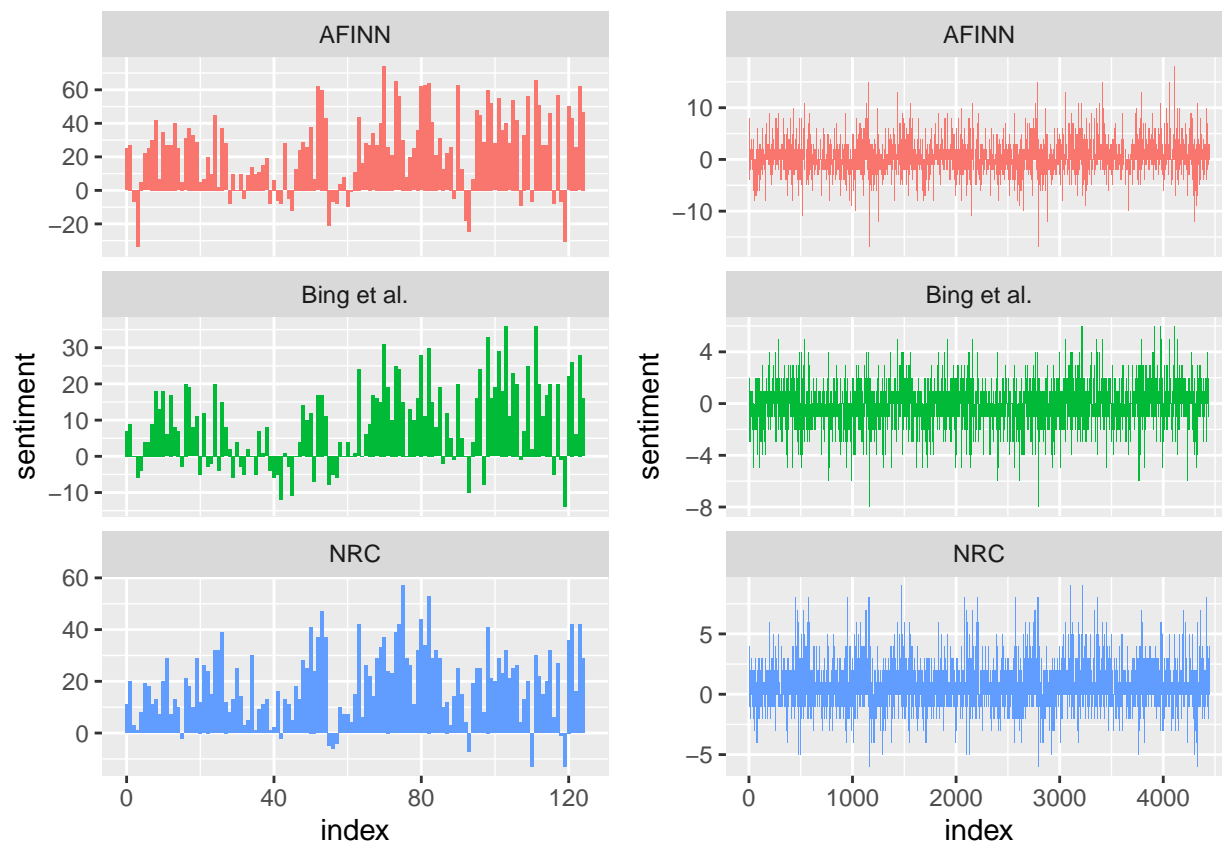
In the figure above, we can see trends, such as positive and negative directions. In conclusion, the emotional trends are roughly the same throughout the whole book, all in a positive direction. However, it is obvious that the level of emotion measured by the emotion lexicon is more positive than the sentiment function.

Then, I decided to look at the difference between top 7 words that contribute to positive and negative sentiment in Heidi by two methods.
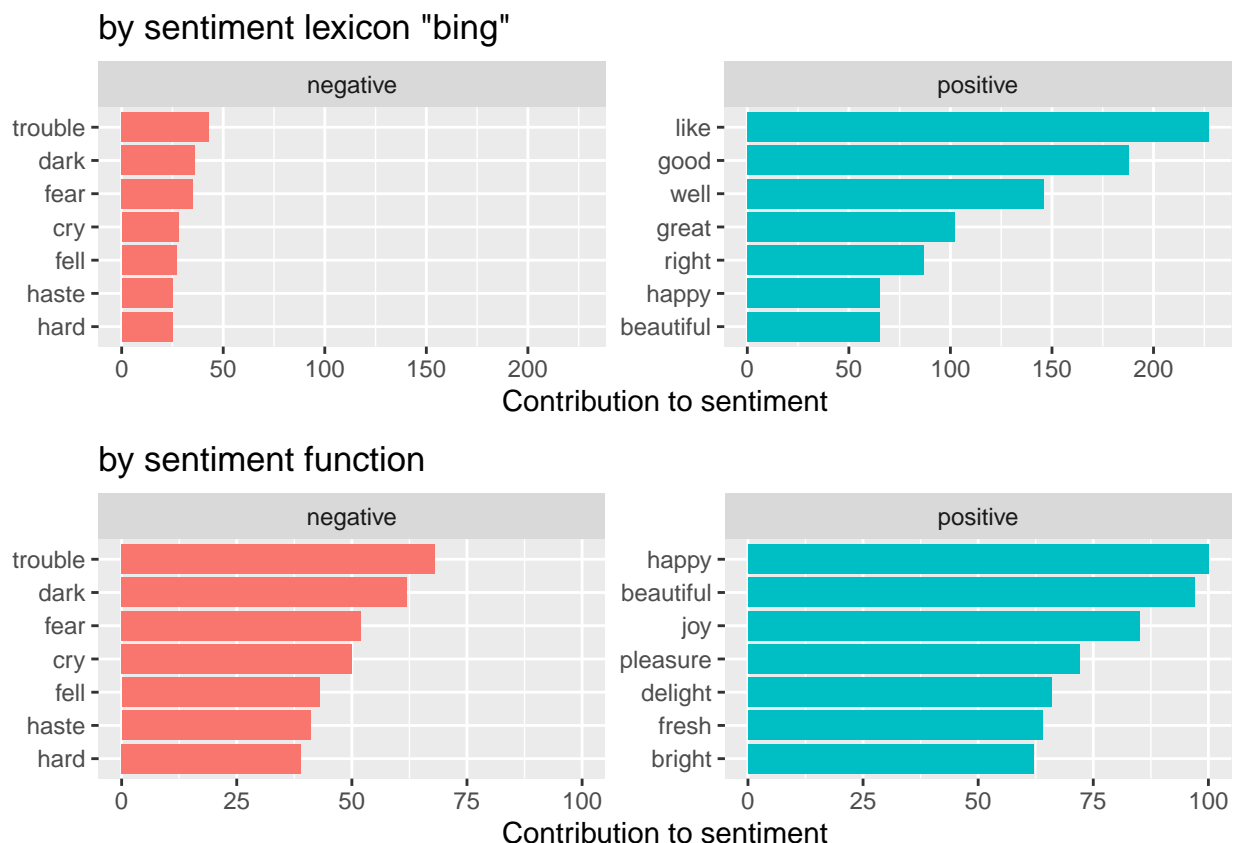
Figure 3: Most common positive and negative words in Heidi by two methods

We found some differences in the terms that contributed the most in the different methods. In the negative module, the top seven terms that contributed most to negative emotions were exactly the same. while in the positive module fast, only happy and beautiful were in the top seven.

Also, we found that in the former method like was used as the word that contributed the most to the positive sentiment, but it should be noted that like is positive when used as a verb, but does not mean anything when used as a preposition, so it is more numerous than any other word because it is counted when used as both a verb and a preposition.

Conclusion

Generally speaking, the conclusions from the sentiment function are a little more objective, since different sentiment dictionaries have different definitions of whether words are positive or negative, and sometimes they even misunderstand some words for positive or negative ones. But there is still some limitation, since this is only after I have analysed the book heidi, and may not be applicable to all text analysis.

Reference

Data Source:
Johanna Spyri, Heidi, (September 1, 19989), from https://www.gutenberg.org/ebooks/1448

Works Cited:

David Robinson, gutenbergr: Search and download public domain texts from Project Gutenberg, (May 28, 2021), from https://cran.r-project.org/web/packages/gutenbergr/vignettes/intro.html

Julia Silge and David Robinson, Text Mining with R: A Tidy Approach, (June 8, 2017), or from https://www.tidytextmining.com/

Github[https://github.com/MA615-Yuli/MA615_assignment4_new/blob/main/TASK3.Rmd], https://github.com/MA615-Yuli/MA615_assignment4_new/blob/main/TASK3.Rmd

Haviland Wright, tnum_instructions and examples, https://learn.bu.edu/ultra/courses/_80585_1/cl/outline