

Task Three (Word-Level Analysis)

Jiayue Lu

“Afinn” lexicon will return a numeric score between -5 and 5. A higher value means the sentiment is more positive, while a lower value means the sentiment is more negative. However, it is difficult to set a weight function for this score while calculate the total level of sentiment of larger unit. So I pick “Bing” and present its result, which directly return positive or negative. (NRC also directly returns the binary label).

Since we will compare the result with Task 3, I need to unify their time axes. So I use the line number in the book as the index and cut them into batch of 85. It means each batch will have 85 lines and the associated data.

“Tidyttext” package is utilized to words tokenization. And I also use anti-join to remove the stop words and inner-join to calculate the sentiment. In my definition, the degree of sentiment is number of positive words in a batch - number of negative words in that batch.

Finally, as shown in the plotting, the sentiment fluctuation is consistent with the plot line of the book.

