Algorithm Aversion

Kopkow, Angermaier, Rohrer, Sonnleitner

February 2022

Contents

Motivation: Erwin 1

Data retrieval 1

Data processing: Mathias 1

Analysis 3

- Progression over the years . . . 4
- Distribution of classes by topic . . . 5

Conclusion: 6

Critique: Alina 6

Github Repository 6

- R Markdown . . . 6
- Including Plots . . . 6

Header 1 6

- Header 2 . . . 6

Motivation: Erwin

Data retrieval

Data processing: Mathias

```
library(tidyverse)
library(tidytext)
library(textdata)
library(dplyr)
library(vader)
library(academictwitteR)
library(data.table)
library(readr)
library(boot)
library(knitr)
```

used libraries

```
## `summarise()` has grouped output by 'alg.data$Year'. You can override using the `.groups` argument.

## `summarise()` has grouped output by 'Year'. You can override using the `.groups` argument.
```

```
listofdfs <- list()

#data group for cummulated dataframe

for (i in c(7:13)){
  alg.data %>%
    filter(alg.data[,i]>=1)%>%
    mutate(Topic = colnames(alg.data)[i], VADERclass=as.factor(VADERclass))%>%
    group_by(Year,Topic,VADERclass, .drop=FALSE)%>%
    summarise(Sent = n(), .groups = "drop") ->listofdfs[[i]]
}
summ.cumm <- bind_rows(listofdfs)


#adds percentage to cummulated DF
summ.cumm <-
  summ.cumm %>%
  group_by(Topic, Year) %>%
  mutate(All = sum(Sent),percent=(100*Sent/All))

summ.cumm
```

```
## # A tibble: 324 x 6
## # Groups:   Topic, Year [84]
##     Year Topic    VADERclass  Sent   All percent
##    <dbl> <chr>    <fct>      <int> <int>   <dbl>
##  1  2010 Business Aversive       2   297   0.673
##  2  2010 Business Negative      14   297   4.71
##  3  2010 Business Neutral      226   297  76.1
##  4  2010 Business Positive      55   297  18.5
##  5  2011 Business Aversive       7   337   2.08
##  6  2011 Business Negative      17   337   5.04
##  7  2011 Business Neutral      217   337  64.4
##  8  2011 Business Positive      96   337  28.5
##  9  2012 Business Aversive       4   357   1.12
```

```
## 10  2012 Business Negative      36   357  10.1
## # ... with 314 more rows
```

```r
#data group for topic oriented dataframe

listofdfs <- list()

for (i in c(7:13)){
  alg.data %>%
    filter(alg.data[,i]>=1)%>%
    mutate(Topic = colnames(alg.data)[i])%>%
    group_by( VADERclass, Topic) %>%
    summarise(Sent = n()) ->listofdfs[[i]]
}
summ.topic <- bind_rows(listofdfs)

#adds percentage to topicwise DF
summ.topic<-
  summ.topic %>%
  group_by(Topic) %>%
  mutate(All = sum(Sent),percent=(100*Sent/All))
```

```r
kable(summ.topic[1:5, ], caption = "Topic Table")
```

Table 1: Topic Table

VADERclass	Topic	Sent	All	percent
Aversive	Business	73	3928	1.858452
Negative	Business	373	3928	9.495927
Neutral	Business	2165	3928	55.117108
Positive	Business	1317	3928	33.528513
Aversive	Social.Media	1601	43621	3.670251

```r
#--------------Test wordgroups---------------------
listofdfs <- list()


for (i in c(7:13)){
  alg.data %>%
    filter(alg.data[,i]>=1)%>%
    mutate(Topic = colnames(alg.data)[i])%>%
    sample_n(size = 20)%>%
  group_by( VADERclass, Topic)  ->listofdfs[[i]]
}
Wordgrouptest<- bind_rows(listofdfs)
```

Analysis

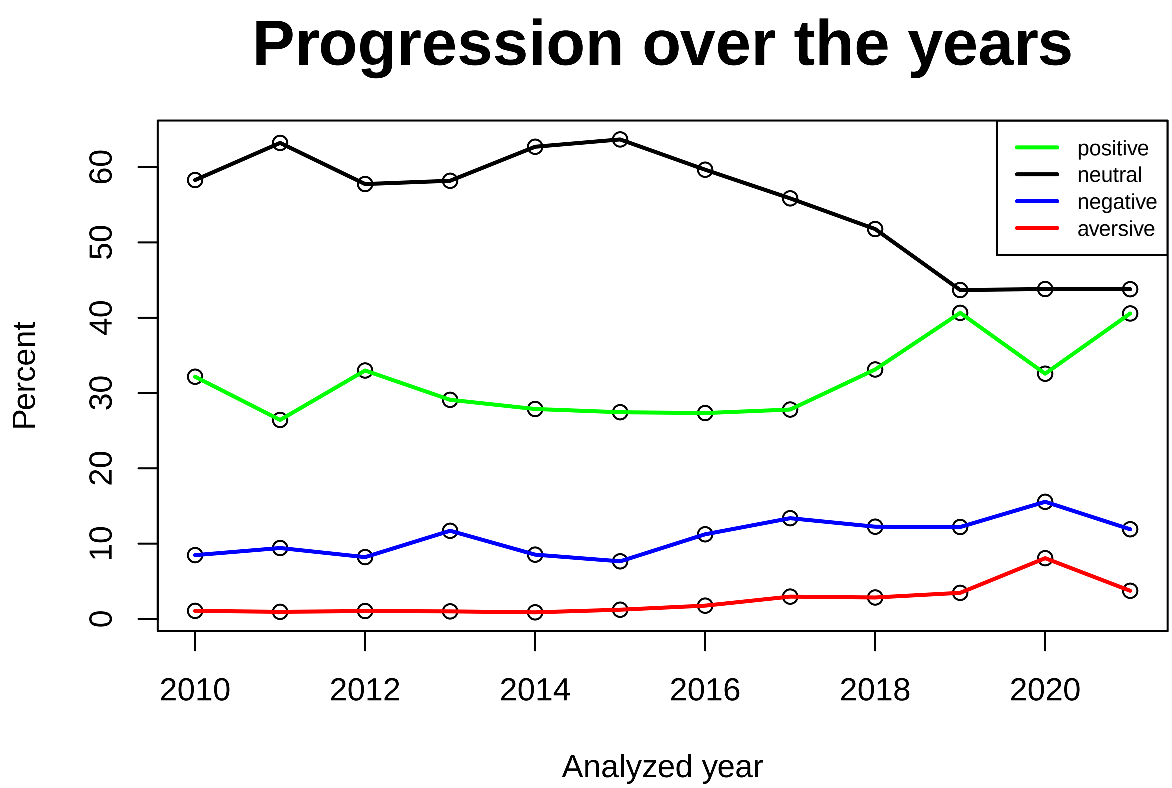
The following chapter covers the results of the sentiment analysis of tweets over the years 2010 to 2021 , for which we used the Vader tool. Values between -0.3 and 0.3 were classified as neutral. Those between 0.3

and 1 were classified as positive, as negative between -0.3 and -0.7 and finally as aversive values between -0.7 and -1.

A Confidence interval based on 10000 bootstrap replications was calculated. It revealed that with a probability of 95% that the mean value of the sentiment analysis falls between 0.0766 and 0.1349.

Progression over the years

The diagram shows the percentage progression of the results of the sentiment analysis of tweets over the years 2010 to 2021.A total of 143271 tweets were analyzed. Since all tweets that could not be evaluated in the sentiment analysis were classified as neutral, the highest number of tweets was yielded in the neutral category while the fewest tweets were classified as aversive.



Over time, the number of neutral tweets decreases. In 2010, there were 6934 neutral tweets which amounts to 58.3% of the yearly total. In 2015, the highest number of neutral tweets were posted, with an absolute of 7641 tweets posted (63.8% of the yearly total). In 2019, the lowest number of neutral tweets (5241, 43.7%) was posted. 2021 yielded similar numbers (5254, 43.8%). Over the last 3 years, the number of neutral tweets remained at the same level.

An increasing trend can be seen in the number of positive tweets. In 2010, there were 3827 positive tweets, representing 32.2% of the yearly total. In 2011, the least amount of tweets 3171 (26.7%) were posted. In 2019, the most positive tweets were posted with absolute 4878 (40.7%) posted. Last year, 2021, a total of 4868 positive tweets were posted, which is 40.6% of the annual total. The number of positive tweets remained high over he last 3 years, with a dip in 2020 (3908 total, 32.6%).

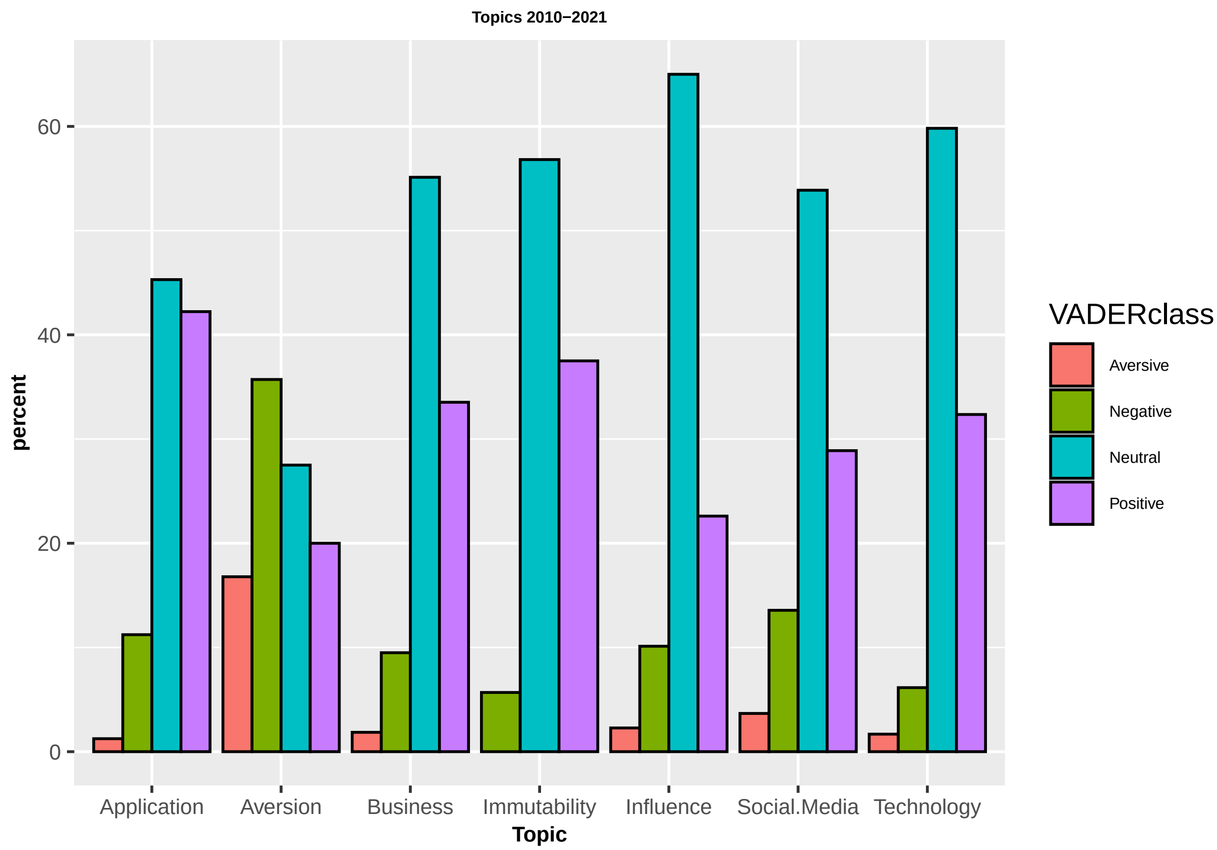
Negative tweets show an increasing trend. In 2010, there were 1008 negative tweets constituting 8.5% of the annual total. In 2020 the highest number of negative tweets were posted with absolute 1867 tweets (15.6%)

posted. In 2015, the least amount of negative tweets was posted (918, 7.7%). In the last year analyzed, 2021, the number of negative tweets remained high with 1429 total, 11.9% of the annual.

The number of aversive tweets behaves similarly to that of negative tweets. 2011 113 0.95% there were the fewest were 2014 105 and 0.88% with the most there were 2020 968 8.1% and 2021 there were 3.7% and 449 tweets.

Distribution of classes by topic

This graph shows the different topics Business, Social Media, Technology, Immutability, Influence, Application and Aversion as well as the percentage distribution of the VADER classes Aversive, Negative, Neutral, and Positive.



The topic Aversion has the most aversive tweets with 16.8% and 35% negative tweets, so there are more negative tweets in this topic than positive 20%. The smallest amount occurs in Application; here, there are no aversive tweets at all. In the topic Social media, 28.9% of the tweets are positive, 13.6% negative and 3.7% aversive. In total, this was also the topic in which the most tweets occurred with a total of 42020.

Conclusion:

Critique: Alina

Github Repository

https://github.com/sonnleit/AlgorithmAversion

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
#summary(cars)
```

Including Plots

You can also embed plots, for example:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Header 1

Header 2

Header 3

Header 4

Header 5

Header 6