Task3Report

Mi Zhang

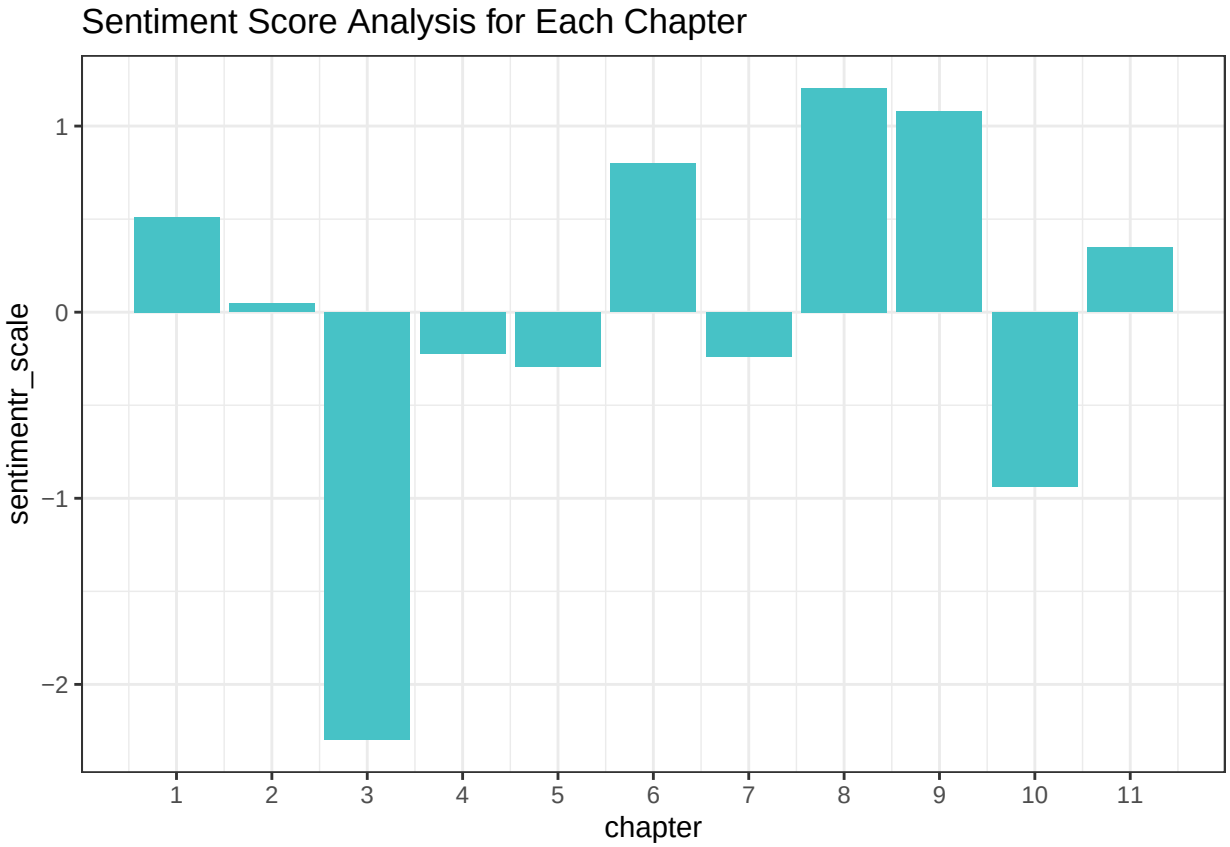
12/7/2021

Task Three

This task is about how we can upload our book into the tnum space and the table below proves that I am able to query my book from tnum space.

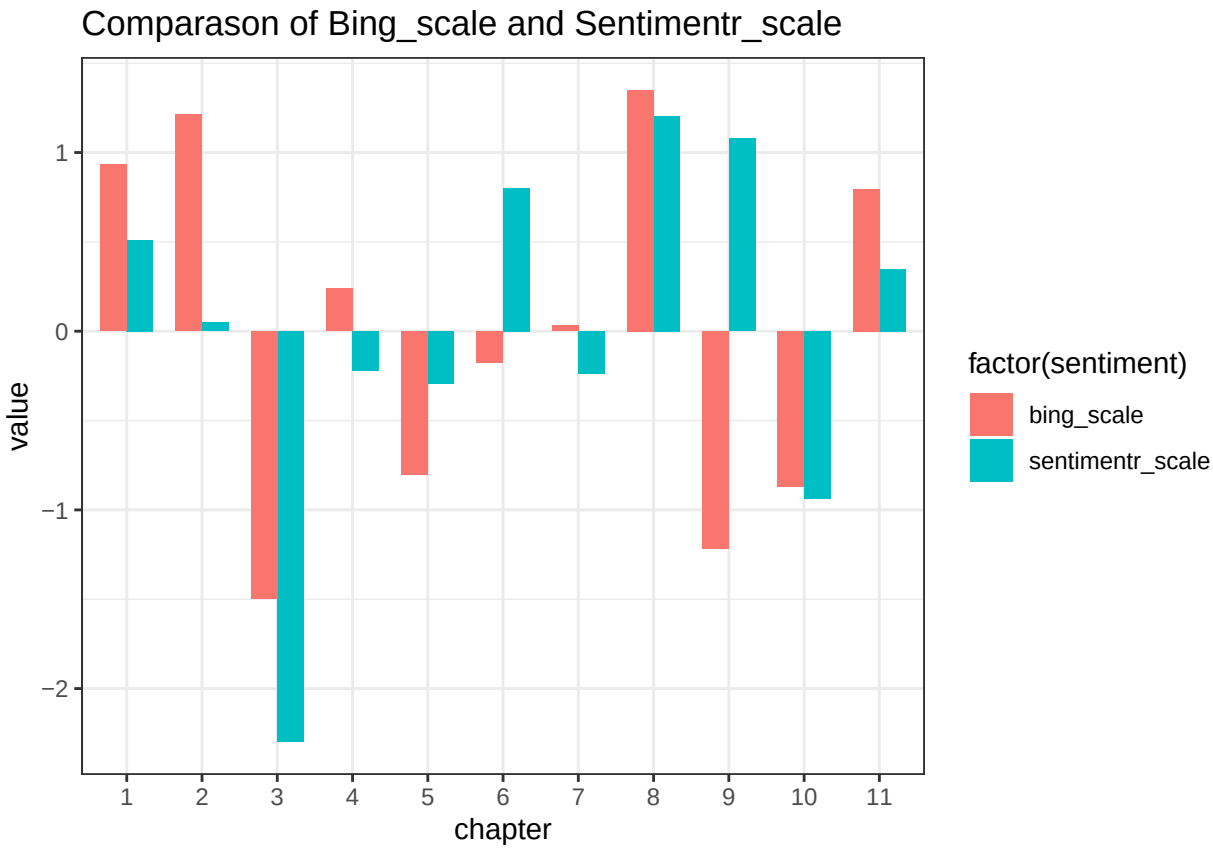
subject	property	string.value
love_paddington/section:0001/paragraph:0001/sentence:0001	text	"Children had been sent off to Sunday school, a
love_paddington/section:0001/paragraph:0001/sentence:0002	text	"As a consequence, in the streets off the main t
love_paddington/section:0001/paragraph:0001/sentence:0003	text	"Praed Street was different. . Praed Street was
love_paddington/section:0001/paragraph:0001/sentence:0004	text	"Praed Street plumed itself on the fact that it
love_paddington/section:0001/paragraph:0002/sentence:0001	text	"Even on a Sunday afternoon, and certainly at
love_paddington/section:0001/paragraph:0002/sentence:0002	text	"the thirst for journals at E."

Now, I am going to using sentimentr to score the sentiment words in the book. The plot below shows an average sentiment score in each chapter. I noticed that chapter has low sentiment score and went back to skim the chapter. It made senses to me because chapter three mentioned that a well prepared trip had been ruined by a storm and several characters in the book refected bad moods due to the storm. In addtion, towards the end of the book, the main character Gertie and Henry struggled about their social class gap and financial problem, so the average sentiment score goes down on chapter ten which made sense to me as the plotline moves on.



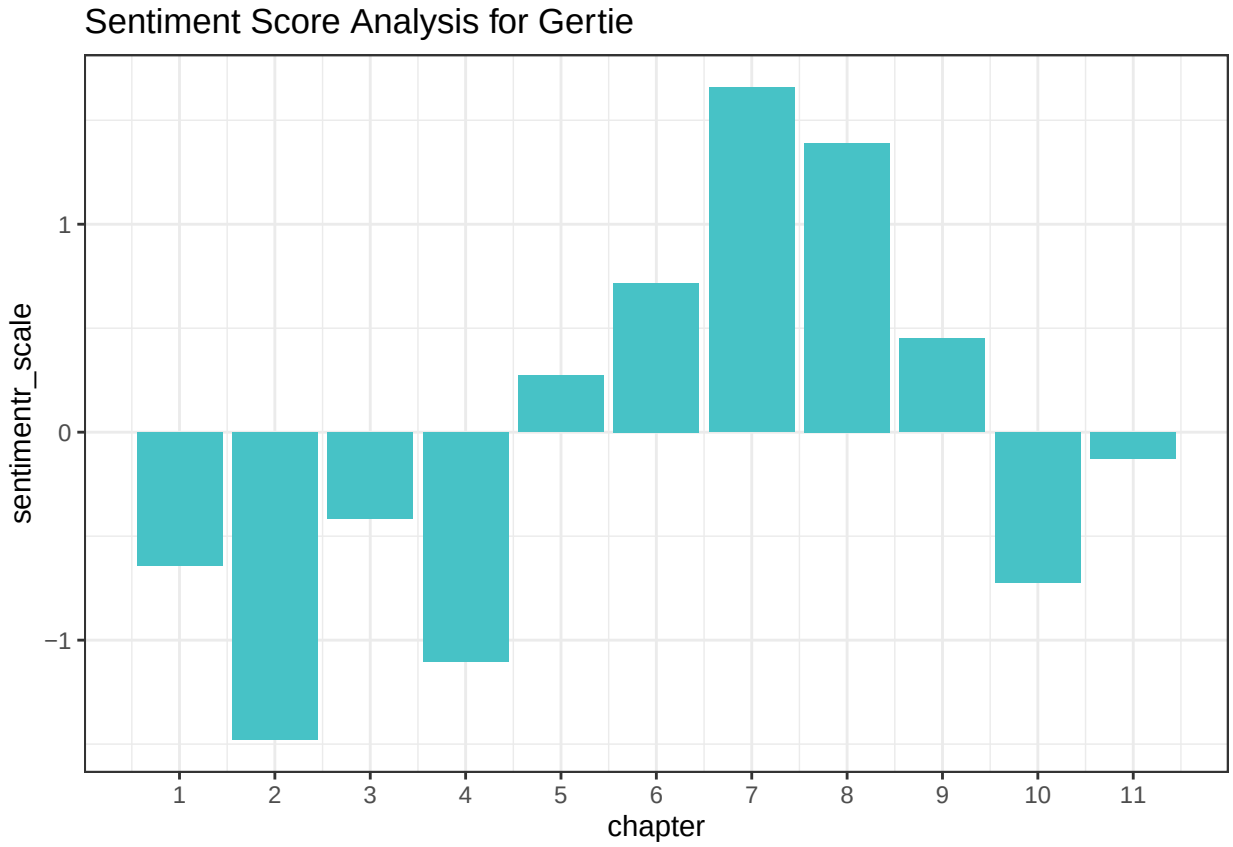
Compare two methods

I choose bing lexicon for the comparison because bing lexicon categorize words in binary fashion and it would easier to compare to the sentiment score in sentimentr. As the plot shows below, I noticed that both methods seems to peak and dip at same places except for Chapter Four, Seven, and Nine. But for the overall plotline, those two sentiment analysis method matched.

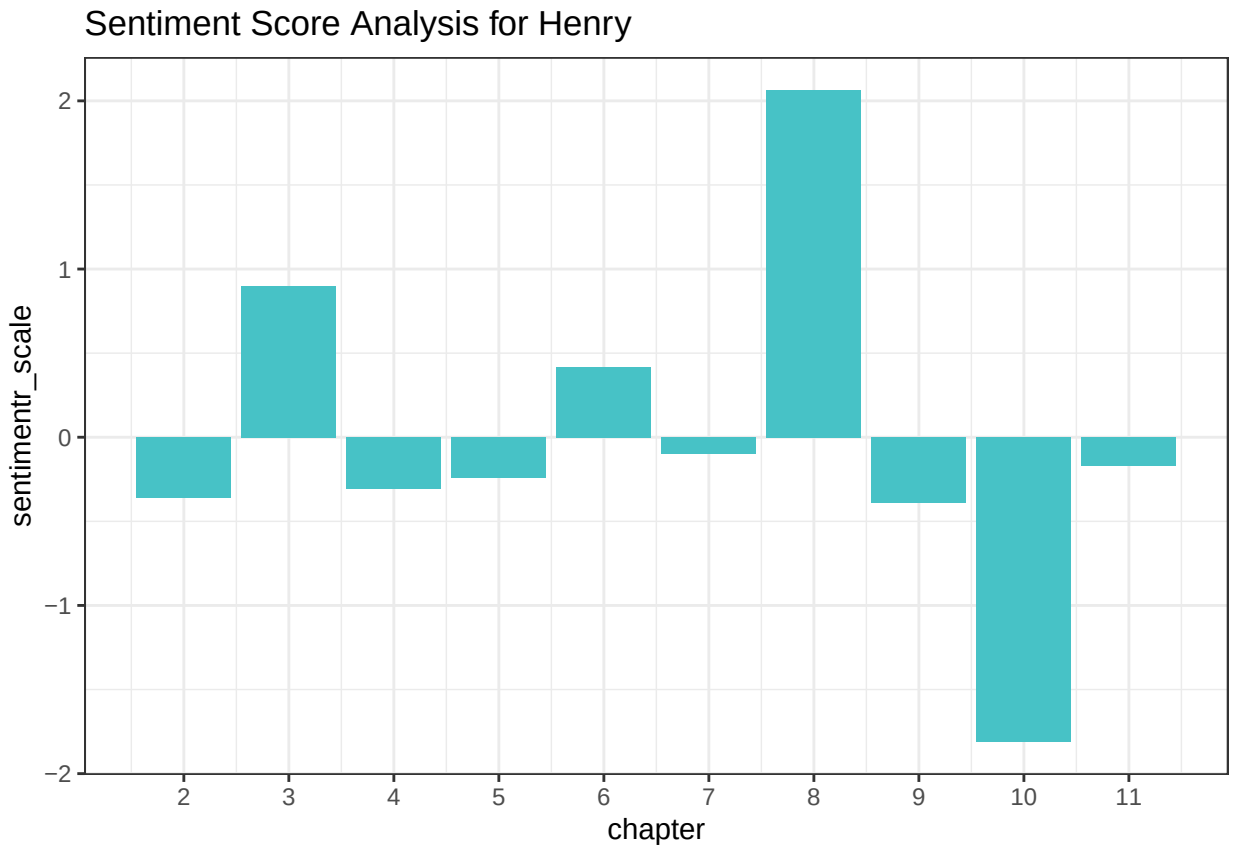

Comparason of Bing_scale and Sentimentr_scale
value
chapter
1
0
−1
−2
1
2
3
4
5
6
7
8
9
10
11
factor(sentiment)
bing_scale
sentimentr_scale

Extra Credit

Gertie and Henry are the two main character of the book, "Love at Paddington", and I am going to do sentiment analysis of them in each chapter using tnum.query. Tnum shows that Gertie's name appeared in the book for 269 times. As we can see the plot below, Gertie involved in a wider range of sentiment change from negative to positive in the book . This is reasonable because the book focus more about how Gertie as a female at her time is able to break people's stereotypical opinion about a middle working-class girl who had the romance relationship with Henry from the upper-class society. Through the sentiment analysis, we can also see how Gertie's sentiment struggles about the relationship.



The plot below is using the same tnum.query technique to analyze sentiments related to Henry in each chapter. Tnum shows that Gertie's name appeared in the book for 96 times.We can see that Henry involved in a small range of sentiment changes throughout the book except in Chapter 8 and 10. I guess the relation between him and Gertie is changing him internally.


Sentiment Score Analysis for Henry
sentimentr_scale
2
1
0
−1
−2
2
3
4
5
6
7
8
9
10
11
chapter

For the plot below, I am just curious about how the sentiment range change between the two characters.

Acknowledgement

I would like to thank my classmates(Yifeng He, Shicong Wang, Boyu Chen) and TA Runqi Zhao who were more than generous with their expertise and precious times. A special thanks to Yuli Jin who explained the project to me and helped me with coding during the entire project.

reference

- https://github.com/orazdow/TrueNum
- https://www.tidytextmining.com/dtm.html
- https://cran.r-project.org/web/packages/sentimentr/sentimentr.pdf
- Professor Haviland's lecture notes