Task3

Jinyu

12/12/2021

Task 1: Pick a Book

The book I chose is written by Charlse Dickens, and the book name is "Hard Times" and shortened as "A Christmas Carol" because we are gonna spend our final time of the year at Christmas.

task 3 sentence-level analysis

Tnum

Now, I input my book, A Christmas Carol, into tnum test2 space, the following tables are part of the book.

Table 1: First 100 Query results in tnum

subject	property	string.value	numeric.value	error	unit	tags	date	guid
charlse_dickens/a_christmas_carol_4/section:0001/paragraph:0001/sentence:0001	section:0001	1	NA	NA			2021-12-08	4666f077-68dd-4216-93b1-3de0cd849c0c
charlse_dickens/a_christmas_carol_4/section:0001/paragraph:0001/sentence:0002	section:0001	7	NA	NA			2021-12-08	254f70d5-2ff6-4fd9-8b44-02ff52379d8c
charlse_dickens/a_christmas_carol_4/section:0001/paragraph:0001/sentence:0003	section:0001	18	NA	NA			2021-12-08	cb674866-deb5-4934-87f4-af11883035f8
charlse_dickens/a_christmas_carol_4/section:0001/paragraph:0001/sentence:0004	section:0001	19	NA	NA			2021-12-08	1769dd67-6cc1-4dec-82b0-b767a103d900
charlse_dickens/a_christmas_carol_4/section:0001/paragraph:0001/sentence:0005	section:0001	8	NA	NA			2021-12-08	3d150658-3124-46ad-9aa5-42f0e30a3350
charlse_dickens/a_christmas_carol_4/section:0001/paragraph:0001/sentence:0006	section:0001	21	NA	NA			2021-12-08	ddf5b936-31c0-4c60-b07d-d7762a91927a

```
# get the text column
df_heading<- tnum.query('Charlse_Dickens/A_Christmas_Carol_4/heading# has text',max=90) %>% tnum.objects
kable(df_heading %>% select(subject:numeric.value),  caption = "Query results with subject heading")
```

Table 2: Query results with subject heading

subject	property	string.value	numeric.value
Charlse_Dickens/A_Christmas_Carol_4/text	heading:001	“CHAPTER I: MARLEY’S GHOST>”	NA
Charlse_Dickens/A_Christmas_Carol_4/text	heading:002		NA
Charlse_Dickens/A_Christmas_Carol_4/text	heading:003		NA
Charlse_Dickens/A_Christmas_Carol_4/text	heading:004		NA
Charlse_Dickens/A_Christmas_Carol_4/text	heading:005		NA

Now I list the sentiment score groupep by these scores with section to get the average result using Sentimentr package.

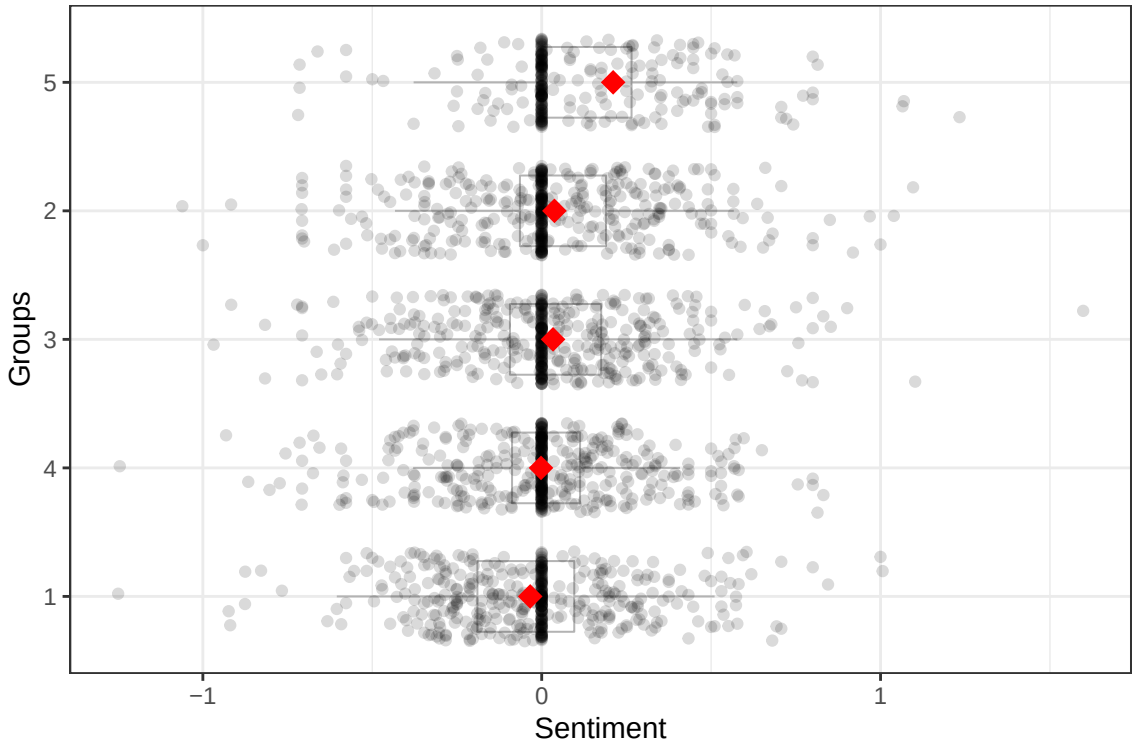

Figure 1: Sentiment Analysis by sentence

As we can see in Figure 6, I make the sentiment analysis by sentence and grouped the score of sentiment into 5 chapters. The red points represent the mean score for each chapter.

Compare this analysis with the analysis you did in Task TWO

We cannot compare the methods of package Sentimentr and Bing lexicon directly because they assess the sentiment in a different scale. At this point, it’s good to try some standardization methods. Here I will scale this 2 kinds of scores and show the sentiment analysis grouped by chapters and then we can tell the sentimental progressions in this fiction with 2 different methods.

In this case there are 3 scale methods I use: 1. standerdization 2. make a rank for these 5 chapter

According to the result, I would like to say the "" method is the best sentiment analysis method for this book.

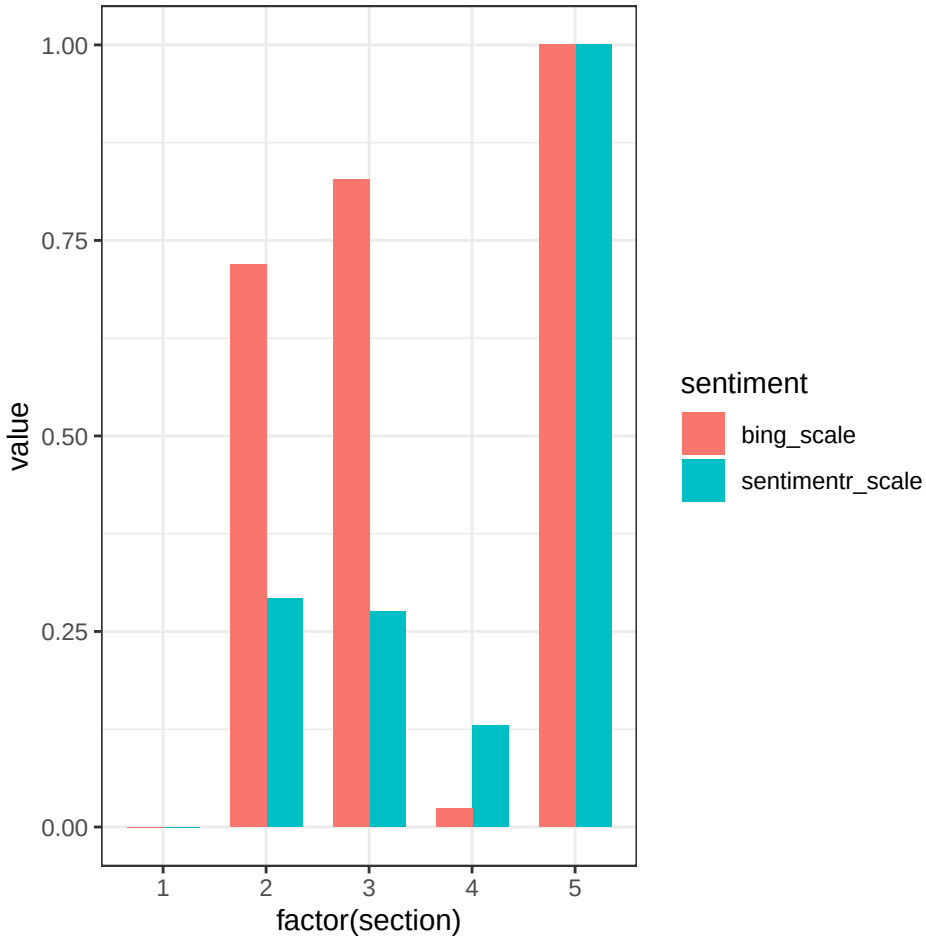

Figure 2: sentiment comparison- normalize

In Figure 7, when we use the normalization to process the scores, we can see there is 0 for both methods in Chapter 1, but for the latter chapters, using bing lexicon shows a more extreme sentiment (whether more positive or more negative), which seems to be more reasonable.

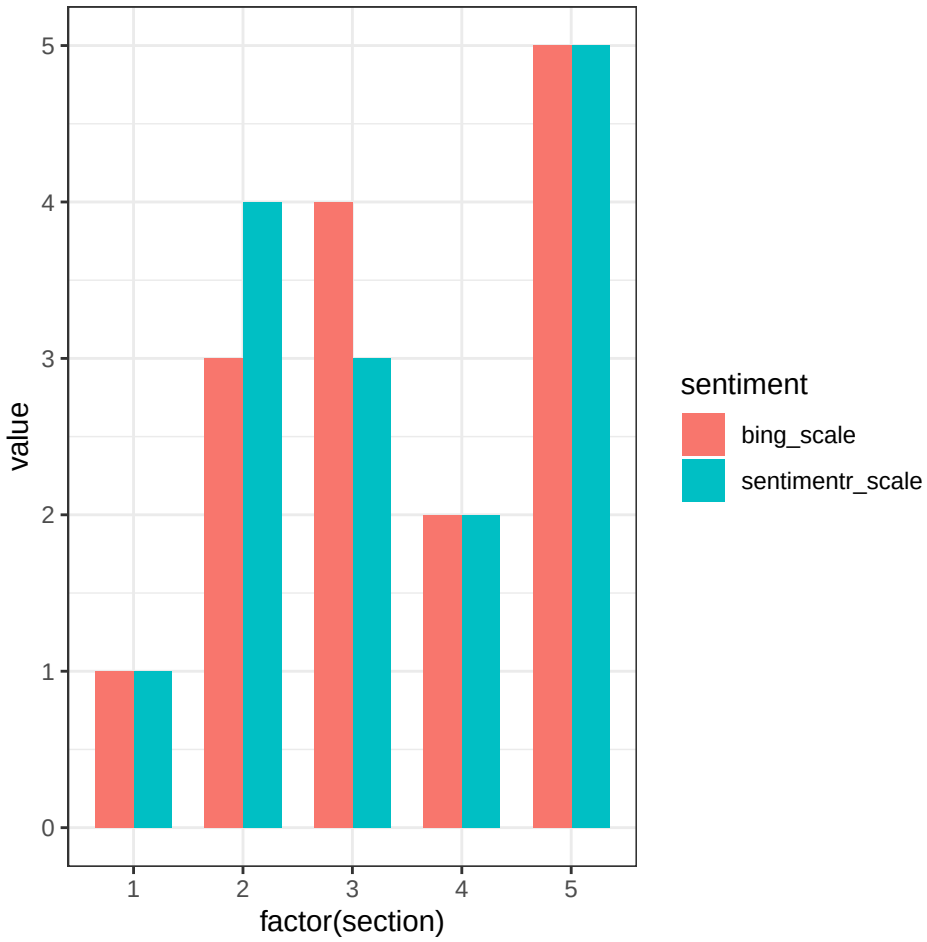

Figure 3: sentiment comparison- rank

In Figure 8, the larger the rank is, the more positive the sentiment is. So we can tell that there is a slight different in Chapter 2 and Chapter 3 in diferent methods and I think it's acceptable for both methods as is shown in this figure.

EXTRA CREDIT: Character Analysis

In this book, as we showed in the word cloud, Scrooge and Bob are 2 leading characters. Now, I prefer to do the character analysis for them:

First, I am going to calculate the frequency for the characters.

The following table in the count number of times each character appears in each chapter:

Table 3: The Count for each Character

section	scrooge	bob
1	120	0
2	66	0
3	81	20
4	34	15
5	38	10

Table 4: The Count when Both Characters appear

section	paragraph	both_appear
3	41	1
3	78	1
3	80	1
4	125	1
5	29	1
5	69	1
5	71	1

Now we can find some information about these 2 characters. For Scrooge, he is the most leading character for this book so we can see him in every chapter. However, for Bob, he only exist in Chapter 3 to 5 and he has some interactions with Scrooge, which is reasonable because Scrooge is his boss.

Reference

1. A Christmas Carol in Prose; Being a Ghost Story of Christmas by Charles Dickens
2. Software Repository for Account and Finance
3. Text Mining with R
4. The ideas and supports from my dear classmate Yuli Jin.