MA615 Assignment4

Text Analysis of Pierre and Jean Task Two

Yanbing Chen

2021/12/6

Task One: Pick a book

According to the task one requirement, I pick a fiction book named Pierre and jean written by Guy de Maupassant on The Gutenburg Project to do text analysis. The book size is 284kB.

Here is the introduction of this book:

The Pierre and Jean , a novel by the French writer Guy de Moposan, which is a serious psychological novel about Pierre and Jean are two brothers. Pierre and Jean, the son of retired businessman Laurent's father, are both adults and graduates of medical and forensics, respectively. Besides,they are both facing the problem of struggling to set up a family. One day, Jean suddenly got an inheritance, which was left to him by a former old friend of the Rolands, Marechar, which made him a rich man overnight. The proporty assigned by the two brothers are unfair, making Pierre began to feel a little jealous, wondered, and finally found that his brother was the illegitimate son of his mother's secret love affair with Marechar, so he was in great pain. His love for his mother turned into resentment, he could not torture her with rude attitude and words, and conflicted with Jean. There was such a deep emotional rift that he could no longer spend time with his family and had to leave his family to work as a doctor on a cruise ship between France and the United States. The whole family sent him offshore, and he had a feeling of eternal farewell. His mother was heartbroken that he would never see her son again.

The plot of the novel is bland, but the whole work is not short of exciting power. Instead of resorting to strange ideas and extraordinary events, the author acquires touching power through the ethical contradictions of a family with tragic colors.

Task Two: Bag of words analysis

In this part, I tidy information in the book to gain a clean data file called 'tidy_pierre'. Because there are many options for sentiment lexicons, I try three methods that are AFINN,BING, and NRC to make box plot to find which measure is the best. As figure1 shown, the BING fits better, therefor, finally I chose to apply NRC method in deeper text analysis.

From Figure 1 it can be seen from this chart that the change of the emotional trend of vocabulary is in line with the development of the storyline of the novel. At the beginning, it shows a positive trend, which corresponds to a plain but happy life for the family. However, as Jean gain a large inheritance and his brother discovers that their mother has an affair with someone else, the emotional direction of the novel begins to turn negative, and after this turning point, the whole thing is in a negative mood untill the end of the story.

Figure 2 counts the number of positive and negative sentiment words in `Pierre and Jean`, and lists the top ten positive and negative emotional words that appear most frequently in the book. Base on the picture's information, we can see that positive words appear more frequently than negative words. In negative words, the most frequent occurrence is "poor", appearing nearly 30 times. In positive terms, "like" appears more than 90 times, and good, well follow by it.
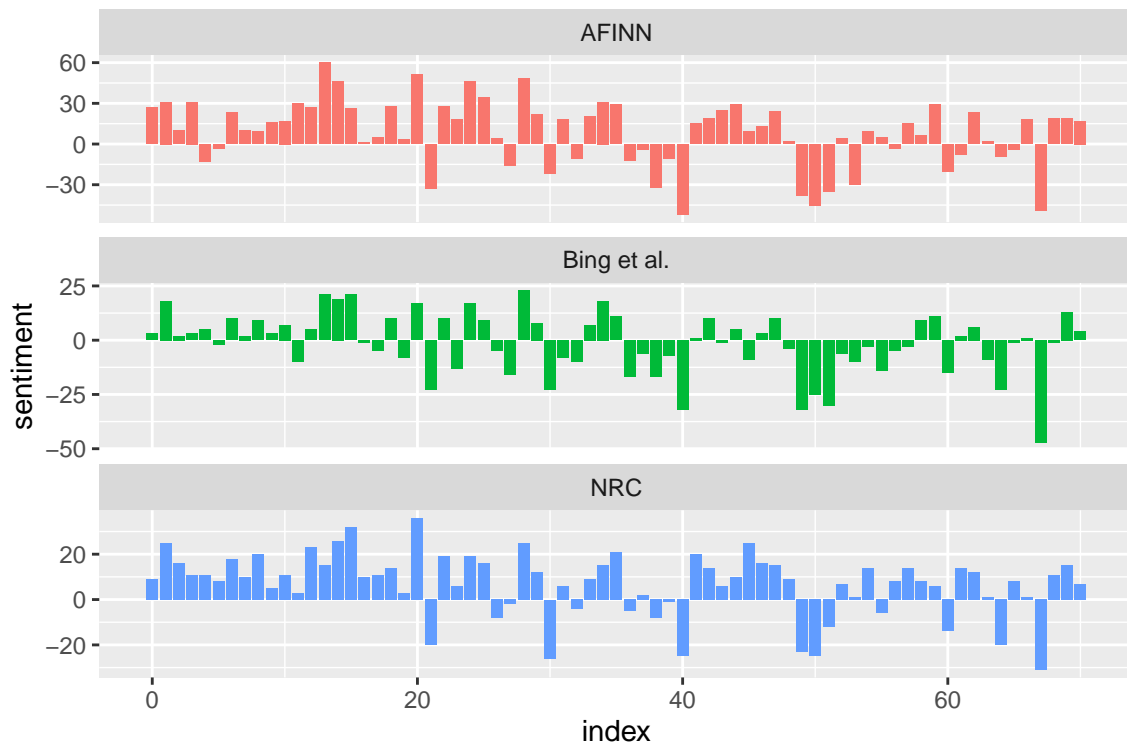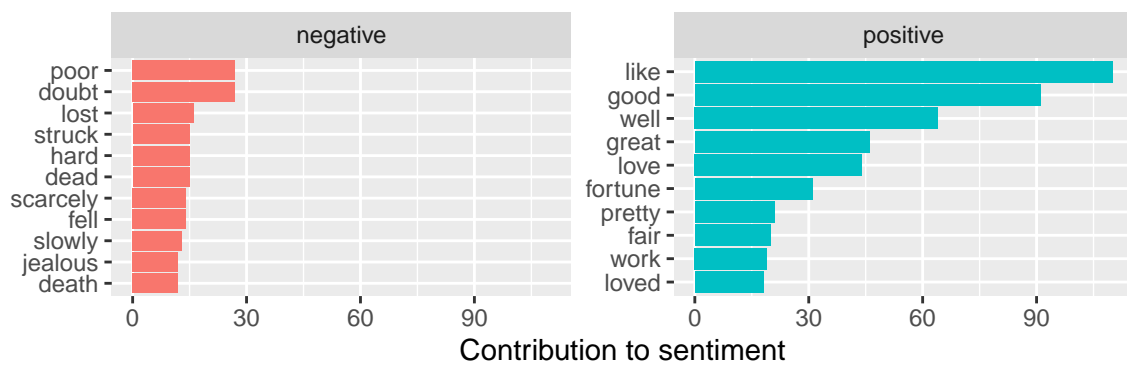
Figure 1: Sentiment plot

Figure 2: negative positive words count

Figure 3: Word cloud

Figure 4: Sentiment word cloud

Figure3 illustrates that the most frequently used words in this book are jean, mother, father, time, son, day and so on. The size of the font and the depth of the color can also tell that words related to a person's name or family members appear more frequently.

Figure4 shows a comparison between positive and negative words in this book, and the size of words text is in proportion to its frequency within its sentiment. Apparently, the font size of positive words are much larger than negative words, which also explains that the frequency of positive words are higher than the negative words'.

Reference

1.https://github.com/MA615-Yuli/MA615_assignment4_new 2.https://www.gutenberg.org/ebooks/3804