

Text analysis task1 and task2

Daniel(Chen) Xu BUID:U49903384

12/07/2021

Task 1 Pick a book

I choose **Twenty Thousand Leagues under the Sea** written by Jules Verne as the book for my text analysis assignment.

TASK 2 bag of word analysis:

I use three types of sentiment analysis methods AFINN, Bing and NRC to plot barchart plots to compare these methods. From the graph below, the AFINN and Bing method fits better. From the sentiment analysis plot, we can see that scores for positive/negative sentiment in **Twenty Thousand Leagues under the Sea** seems balanced throughout the whole book.

Twenty Thousand Leagues Under the Sea tells the story of marine biologist Pierre Aronnax, his manservant Conseil and harpoonist Ned Land, who – after joining the hunt for a mysterious sea monster – are thrown overboard when the monster attacks and find themselves prisoners of Captain Nemo, probably one of Verne’s most memorable yet elusive characters. On board the Nautilus, a technologically advanced submarine that everyone has mistaken for a sea monster, the three companions get to experience the vast and endlessly fascinating world under the sea. In 1866, a monster suspected of being a narwhal was discovered on the sea, and Professor Aronnax and his servant Conseil were invited to join the hunt. In pursuit, they and Their harpooner Ned Land fell overboard and landed on the monster’s back. They discovered that the monster was not a narwhal, but a strangely constructed submarine. Nemo secretly built the submarine on a deserted island in the middle of the ocean. It was sturdy and powered by seawater. Captain Nemo invited Aronnax on a voyage under the sea. They set out from the Pacific, passed coral islands, the Indian Ocean, the Red Sea, the Mediterranean Sea, the Atlantic Ocean, and saw many rare plants and animals and strange sights in the sea. On the way, I also experienced a lot of dangerous situations, such as stranding, aboriginal siege, shark fighting, iceberg blocking, octopus attack and so on. Finally, when the submarine reached the Norwegian coast, the three men left without saying goodbye and returned to his hometown.

That is to say, at the beginning of the book, the sentiment of the book is negative, but soon it converts into positive sentiment. And converts to be negative when goes to the end of the book. However, it is difficult to identify which of the two methods is better. In the following task, I use Bing method to conduct further analysis.

Figure 2 shows negative and positive word count of each word. For the negative chart, stuck is the most common word throughout the whole book. Lost and monster rank the second and third place respectively. For the positive chart, like is the most common words throughout the whole book. Well and great rank the second and third place respectively.

Figure 3 displays word cloud which shows the frequency. As we can see, captian, nautilus, nemo are the words have the highest frequency among all the words. It is reasonable because they are the main characters in that fiction book. In task 3, I will use two of three characters to conduct further analysis.

Figure 4 generally converts Figure 2’s information into word cloud.

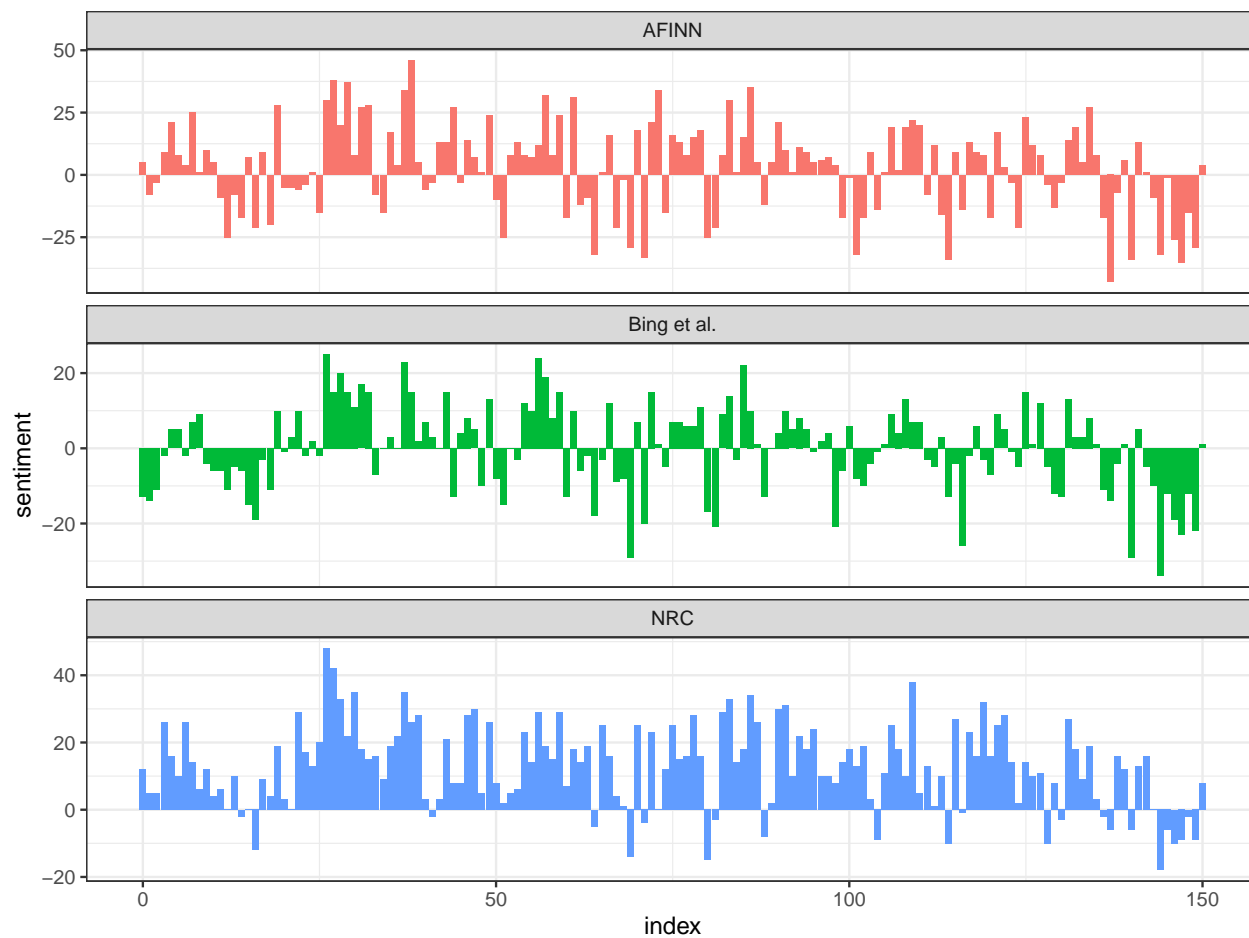


Figure 1: sentiment plot

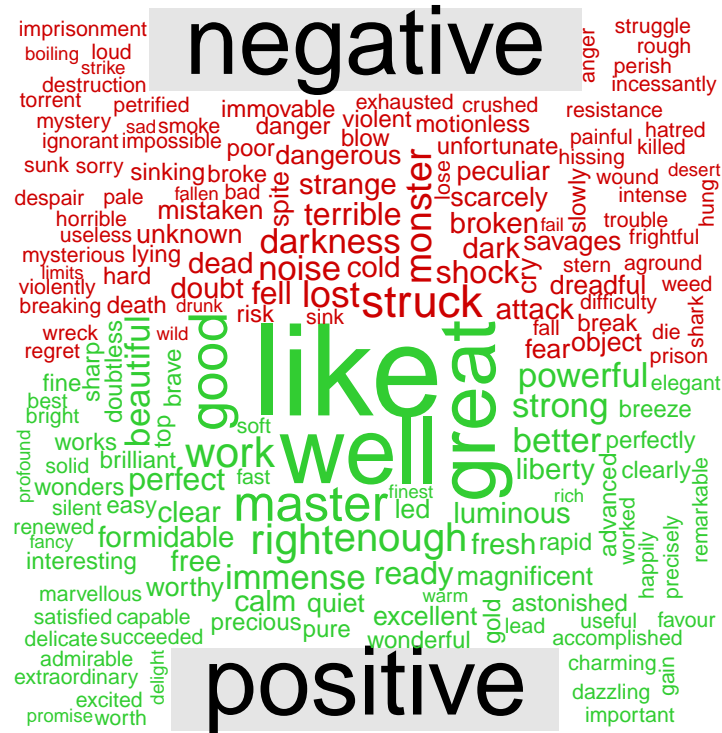


Figure 4: sentiment word cloud

Task 2 Extra Credit:

- Reference for extra credit:
 1. https://emilhvittfeldt.github.io/textdata/reference/lexicon_loughran.html
 2. <https://sraf.nd.edu/textual-analysis/resources/#Master%20Dictionary>
 3. https://search.r-project.org/CRAN/refmans/textdata/html/lexicon_loughran.html
- What is Loughran-McDonald:

English sentiment lexicon created for use with financial documents.

As we can see from the below figure, the sentiment values generate from Loughran-McDonald lexicon text analysis method are biased to negative side, which does not accord with the full text sentiment trend.

