

Assignment_4_task3

Guangze Yu

12/7/2021

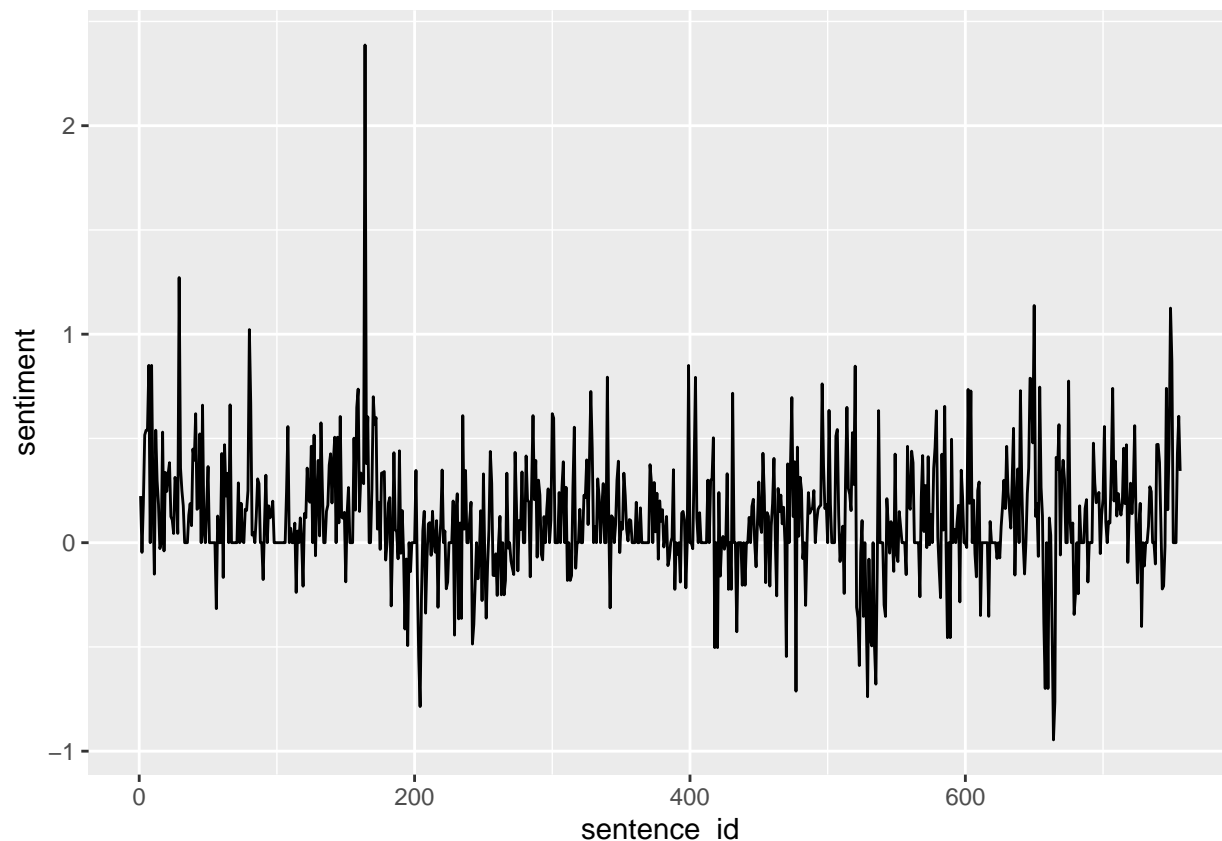
TASK three:sentence-level analysis

```
library(tnum)
tnum.authorize("mssp1.bu.edu")
tnum.setSpace("test2")
source("Book2TN-v6A-1.R")
Achievement <- gutenbergs_download(gutenberg_id = 38783)
# write.table(Achievement, "Achievement.txt", row.names = FALSE, col.names = FALSE)
Achievement <- read.table("Achievement.txt", header = T)
# tnBooksFromLines(Achievement$text, "Benjamin/Achievement_test3")
# tnum.getDBPathList(taxonomy = "subject", levels = 2, max=1000)

q4 <- tnum.query("benjamin/achievement_test3# has text", max=1000)
qdf4 <- tnum.objectsToDf((q4))
para_text4 <- qdf4 %>% pull(string.value) %>%
  str_replace_all("\\\"", "\"") %>%
  str_flatten(collapse = " ")
sentences_ach <- get_sentences(para_text4)
```

Below is the sentiment scores based on sentimentr package.

```
sentiment(sentences_ach) %>%
  ggplot(aes(sentence_id, sentiment)) +geom_line()
```



Below is the sentences level analysis based on TASK 2: the bag of word analysis. We can see that the sentiment scores are not the same for each lexicon. The trend is approximately the same for different sentences analysis.

```
text_df_task_3 <- qdf4%>%
  mutate(string.value = as.character(string.value)) %>%
  mutate(linenumber = row_number()) %>%
  unnest_tokens(word, string.value)

afinn <- text_df_task_3 %>%
  inner_join(get_sentiments("afinn")) %>%
  group_by(index = linenumber) %>%
  summarise(sentiment = sum(value)) %>%
  mutate(method = "AFINN")

bing_and_nrc_lou <- bind_rows(
  text_df_task_3 %>%
    inner_join(get_sentiments("bing")) %>%
    mutate(method = "Bing et al."),
  text_df_task_3 %>%
    inner_join(get_sentiments("nrc")) %>%
    filter(sentiment %in% c("positive",
                          "negative"))
) %>%
  mutate(method = "NRC"),
text_df_task_3 %>%
  inner_join(get_sentiments("loughran")) %>%
```

```

        filter(sentiment %in% c("positive",
                                "negative"))

    ) %>%
    mutate(method = "loughran") %>%
    count(method, index = linenumber, sentiment) %>%
    pivot_wider(names_from = sentiment,
                values_from = n,
                values_fill = 0) %>%
    mutate(sentiment = positive - negative)

bind_rows(afinn,
          bing_and_nrc_lou) %>%
ggplot(aes(index, sentiment, fill = method)) +
geom_col(show.legend = FALSE) +
facet_wrap(~method, ncol = 1, scales = "free_y")

```

