Text Analysis - Group Exercise

Bamberg Summer Institute in Computational Social Science

Carsten Schwemmer, University of Bamberg

2019-08-01

Many thanks to Chris Bail for providing material for this lecture

Today we covered a broad range of techniques for analyzing textual data. For today's exercise, there are two options among which you can choose:

- a more structured group exercise designed to compare the strengths and weaknesses of different text analysis techniques when they are applied to the same dataset.
- a "freestyle" exercise using the *DonorsChoose* dataset.

You can make this choice after dividing yourselves into groups of four.

Option 1 - Trump tweets (structured)

1) Load the dataframe of tweets by President Trump that we analyzed with dictionaries:

```
library(tidyverse)
load(url("https://cbail.github.io/Trump_Tweets.Rdata"))
```

2) Use at least two of the techniques we discussed this morning to pull out features from the text of Trump's tweets (e.g. substantive themes, topics, sentiment).
3) Work together to identify whether any features of Trump's Twitter language predict the number of retweets or likes his messages receive.

Optione 1 - Trump tweets (structured)

4) Load the dataframe of daily approval ratings for President Trump:

```
url <- "https://projects.fivethirtyeight.com/trump-approval-data/approval_topli
trump_approval <- read_csv(url)
```

5) Work together to determine whether there are any features of Trump's Twitter language that have an association with his overall approval ratings.

Option 2 - DonorsChooose (unstructured)

Use the dataset DonorsChoose dataset to apply any of the techniques we discussed today. Examples of questions you might ask:

- how has classroom technology use changed over time? How does it differ by geographic location and the age of students?
- how do the requests of schools in urban areas compare to those in rural areas?
- what predicts whether a project will be funded?
- how do the predictors of funding success vary by geographic location? Or by economic status of the students?

Share your findings

Regardless which exercise option you have chosen, produce one visualization that describes the findings of your analysis. Share your visualization on Slack until 3.45 PM.