task3_truenumber

zby

12/7/2021

MA615 Assignment4 Text Analysis_Task3

Introduction of The Cash Boy

The book I choose is The Cash Boy,a juvenile fiction. Here is the link https://www.gutenberg.org/ebooks/296. I used the gutenbergr package which provides access to the public domain works from Project Gutenberg collection to download this book.

The story is about a little poor boy, which lost everyone he loved except his little sister. Because the situation with no parents and income, the boy moved to New York for himself, so he could make some money for him and his sister.

Installing the R tnum package

```
##
##      checking for file '/private/var/folders/jt/pwyby9051bn5lfvmx1kkptm00000gn/T/RtmpvuLpwt/remotes4
## -  preparing 'tnum':
##      checking DESCRIPTION meta-information ...  v  checking DESCRIPTION meta-information
## -  checking for LF line-endings in source and make files and shell scripts
## -  checking for empty or unneeded directories
##      Omitted 'LazyData' from DESCRIPTION
## -  building 'tnum_0.3.3.tar.gz'
##
##
```

Load the tnum package and authorize the mssp1 server

I download The Cash Boy and put it into tnum.

```
##  [1] ""                       "wells/hw_time"
##  [3] "wells/hw_time_1"        "Twain/Tom_Sawyer"
##  [5] "twain/tom_sawyer"       "cervantes/don_quixote"
##  [7] "Dickens/david_copperfield" "dickens/david_copperfield"
##  [9] "well/hw_time_1"         "hw_time_1/heading"
## [11] "hw_time_1/section"      "hw_time_2/heading"
## [13] "hw_time_2/section"      "Alcott/Little_women"
## [15] "alcott/little_women"    "Kafka/Metamorphosis"
## [17] "kafka/metamorphosis"    "bronte/wuthering_heights"
## [19] "twain/wz_huck_1"        "Thomas"
## [21] "thomas"                 "Thomas/Tess"
## [23] "thomas/tess"            "wells/hw5"
## [25] "wells5/hw5"             "Kipling/junglebook"
```

```
## [27] "kipling/junglebook"        "cervantes/don_quixote_v2"
## [29] "jane_eyre/heading"         "jane_eyre/section"
## [31] "wells6/hw6"                "wells7/hw7"
## [33] "wells8/hw8"                "wells9/hw9"
## [35] "Twain/Tom_Sawyer_TEST"     "twain/tom_sawyer_test"
## [37] "barrie/Assignment4"        "barrie/assignment4"
## [39] "wilde/section"             "gatsby/hw_1"
## [41] "zweig/test1"               "tnum_pp/assignment4"
## [43] "randb/heading"             "randb/section"
## [45] "cervantes/don_quixote_v3"  "cervantes/don_quixote_v4"
## [47] "Zweig/test1"               "Twain/Tom_Sawyer2"
## [49] "twain/tom_sawyer2"         "Zweig/test2"
## [51] "zweig/test2"               "glasser/the_martian"
## [53] "wells10/hw10"              "wells11/hw11"
## [55] "wells12/hw12"              "hssb/book"
## [57] "hssb/book1"                "glasser/the_martian_1"
## [59] "Glasser/the_martian_1"     "jv_ttluts/textanalysis"
## [61] "Austen/persuation"         "austen/persuation"
## [63] "hvsb/book3"
```

use query to check INs
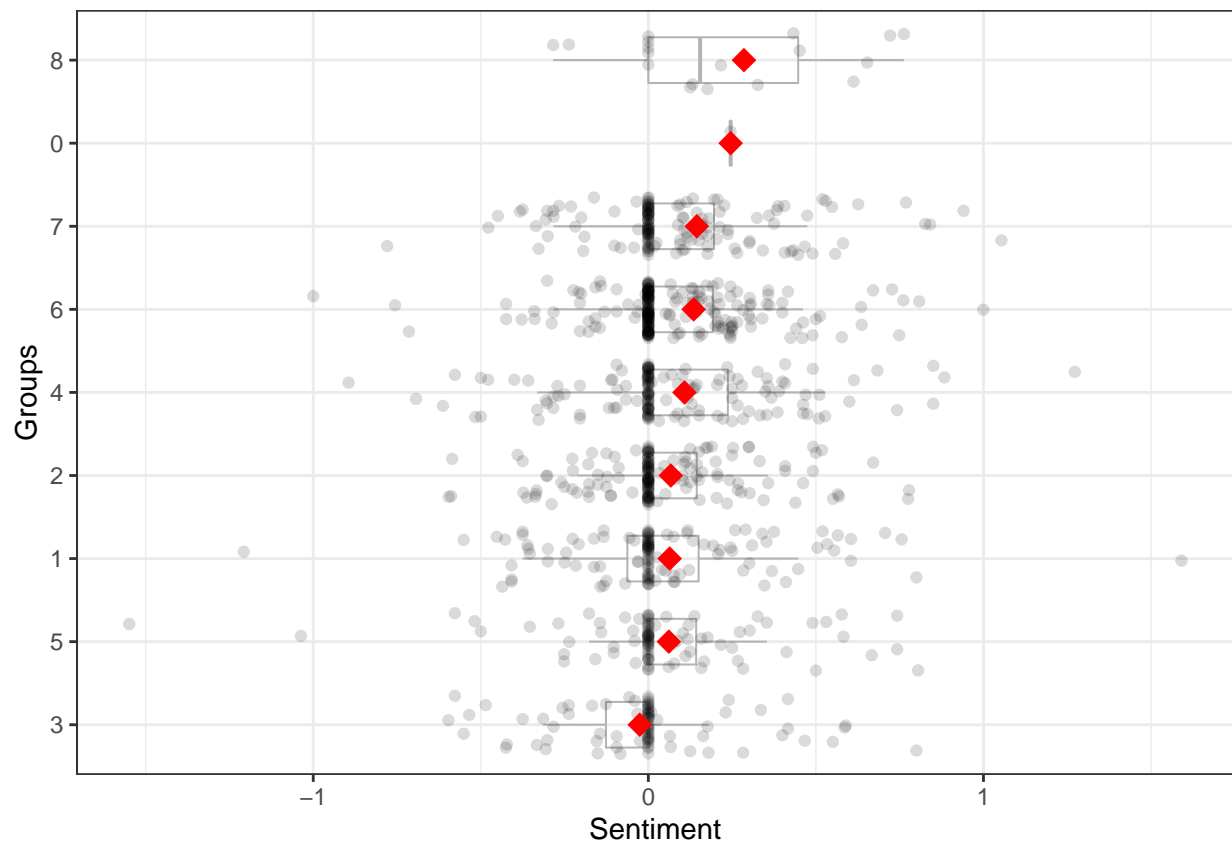
I show the tnum database of The Cash Boy.

subject	property	string.value	numeric.value
thecashboy1/zzzby/section:0000/paragraph:0001/sentence:0001count:word	NA		8
thecashboy1/zzzby/section:0001/paragraph:0003/sentence:0001count:word	NA		26
thecashboy1/zzzby/section:0001/paragraph:0004/sentence:0001count:word	NA		44
thecashboy1/zzzby/section:0001/paragraph:0004/sentence:0002count:word	NA		38
thecashboy1/zzzby/section:0001/paragraph:0004/sentence:0003count:word	NA		15
thecashboy1/zzzby/section:0001/paragraph:0005/sentence:0001count:word	NA		11

I use query section and heading text to show the head

```
##                                                    subject   property
## 1 thecashboy1/zzzby/section:0000/paragraph:0001/sentence:0001 count:word
## 2 thecashboy1/zzzby/section:0001/paragraph:0003/sentence:0001 count:word
## 3 thecashboy1/zzzby/section:0001/paragraph:0004/sentence:0001 count:word
## 4 thecashboy1/zzzby/section:0001/paragraph:0004/sentence:0002 count:word
## 5 thecashboy1/zzzby/section:0001/paragraph:0004/sentence:0003 count:word
## 6 thecashboy1/zzzby/section:0001/paragraph:0005/sentence:0001 count:word
##   string.value
## 1         <NA>
## 2         <NA>
## 3         <NA>
## 4         <NA>
## 5         <NA>
## 6         <NA>
```
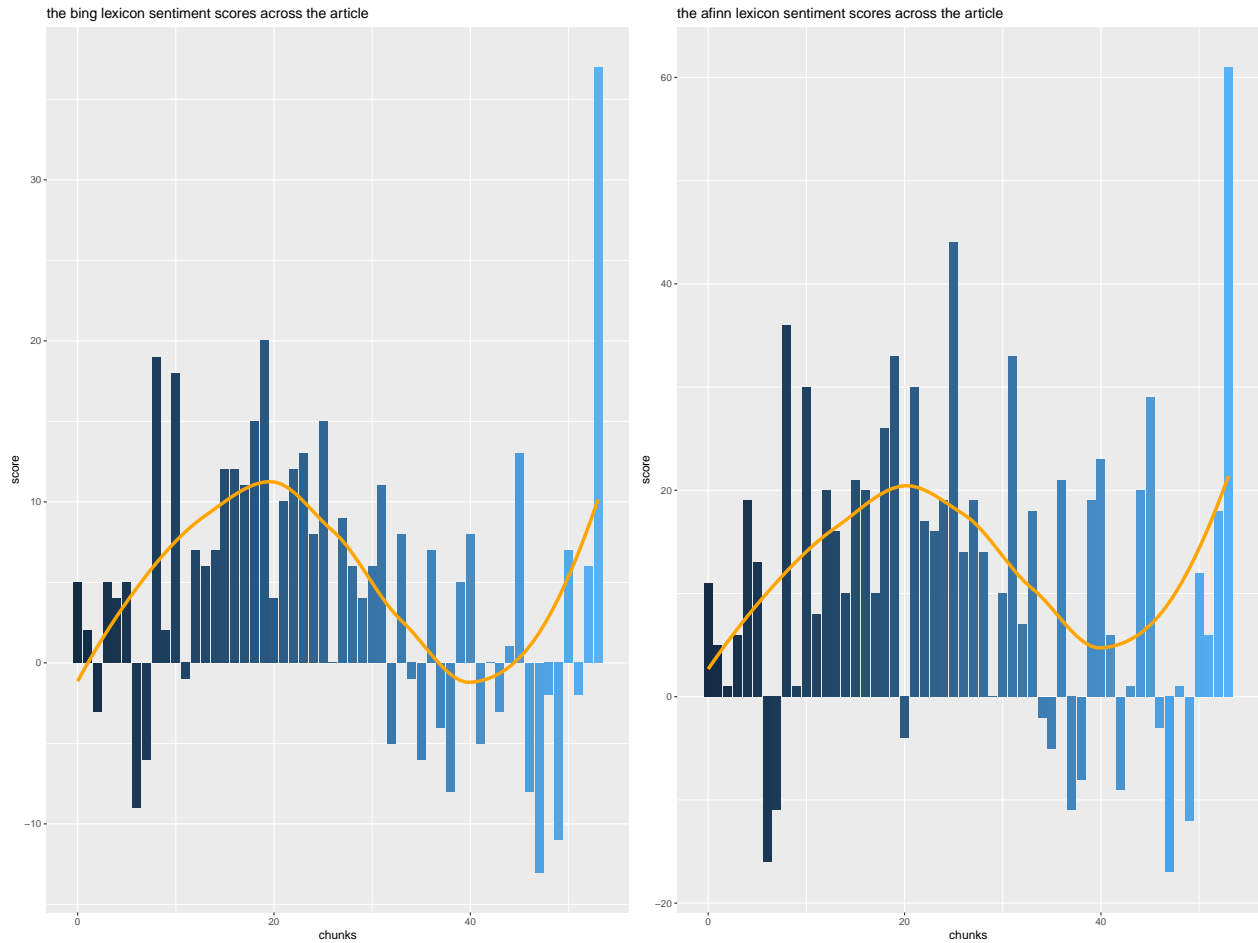
Sentiment Annalysis

I used sentimentr package to do sentiment analysis and get sentiment score.

Comparative Analysis

I used four lexicons to analyze The Cash Boy. Now I have the sentence-level scores, I want to do comparative analysis.

the bing lexicon sentiment scores across the article

the afinn lexicon sentiment scores across the article

According to the above plots, the trend of sentence-level scores and previous sentiment scores are the same. At the begin, Frank lost parents, so he experienced sadness. Sentiment Scores change from positive to negative twice. After Frank arrived New York, he makes some money so that he can support himself and little sister. There are joys and sorrows in New York, so there are positive and negative scores. At the end of sotry, Frank walked out of sorrow, sentiment scores are positive.

Extra Credit

Frank is the main character in The Cash Boy. Firstly, I find Frank. Secondly, I calculated the times Frank appeared in each chapter. Finally, I used table to show the section Frank appeared.

section	Frank
section:0000	0
section:0001	20
section:0002	13
section:0003	11
section:0004	18
section:0005	13
section:0006	24
section:0007	23
section:0008	2

Reference

Julia Silge & David Robinson. (2016). Welcome to Text Mining with R [online]. Available from: https://www.tidytextmining.com/index.html [accessed 9 December 2021].