

Text mining - Task Two

Qiannan Shen

12/6/2021

Task one

I choose the fiction *The Odyssey* by Homer as the book I focus on my text analysis. This book has 24 BOOKS which I regard as 24 chapters in a whole book, and each chapter has paragraphs and sentences that it is available of a short analysis. After picking the target book, I downloaded the book source using gutenbergr package.

Task two

Sentiment analysis by three methods

First, the analysis uses the function `unnest_tokens()` to take the text of the novels and convert the text to a tidy format. Then, I use the sentiment lexicon methods Bing, Affin, and NRC in the `tidytext` package to do the text analysis to digest the sentiment through the narrative of your book. The function `get_sentiments()` allows us to get specific sentiment lexicons and `inner_join()` to perform the sentiment analysis. And eventually, I could compute the sentiment score and do visualizations through the storyline of the book. The result comparing the three sentiment lexicon methods is shown below in Figure 1. The overall sentiment result by Affin and Bing methods tend to be similar, while the result by NRC would show a little different. At the beginning of the book, the sentiment mostly shows the positive sentiment. Though, the sentiment through the following part of the book fluctuates frequently that intersperses with positive and negative sentiment.

The *Odyssey* tells the story of a heroic Odysseus who fights with many antagonists on the arduous journey home from war for 10 years. On his way back home, we could explore the inspiration about fate, recitation, and the power of civilization and barbarism. He finally returned to his hometown Ithaca, defeated the rude suitors who camped out in his palace, and reunited with his loyal wife Penelope. And this could well explain the fluctuated sentiment change through the book since he suffered the 10-year-long arduous journey and he had but overcome lots of challenges with pain and joy.

Positive & Negative sentiment plot

Then, the report wants to figure out the frequency of the negative and positive words shown in the book, and the result is shown in Figure 2. For the positive chart, “good” is the most frequent word shown in the whole book, and meanwhile “great”, “like”, and “heaven” is also the common positive words that appeared. For the negative chart, “stranger” ranks first in the count of the negative words, and the words “killed”, “dead”, “fell”, “death” is also common in the book. It is obvious that on his way back home Odysseus experienced many horrible and magical things and people fight and died around him, so that the words like “killed”, “dead” appear so frequently.

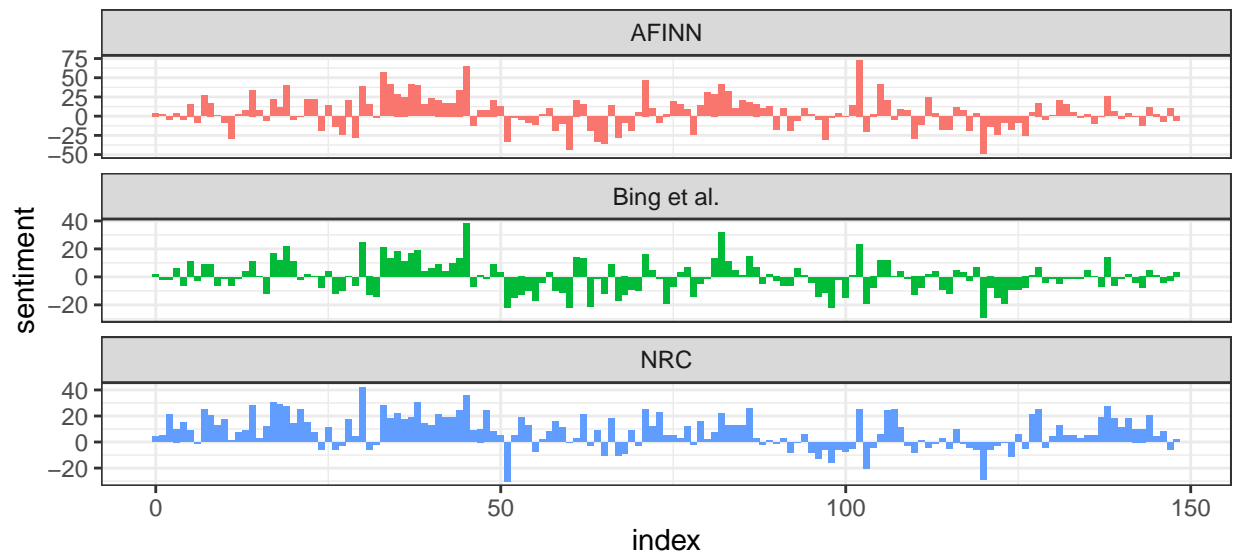


Figure 1: Sentiment Plot by Three Methods

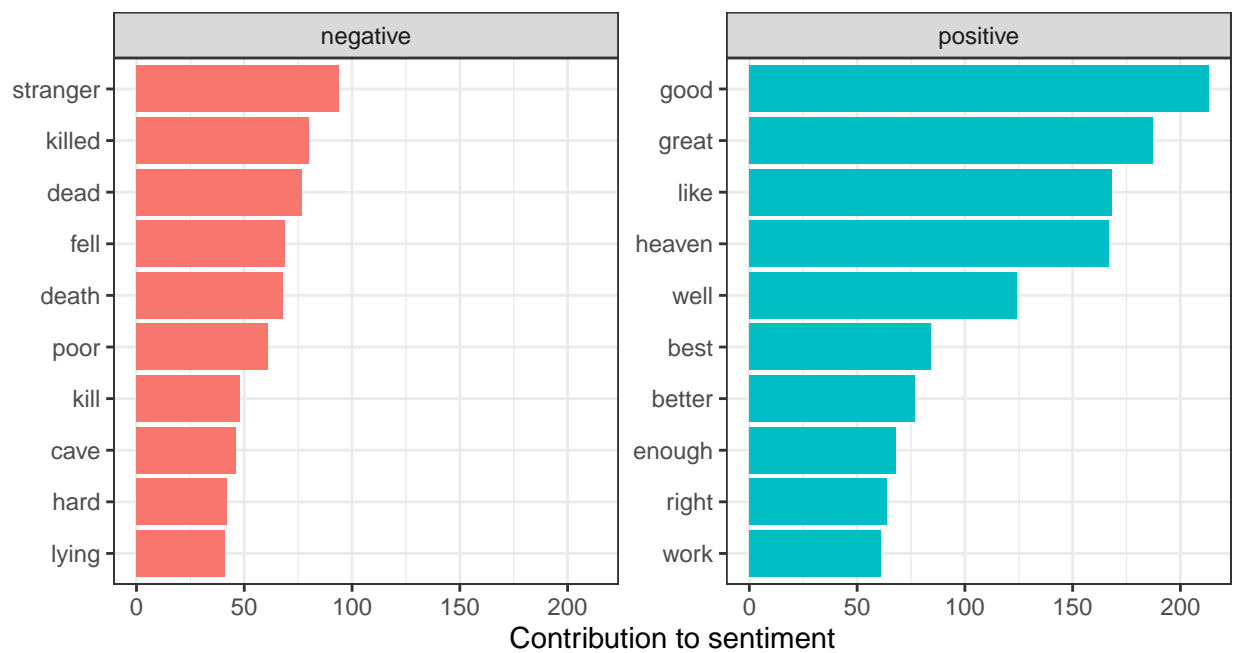


Figure 2: Negative & Positive charts

Figure 3 shows the frequency of the words that appear in the book. It indicates that the “Ulysses” which is the Latin name of Odysseus, and his son’s name “Telemachus” are the most common words in the book. They are also the main characters in the story. Other words like “house”, “home”, “son”, “father”, and “suitors” also show frequent, which tells his desire of going home.



Figure 4 displays the sentiment words frequency where red words represent the negative sentiment and the blue words represent the positive sentiment. This is the word cloud version for data in the Figure 2.

The lexicon method Loughran categorizes words in a binary fashion into categories of positive, negative, uncertainty, litigious, constraining, superfluous. I only choose the category of positive and negative, and calculate the sentiment score. Then I plot the Loghran lexicon method with other three methods to do the comparison. The plot indicates the sentiment are mostly positive in the book which is much different with the other three. So I could suggest that the Loughran lexicon method fits not really well in this book.



Figure 4: Sentiment Wordcloud Plot

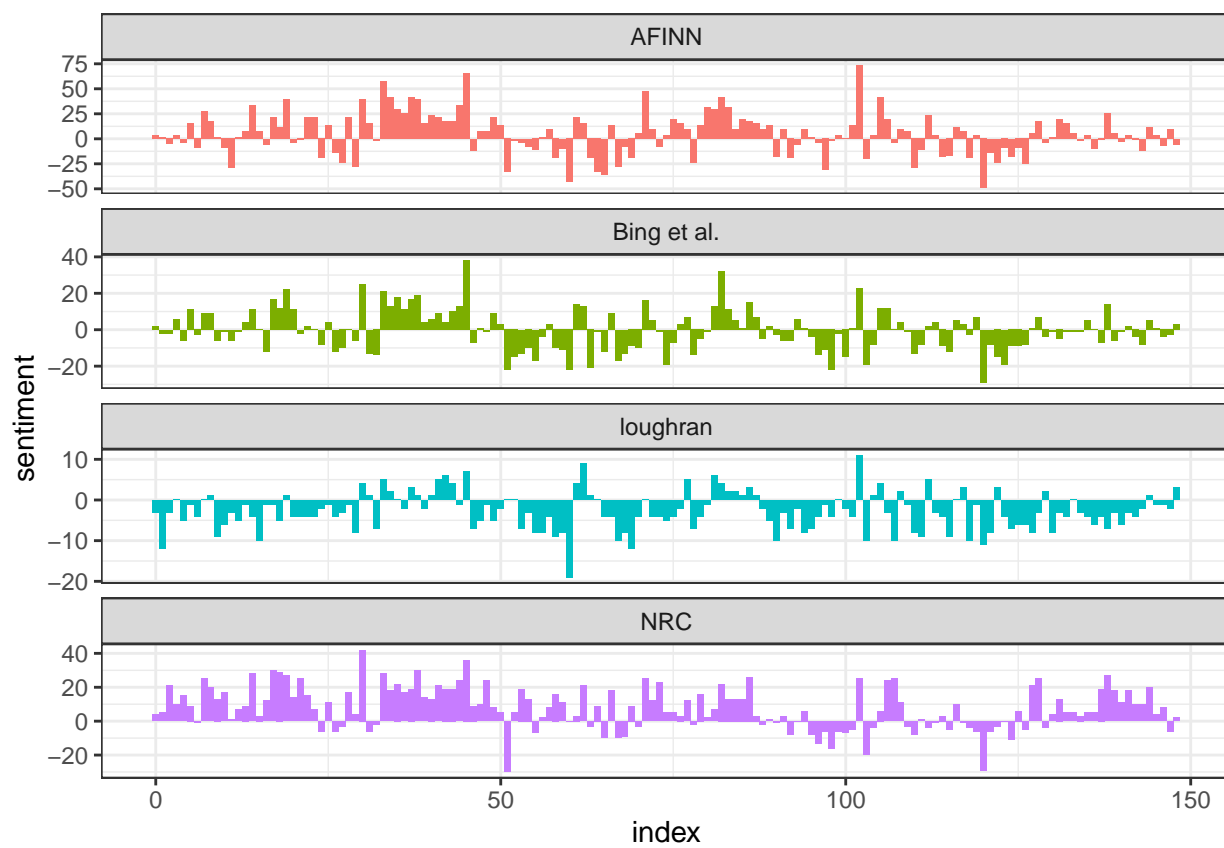


Figure 5: Sentiment Plot by Four Methods

Citation

Text Mining with R <https://www.tidytextmining.com>