Algorithm Aversion

Kopkow, Angermaier, Rohrer, Sonnleitner

February 2022

Contents

Motivation

Data retrieval

Data retrieval and analysis was performed entirely in R using the respective framework RStudio. To obtain a sufficient amount of data over a period of 11 years, the "academictwitteR" package was used. Its function get_all_tweets() allows not only a search for Tweets containing a specific word, but also to set up a timespan, from where Tweets should be obtained. In order to reduce seasonal effects, 1000 tweets were retrieved from each month on a random day from 2010 to 2021 included. The resulting dataset thus includes 143,721 tweets (on some days, not exactly 1000 tweets were found).

As text analysis tool, we decided to use the unsupervised "VADER" method, as it turned out in the lecture and one according exercise, that it has advantage in precision and reliability compared to other tools like "Syuzhet". With its capabilities, we analyzed a sentiment value for every tweet, ranging from -1 to 1.

A further important resource is the "tidytext" library, especially the unnest_tokens() function in order to categorize tweets into various topics. After searching for words on the Merriam-Webster thesaurus, that do indicate the following topics: Business, Technology, Social Media, Immutability, Changeability, Application, Aversion. Those word patterns where used to identify a tweet. To do this, the text was disassembled,

stopwords were erased in order to maintain better efficiency, and the remaining text was compared to the before initialised word patterns.

Data processing

The value results from the VADER analysis was categorized into four sentiments. A value higher than 0.3 indicates a positive sentiment, between 0.3 and -0.3 a neutral sentiment, between -0.3 and -0.7 a negative sentiment and lower than 0.7 was considered an aversive sentiment. The treshold for the aversive category was set through trying various values, reading the corresponding tweets to find out, what threshold offers an accurate indicator for aversiveness. The aformentioned word pattern comparison uses the number of indicator words in a tweet and saves that value in the respective topic column. With that information, we can than link the tweet to one specific topic. Thanks to this procedure, we are able to group the tweets into the existing topics, which is crucial for research question 2. Research question 1, the development over the years, could be answered by getting the year of a tweet's timestamp, save it in a corresponding column and then group by those years.

The topic categorization needed to be further validated, as our group wanted to make sure, that tweets were ordered accordingly and the significance of such a content check is assured. In order to do so, 30 random tweets of every topic were read and than decided, whether they truly fit the topic. Especially the topics Business, Technology and Social Media were accurately assigned in all of the 30 tweets. The topics Immutability, Influence and Aversion had some ambiguities in a few tweets, but overall still mostly were realted to the topic. Only the topic aversion showed inconsitencies and expressed a different meaning as we intended it to be.

With all that information we have every tweet's topic, sentiment, and year. This enables according filtering and grouping. As mentioned before, research question 1 is analyzed via grouping the years for the overall development. We then calculated the ratio of the four sentiments for every year and created a plot that shows the development via a line plot. To answer research question 2, grouping the topics was necessary. By calculating the percentage of every sentiment for every topic we could create a bar chart to indicate every topic's sentiment ratio. A further interesting look was on the topical sentiment ratio over time. This can be seen as a combination of both research questions. We created a line plot over time but for every topic separately. This enables deeper insight if algorithm aversion developed differently for specific topics. As we had to limit this report, the results can be found in our Git repository.

```
#--------------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.

Progression over the years
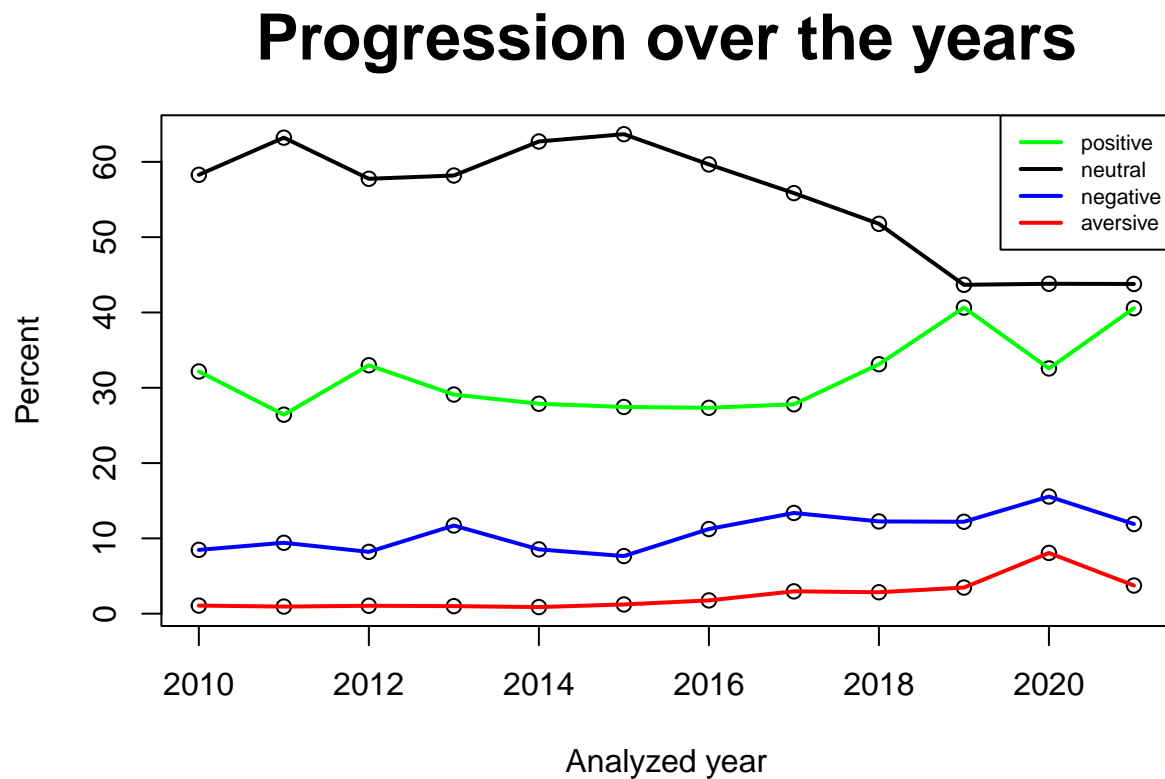
Over time, the number of neutral tweets decreases. In 2010, there were 6934 neutral tweets which amounts to 58.3% of the yearly total. In 2015, the highest number of neutral tweets were posted, with an absolute of 7641 tweets posted (63.8% of the yearly total). In 2019, the lowest number of neutral tweets (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.
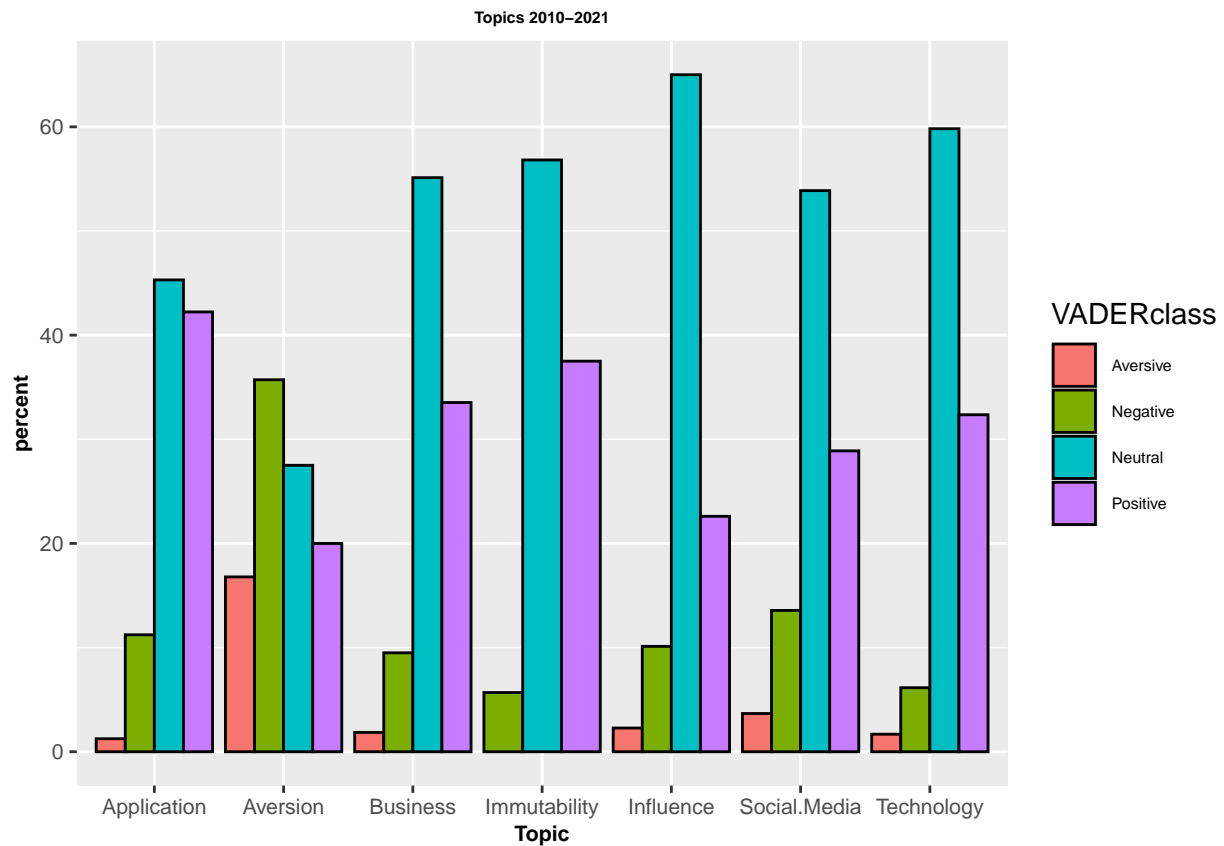
The topic Aversion has the most aversive tweets with 16.8% and 35% negative tweets, so there are more negative tweets in this topic than positive 20%. The smallest amount occurs in Application; here, there are no aversive tweets at all. In the topic Social media, 28.9% of the tweets are positive, 13.6% 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