DataLab

COVID & Social Unrest

12th August 2020
Outline

→ Data
→ Methodology
  ◆ Topics
  ◆ Sentiment
→ Shiny App
Data

615,862 Tweets
127 countries
137 media coverage
4 Languages:
- English
- Spanish
- French
- Portuguese
Feb - Now time range
A tweet posted by any media tracked by our API. It is collected and then stored in a NoSQL DB.

Language detection:
- It detects if the tweet language is in: EN, ES, FR or PT.
- The other languages are kept in the DB, but not used further steps.

Text processing:
- It removes the “stop words”: the, a, that, etc...
- Word stemming
- Bag of words: text is broken down into words

TOPIC keywords:
- Each topic is represented by a list of keywords.
- The list of keywords starts with a short list, and then the tool enlarge it with synonyms, related words.

A dichotomous variable is created in the DB for each topic.
- 1: Presence
- 0: Absence

Target topics:
- Social Unrest
- COVID
- COVID & Social Unrest
The tweets are **classified** into one of the **four categories** shown in the table below.

<table>
<thead>
<tr>
<th>Social Unrest</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>COVID</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Other</td>
<td>Social Unrest</td>
</tr>
<tr>
<td>Yes</td>
<td>COVID</td>
<td>COVID &amp; Unrest</td>
</tr>
</tbody>
</table>
Simple sentiment index based on the Python library: polyglot

The library has polarity lexicons for 136 languages.

The scale of the words’ polarity consisted of three degrees:
  ○ +1 for positive words,
  ○ -1 for negatives words.
  ○ Neutral words will have a score of 0.

There is a polarity dictionary for each language.
  ○ The dictionary is built from a set of seed words, and then the polarity of
    the other words is estimate via graph propagation.
Sentiment Index: polarity

- Simple sentiment index based on the Python library: *polyglot*
- The library has **polarity lexicons for 136 languages**.
- The scale of the words’ **polarity** consisted of three degrees:
  - +1 for positive words,
  - -1 for negatives words.
  - Neutral words will have a score of 0.

**Example:** “The movie was really good.”

<table>
<thead>
<tr>
<th>Word</th>
<th>Polarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>The</td>
<td>0</td>
</tr>
<tr>
<td>movie</td>
<td>0</td>
</tr>
<tr>
<td>was</td>
<td>0</td>
</tr>
<tr>
<td>really</td>
<td>0</td>
</tr>
<tr>
<td>good</td>
<td>1</td>
</tr>
<tr>
<td>.</td>
<td>0</td>
</tr>
</tbody>
</table>
The sentiment index for a tweet is the rate of positive words in the full text.

\[ SI = 100 \times \frac{\# \text{ positive words}}{\# \text{ positive words} + \# \text{ negative words}} \]
Shiny App

https://foodandagricultureorganization.shinyapps.io/sentiment/