Wk1_SLP3

March 8, 2021

0.1 Sentiment Analysis of Political Tweets

As discussed in class, sentiment analysis tries to measure the positive or negative sentiment expressed in a message. It is commonly used to track the mood of the nation on all kinds of topics through the analysis of Twitter data.

This is a simple example of using the NLTK sentiment analysis package to look at sentiment of tweets about the NSW Liberal and Labor parties. Unfortunately I've not been able to collect very much data from twitter because of their rate limits, so we only have around 250 tweets in each case. However, we can use this as a sample to test the basic methodology.

First we'll read in the tweets from a CSV file:

```python
import pandas as pd
import nltk
from nltk.sentiment.vader import SentimentIntensityAnalyzer
%matplotlib inline

#liberal = pd.read_csv('files/nswliberal.csv')
#labor = pd.read_csv('files/nswlabor.csv')

# updating for own file location
liberal = pd.read_csv('Wk1_SLP3_nswliberal.csv')
labor = pd.read_csv('Wk1_SLP3_nswlabor.csv')

liberal.head()
```

[6]:

```
                          created  \
0  Thu Oct 12 10:28:13 +0000 2017
1  Thu Oct 12 07:39:23 +0000 2017
2  Thu Oct 12 07:31:30 +0000 2017
3  Thu Oct 12 06:43:38 +0000 2017
4  Thu Oct 12 05:38:03 +0000 2017


                                               text
0  @WesFangNats @keredm @sffAustralia I'd hazard …
1  RT @venusboy1977: @AustralianLabor @RoseBrenna…
2  @AustralianLabor @RoseBrennan12 @BernardKeane …
3  The latest The Wollongong Happenings Daily! ht…
4  @sffAustralia @NSWNationals @LiberalNSW Two of…
```

```
[7]: labor.head()
```

```
[7]:                              created  \
     0  Thu Oct 12 20:27:43 +0000 2017
     1  Thu Oct 12 20:10:27 +0000 2017
     2  Thu Oct 12 10:01:23 +0000 2017
     3  Thu Oct 12 08:58:52 +0000 2017
     4  Thu Oct 12 07:52:19 +0000 2017


                                                    text
     0  #GladysBerejiklian Donations For Access Rife i…
     1  I don't think @NSWLabor got the message on gun…
     2  @stukhan @NSWLabor But it makes for one hell o…
     3  .@NSWLabor supports keeping Springvale mine op…
     4  RT @NSWLabor: What do you think ? https://t.co…
```

We can use the NLTK python package to do sentiment analysis. It contains all of the tools you need to build your own analyisis engine based on training data you have labelled yourself, but it also has a pre-built sentiment analysis engine called VADER. VADER uses word lists of positive and negative words and returns a set of ratings that count the proportion of positive and negative words in a text. A final metric *compound* is a combination score that is +1 for a very positive text and -1 for a negative one.

We first need to download the lexicon (word list) for NLTK to work, we then make a sentiment analyser and apply it to every row of the data with the apply method.

```
[8]: nltk.download('vader_lexicon')
     sid = SentimentIntensityAnalyzer()

     liberal_sent = liberal.apply(lambda r: sid.polarity_scores(r.text), axis=1)
     labor_sent = labor.apply(lambda r: sid.polarity_scores(r.text), axis=1)
```

```
     [nltk_data] Downloading package vader_lexicon to
     [nltk_data]     C:\Users\nikki\AppData\Roaming\nltk_data…
```

```
[9]: # turn the result into a data frame and add as columns to the original data
     df = pd.DataFrame(list(liberal_sent))
     liberal_s = liberal.join(df)

     df = pd.DataFrame(list(labor_sent))
     labor_s = labor.join(df)

     labor_s.head()
```

```
[9]:                              created  \
     0  Thu Oct 12 20:27:43 +0000 2017
     1  Thu Oct 12 20:10:27 +0000 2017
     2  Thu Oct 12 10:01:23 +0000 2017
```

2

```
3   Thu Oct 12 08:58:52 +0000 2017
4   Thu Oct 12 07:52:19 +0000 2017


                                          text    neg     neu     pos  \
0   #GladysBerejiklian Donations For Access Rife i…  0.000  1.000  0.000
1   I don't think @NSWLabor got the message on gun…  0.124  0.876  0.000
2   @stukhan @NSWLabor But it makes for one hell o…  0.464  0.536  0.000
3   .@NSWLabor supports keeping Springvale mine op…  0.168  0.745  0.087
4   RT @NSWLabor: What do you think ? https://t.co…  0.000  1.000  0.000


      compound
0       0.0000
1      -0.3400
2      -0.8316
3      -0.3898
4       0.0000
```

[10]: `liberal_s.head()`

[10]:
```
                       created  \
0   Thu Oct 12 10:28:13 +0000 2017
1   Thu Oct 12 07:39:23 +0000 2017
2   Thu Oct 12 07:31:30 +0000 2017
3   Thu Oct 12 06:43:38 +0000 2017
4   Thu Oct 12 05:38:03 +0000 2017


                                          text  neg     neu     pos  \
0   @WesFangNats @keredm @sffAustralia I'd hazard …  0.0  1.000  0.000
1   RT @venusboy1977: @AustralianLabor @RoseBrenna…  0.0  1.000  0.000
2   @AustralianLabor @RoseBrennan12 @BernardKeane …  0.0  1.000  0.000
3   The latest The Wollongong Happenings Daily! ht…  0.0  0.803  0.197
4   @sffAustralia @NSWNationals @LiberalNSW Two of…  0.0  0.885  0.115


      compound
0       0.0000
1       0.0000
2       0.0000
3       0.4926
4       0.3244
```

So, now you have sentiment data for the NSW Liberal and Labor parties. Your task now is to generate some useful intelligence from this data. You might ask:

- what is the spread of sentiment for each party
- who is ahead on the sentiment scale - how do they compare?
- what are the most positive and negative issues in each case?

Your checkpoint this week is to show us a useful plot or summary statistic based on this data along

with your interpretation.

```
[109]:  #print(labor_s['compound'].describe())
        #print(liberal_s['compound'].describe())

        import numpy as np

        def new_columns(df, party):
            df['sentiment'] = np.where(df['compound'] <= -0.05, "neg",
                                np.where(df['compound'] >= 0.05, "pos", "neu"))
            df['party'] = party

        new_columns(labor_s, "labor")
        new_columns(liberal_s, "liberal")

        all_s = labor_s.append(liberal_s)
        all_s.head()

        fig = all_s.boxplot(by = ['sentiment', 'party',], column = 'compound', figsize␣
         ␣= (12,8))
        fig.set_title('Compound sentiment score by tweet label and political party')
        fig.set_xlabel('')
        fig.set_ylabel('Sentiment score')
        fig.xaxis.grid(False)
        fig.yaxis.grid(False);
```
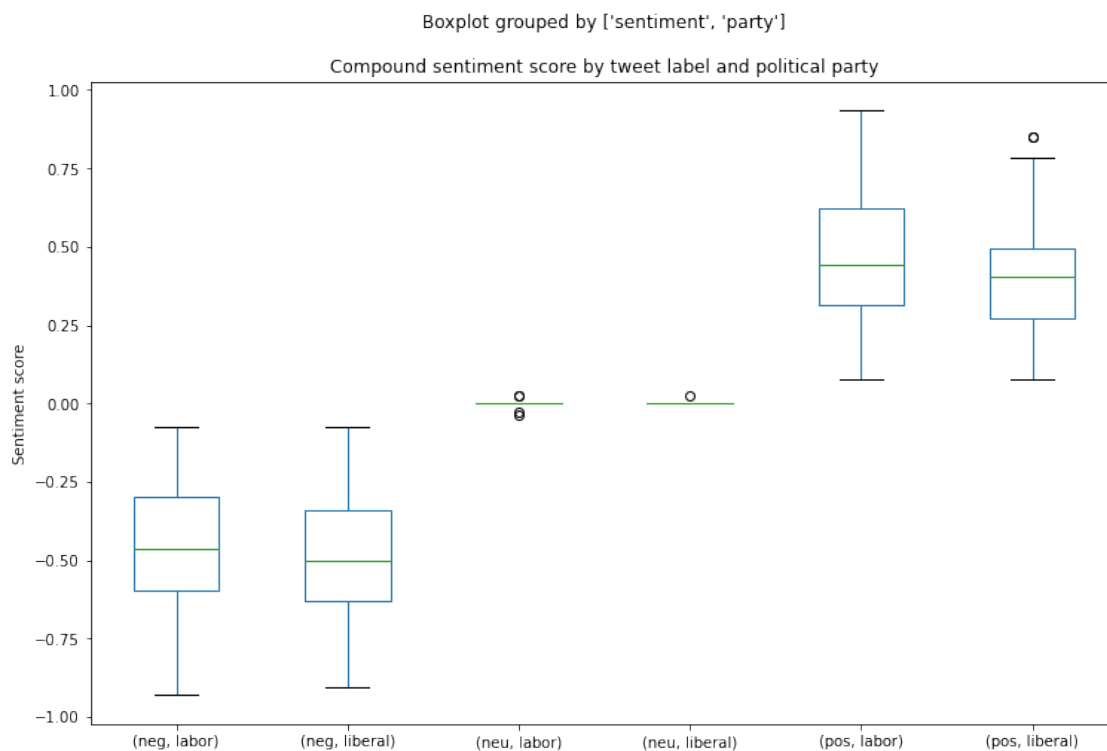
The boxplot indicates that there is just about no difference in the sentiment of the tweets from the labour and liberal parties for this corpus. That being said, the average sentiment of the liberal party tweets appears to be ever-so-slightly more negative than the average sentiment of the labor party.