

Sentiment Analysis

Max Callaghan

2022-11-10

Introduction and Objectives

Assignment 2

Assignment 2 is still live. If you have issues, or encounter difficulties, raise an issue on the Github repository, or write me an email!

Assignment 3

Assignment 3 is approaching, and you should have a clear idea of what you want to do by the end of next week.

Feel free to ask for quick feedback on any ideas you have in the coming days

Objectives

By now we have spent a long time understanding how to **represent** texts in simple and more complex ways.

We've also started asking questions about texts. Viz. What is it about?

Today we will ask a new question about texts: what sentiment does it express?

Intro to sentiment analysis

What is a sentiment?

The emotion embodied in a text. Often reduced to positive-negative, but can encompass a more complex range of emotions like joy, sadness, anger.

Sentiment analysis as classification

In some ways, sentiment analysis is a special case of the broader **classification** task, where we ask:

- Is a text of class A or not? or,
- Is a text of class A or B

With sentiment analysis, the classes are inherent features of how humans use language which *generalise* (to a certain extent) across contexts.

An overview of techniques to do sentiment analysis

Doing sentiment analysis usually involves rule-based or statistical techniques

- Assessing sentiment based on counting words have a predefined sentiment

An overview of techniques to do sentiment analysis

Doing sentiment analysis usually involves rule-based or statistical techniques

- Assessing sentiment based on counting words have a predefined sentiment
- Using a classifier that has been trained to identify sentiment with text examples that have been labelled.

Introduction and Objectives
○○○○

Intro to sentiment analysis
○○○○

Lexicon-based sentiment analysis
●○○○○○○○○○○○○○○

Fancy sentiment analysis
○○○

Validation
○○○

Examples in research
○○○○○○○

Wrapup and Outlook
○○○○

Lexicon-based sentiment analysis

Positive and negative words

We know about the “bag of words” model of representing texts.

We also know that some words are rather positive, whereas some are rather negative.

Consider the texts:

```
texts <- c(  
  "Elon Musk is a champion of free speech",  
  "It's a terrible shame to see mashed potato thrown at art"  
)
```

Do they express positive or negative sentiment? How can we tell?

Using Lexicons in R

We can import a lexicon in R using tidytext. Each row, contains a word and its value

```
library(tidytext)
library(dplyr)
lex <- get_sentiments("afinn")
sample_n(lex, 5)
```

```
## # A tibble: 5 x 2
##   word      value
##   <chr>    <dbl>
## 1 hooligan    -2
## 2 damn       -4
## 3 slicker     2
## 4 brainwashing -3
## 5 faggot     -3
```

Using Lexicons in R

Note that the Afinn lexicon is not the newest version. We can just read this in directly from the author's Github page.

```
library(readr)
lex <- read_tsv(
  "https://raw.githubusercontent.com/fnielsen/afinn/master/afinn/data/AFINN-en-165.txt",
  col_names=c("word", "value")
)

lex
```

```
## # A tibble: 3,382 x 2
```

```
##   word      value
```

```
##   <chr>    <dbl>
```

```
## 1 abandon      -2
```

```
## 2 abandoned    -2
```

```
## 3 abandons     -2
```

```
## 4 abducted     -2
```

```
## 5 abduction     2
```

Using Lexicons in R

There are a few different lexicons, compiled by different authors, using different techniques involving amazon turk and author knowledge, which encode different types of emotions.

```
library(tidytext)
library(dplyr)
lex <- get_sentiments("nrc")
head(lex)
```

```
## # A tibble: 6 x 2
##   word      sentiment
##   <chr>     <chr>
## 1 abacus    trust
## 2 abandon   fear
## 3 abandon   negative
## 4 abandon   sadness
## 5 abandoned anger
## 6 abandoned fear
```

Using Lexicons in R

We can also put our usual document feature matrix into a similar format

```
library(quantda)
dfmat <- texts %>%
  tokens %>%
  dfm()

text_tokens <- tidy(dfmat)
head(text_tokens)
```

```
## # A tibble: 6 x 3
##   document term      count
##   <chr>    <chr>    <dbl>
## 1 text1    elon          1
## 2 text1    musk          1
## 3 text1    is            1
## 4 text1    a             1
## 5 text2    a             1
## 6 text1    champion      1
```


Tidy lexicons

Now we can join these to see which words in the texts have what sentiment

```
lex <- read_tsv("https://raw.githubusercontent.com/fnielsen/afinn/master/afinn/data/AFINN-en-165.txt", col_names=c("term", "value"))
dfmat <- texts %>%
  tokens %>%
  dfm()

text_tokens <- tidy(dfmat) %>%
  inner_join(lex, by=c("term" = "word"))

text_tokens
```

```
## # A tibble: 4 x 4
##   document term      count value
##   <chr>    <chr>    <dbl> <dbl>
## 1 text1    champion     1      2
## 2 text1    free        1      1
## 3 text2    terrible    1     -3
## 4 text2    shame       1     -2
```

Tidy lexicons

We can then just sum word scores for each document to get a sentiment score for that document

```
doc_sentiments <- tidy(dfmat) %>%
  inner_join(lex, by=c("term" = "word")) %>%
  mutate(value=value*count) %>%
  group_by(document) %>%
  summarise(value = sum(value))
```

```
doc_sentiments
```

```
## # A tibble: 2 x 2
##   document value
##   <chr>     <dbl>
## 1 text1         3
## 2 text2        -5
```

VADER

VADER represents just about the state of the art in lexicon-based sentiment analysis, and is especially suitable for social media texts.

It also incorporates rules that extend it beyond the bag-of-words model

5 Heuristics

The Vader [paper](#) identifies 5 heuristics that extend just counting words from a lexicon, and implements these in their algorithm.

- Punctuation (!) increases the magnitude of the sentiment: "Food here is good!!" > "Food here is good"

5 Heuristics

The Vader [paper](#) identifies 5 heuristics that extend just counting words from a lexicon, and implements these in their algorithm.

- Punctuation (!) increases the magnitude of the sentiment: "Food here is good!!" > "Food here is good"
- CAPITALIZATION increaeses the magnitude of the sentiment: "Food here is GREAT" > "Food here is great"

5 Heuristics

The Vader [paper](#) identifies 5 heuristics that extend just counting words from a lexicon, and implements these in their algorithm.

- Punctuation (!) increases the magnitude of the sentiment: "Food here is good!!" > "Food here is good"
- CAPITALIZATION increaeses the magnitude of the sentiment: "Food here is GREAT" > "Food here is great"
- Degree modifiers impact intensity > or <. "Service is marginally good" < "service is good" < "service is extremely good".

5 Heuristics

The Vader [paper](#) identifies 5 heuristics that extend just counting words from a lexicon, and implements these in their algorithm.

- Punctuation (!) increases the magnitude of the sentiment: "Food here is good!!" > "Food here is good"
- CAPITALIZATION increaeses the magnitude of the sentiment: "Food here is GREAT" > "Food here is great"
- Degree modifiers impact intensity > or <. "Service is marginally good" < "service is good" < "service is extremely good".
- "But" signals shift in sentiment, and that second clause is stronger: "Food here is good, but the service is bad" -> Overall more negative than positive

5 Heuristics

The Vader [paper](#) identifies 5 heuristics that extend just counting words from a lexicon, and implements these in their algorithm.

- Punctuation (!) increases the magnitude of the sentiment: "Food here is good!!" > "Food here is good"
- CAPITALIZATION increaeses the magnitude of the sentiment: "Food here is GREAT" > "Food here is great"
- Degree modifiers impact intensity > or <. "Service is marginally good" < "service is good" < "service is extremely good".
- "But" signals shift in sentiment, and that second clause is stronger: "Food here is good, but the service is bad" -> Overall more negative than positive
- Negations in a tri-gram preceeding a sentiment-laden feature flip the polarity

VADER in practice

Let's load a [dataset](#) of tweets from the VoteYes campaign from the Scottish independence referendum. We can calculate sentiment for each tweet using `vader_df()`.

Let's look at the most positive tweets

```
library(vader)
yes <- read_csv("../datasets/YesScotlandTweets_cleaned.csv")
yes$campaign <- "YesScotland"
no <- read_csv("../datasets/UkTogetherTweets_cleaned.csv")
no$campaign <- "UkTogether"
tweets <- rbind(yes, no)

sentiments <- vader_df(tweets$text)
tweet_sentiment <- cbind(tweets, select(sentiments, -text))

pos <- tweet_sentiment %>% arrange(desc(compound)) %>%
  head()

for( i in rownames(pos) ) {
  print(pos[i, "text"])
  print(pos[i, "compound"])
}
```

```
## [1] "RT @dobbs_michael: Off to Edinburgh. Yesterday great campaigning in W Scotland, smiles, great support. Hope
```

```
## [1] 0.965
```

```
## [1] "A Yes means greater financial security for families - we can expand free childcare, safeguard free education
```

VADER in practice

Let's load a [dataset](#) of tweets from the VoteYes and UkTogether campaigns from the Scottish independence referendum. We can calculate sentiment for each tweet using `vader_df()`.

Let's look at the most negative tweets

```
neg <- tweet_sentiment %>% arrange(compound) %>%
  head()
```

```
for( i in rownames(neg) ) {
  print(neg[i, "text"])
  print(neg[i, "compound"])
}
```

```
## [1] "A statement: There is ABSOLUTELY no place for attacks - be they abuse, graffiti, vandalism or physical assa
## [1] -0.934
## [1] "Westminster wants to waste our resources on renewing obscene and dangerous weapons of mass destruction. Sco
## [1] -0.925
## [1] "RT @AlexSalmond: The murder of David Haines shows a degree of brutality which defies description. Thoughts
## [1] -0.866
## [1] "Damaging Westminster cuts are threatening Scotland\u008a\u0097Ès public services. #indyref #VoteYes http://
## [1] -0.836
## [1] "RT @martin_compston: More ridiculous scare stories this regarding losing BBC shows I'm in a hotel in Ireland
## [1] -0.835
## [1] "RT @StephenNoon: Hearing that a truly desperate & shameful scare story is coming @DHgovuk about to thre
## [1] -0.813
```

Sentiment over time

We can also look at how sentiment changed over time by taking the mean compound score in each time period, for each campaign group. Given the regular week-weekend variation, it also makes sense to show the 7 day rolling mean.

```
library(tidyr)
tweet_sentiment$date <- as.Date(tweet_sentiment$created)
wide_sentiment <- tweet_sentiment %>%
  group_by(campaign, date) %>%
  summarise(score = mean(compound)) %>%
  pivot_wider(names_from=campaign, values_from=score)

days <- data.frame(date=seq(as.Date("2014-06-01"),as.Date("2014-09-18"),1))

daily_sentiment <- days %>% left_join(wide_sentiment) %>%
  pivot_longer(cols=-date, names_to="campaign", values_to="score") %>%
  group_by(campaign) %>% arrange(date) %>%
  mutate(score7 = data.table::frollmean(score, 7))

daily_sentiment %>% head()
```

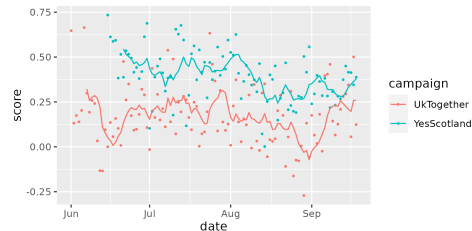
```
## # A tibble: 6 x 4
## # Groups:   campaign [2]
##   date      campaign    score score7
##   <date>    <chr>      <dbl> <dbl>
## 1 2014-06-01 UkTogether  0.647    NA
```

Sentiment over time

Now that we have the data in the format we want, we can plug this into ggplot2

```
library(ggplot2)
ggplot(daily_sentiment, aes(date, colour=campaign)) +
  geom_point(aes(y=score), size=0.5) +
  geom_line(aes(y=score7))

ggsave("plots/sentiment_time.png", width=6, height=3)
```



Introduction and Objectives
○○○○

Intro to sentiment analysis
○○○○

Lexicon-based sentiment analysis
○○○○○○○○○○○○○○○○

Fancy sentiment analysis
●○○

Validation
○○○

Examples in research
○○○○○○○

Wrapup and Outlook
○○○○

Fancy sentiment analysis

Fancy sentiment analysis

Fancy NLP does not apply rules that we give it. It *learns* rules from training data.

Complex models, which encode text in complex ways, have outperformed lexicon-based sentiment analysis *on the main benchmarked tasks for which they are often optimized*.

Sentiment datasets are often comprised of movie or product reviews.

Fancy sentiment analysis

We will learn more about how training such models work in the next sessions, but you can access one of many such models [here](#)

Introduction and Objectives
○○○○

Intro to sentiment analysis
○○○○

Lexicon-based sentiment analysis
○○○○○○○○○○○○○○○○

Fancy sentiment analysis
○○○

Validation
●○○

Examples in research
○○○○○○○

Wrapup and Outlook
○○○○

Validation

Validation

Almost all methods for sentiment analysis are validated, but almost none are validated on your dataset. Unless your dataset is very similar to the validation dataset, you should validate yourself.

This means selecting a random sample of your texts, labelling the sentiment of these texts by hand, then comparing the label you gave with the score given by your method.

If your method gives the same label as you in 100% of cases, then you have an accuracy of 100%.

We will explore validation metrics in the next session on supervised learning.

Exercise - validating on your own data

- In a spreadsheet, write 5 texts that express your feelings about recent political events, with each text in a new row of the same column
- Ask your neighbour to rate the sentiment of your texts in the neighbouring column. The value should be between -1 (completely negative) and 1 (completely positive).
- Apply one of the techniques we have learnt today to your texts and compare with the human annotation
- What is the average (absolute) difference between your human label and the automated technique?
- For what text is the divergence greatest?

Introduction and Objectives
○○○○

Intro to sentiment analysis
○○○○

Lexicon-based sentiment analysis
○○○○○○○○○○○○○○○○

Fancy sentiment analysis
○○○

Validation
○○○

Examples in research
●○○○○○○

Wrapup and Outlook
○○○○

Examples in research

Tones from a Narrowing Race

In [Tones from a Narrowing Race: Polling and Online Political Communication during the 2014 Scottish Referendum Campaign](#), Evelyne Brie and Yannick Dufresne set out to investigate the dynamics of how political campaigns use negative communication.

They use the data that we've just seen: tweets sent by the pro-independence and pro-union campaigns in the Scottish independence referendum of 2014.

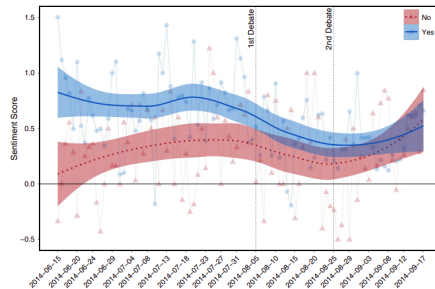
Tones from a Narrowing Race: Obectives

The authors want to provide evidence on negative campaigning *in practice* that complements political theory about how and why actors use negative campaigning.

They are interested in how real organisations use twitter, meaning their findings are not subject to the same external validity concerns generated by the unrepresentativeness of twitter.

Tones from a Narrowing Race: Results

The authors' analysis, which uses the Opinion Observer lexicon (I think available with the “bing” option in tidytext's `get_sentiments()`) finds a similar pattern of sentiment scores in the different campaigns over time



Tones from a Narrowing Race: Results

They find a significant relationship between the no campaign leading in the polls and the negativity of their tweets. The more NO leads, the more negatively they tweet.

In a simpler model, there is a significant relationship between the number of days to the election and the negativity of the YES campaign. That is, the decline in positivity as the election is statistically significant.

There is also evidence that the YES campaign tweeted more negatively the day after each debate (which they were said to have lost)

TABLE 1 *Testing the effect of public opinion and timing on sentiment score*

	Sentiment Score (Daily Average)				
	YES Campaign		NO Campaign		
Poll Lead	0.003 (0.01)		-0.01 (0.01)	-0.04* (0.01)	-0.05** (0.02)
Momentum		-0.001 (0.01)	-0.001 (0.01)	-0.01 (0.01)	0.01 (0.01)
Post-debate 1			-0.31* (0.15)		0.08 (0.19)
Post-debate 2			-0.38* (0.18)		-0.20 (0.24)
Poll Day	0.14 (0.07)	0.15* (0.07)	0.15* (0.07)	0.03 (0.09)	0.09 (0.10)
Sentiment Score _{t-1}	0.20* (0.10)	0.20* (0.10)	0.13 (0.10)	0.005 (0.10)	0.08 (0.10)
Days to Referendum	0.003** (0.001)	0.003* (0.001)	-0.002 (0.003)	0.0002 (0.002)	0.001 (0.002)
Constant	0.27*** (0.08)	0.27** (0.08)	0.74** (0.22)	0.29* (0.10)	0.23* (0.10)
N	94	94	94	94	94
R ²	0.18	0.18	0.23	0.11	0.03

*p < 0.05; **p < 0.01; ***p < 0.001.

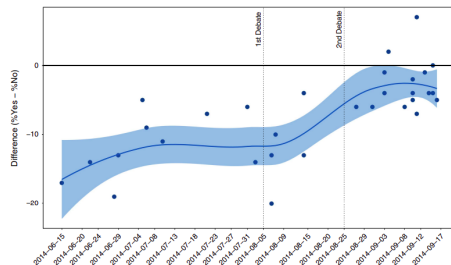
Source: Official Twitter accounts of *Better Together* (N = 1,230) and of *Yes Scotland* (N = 3,078), and *What Scotland Thinks* (16 June to 17 September 2014).

Tones from a Narrowing Race: Results

They find a significant relationship between the no campaign leading in the polls and the negativity of their tweets. The more NO leads, the more negatively they tweet.

In a simpler model, there is a significant relationship between the number of days to the election and the negativity of the YES campaign. That is, the decline in positivity as the election is statistically significant.

There is also evidence that the YES campaign tweeted more negatively the day after each debate (which they were said to have lost)



Discussion

What does this tell us about the scottish election campagin?

What does this tell us about negative communication in political campaigns in general?

What potential confounding factors might explain the results?

Other examples

- [Baylis and Obradovich et al.](#) explore how weather affects sentiment
- [Schwalbach](#) looks at how political debate varies in sentiment through the legislative period, depending on their involvement in the government.
- [Wang et al.](#) explore how the COVID-19 pandemic affected how people expressed sentiment
- [Arratia et al.](#) describe how sentiment analysis is used to provide information about financial markets

Introduction and Objectives
○○○○

Intro to sentiment analysis
○○○○

Lexicon-based sentiment analysis
○○○○○○○○○○○○○○○○

Fancy sentiment analysis
○○○

Validation
○○○

Examples in research
○○○○○○○

Wrapup and Outlook
●○○○

Wrapup and Outlook

Wrapup

So far in this course, we have covered:

- How to get texts from tricky places
- How to process them and manipulate them
- How to represent them in different ways
- How to find the similarity between texts
- How to find what texts are about

Now we know how to find out what emotions are present in texts

Outlook

We said sentiment analysis is a special case of classification. We will explore this in detail next week when we cover **supervised text classification**.

We'll be training machine learning classifiers to assign documents to predefined classes, and learning how to evaluate how well these work.

This is the last major technique we will learn, before the penultimate session where we will find about how to apply a range of the techniques we have seen using the fanciest most up-to-date methods.

Introduction and Objectives
○○○○

Intro to sentiment analysis
○○○○

Lexicon-based sentiment analysis
○○○○○○○○○○○○○○○○

Fancy sentiment analysis
○○○

Validation
○○○

Examples in research
○○○○○○○

Wrapup and Outlook
○○○●