Sentiment and Emotion analysis with topic modelling of COVID-19 tweets – PROJECT 1

Chilukuri, Susmitha
Supervisor: Yu, Fei
Advisor: Gibson, Amelia
Introduction

- Infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
- First identified in December 2019 – Wuhan, Hubei, China
- 2/11/2020 – name announced as COVID-19
- 3/2020 – Pandemic declaration by WHO
Impact of COVID-19

https://covid19.who.int/

Effects of COVID-19

- Physiologically
- Psychologically
- Economically
- Environmentally
- Politically, etc.

As per 11/8/2020, 4.30 pm CET, 49,578,590 confirmed cases 1,245,717 deaths
Objective

To conduct sentiment analysis of information posted on a social media platform about COVID-19 pandemic.
Research questions

• What has been posted and shared among the public?
• What are the publics major interests and perceptions?
• Strategies to use social media to promote health education and fight misinformation and disinformation
Why sentiment and emotion analysis?

Overview of wider public opinion
Method:

- Python
- Twitter - Twitter API - GetOldTweets3
- 1/1/2020 – 7/29/2020
- More than 200,000 tweets
- “Coronavirus” OR “COVID” OR “Hubei” OR “SARS-CoV-2” OR “2019nCoV” OR “Wuhan” OR “SARS-Coronavirus-2” OR “SARSCoV2” OR “nCoV2019” OR “Pandemic” OR “nCoV-2019”
Project workflow model

Data collection and extraction → Data pre-processing/text preparation → Topic modelling and labeling → Sentiment detection of topics → Sentiment classification of topics → Analysis and visualization

 → Emotion detection → Emotion classification of topics → Analysis and visualization
Data extraction and preprocessing

222,791 Tweets are collected using GetOldTweets3 from Python

216,022 are left after removing duplicates

6657 irrelevant tweets were removed leaving with 209441 for further analyses

Removing #, @, URL/Links(https), punctuations(special characters)

Tokenization

Removing stop words

Stemming and Lemmatization
# Topic modelling

19 topics – grouped into 3 themes

<table>
<thead>
<tr>
<th>CHINA</th>
<th>EFFECTS OF COVID-19 ON PEOPLE</th>
<th>GLOBAL EFFECTS OF COVID-19</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outbreak/Pneumonia</td>
<td>Travel restrictions</td>
<td>Lockdown</td>
</tr>
<tr>
<td>Origin</td>
<td>Healthcare workers</td>
<td>Economy</td>
</tr>
<tr>
<td></td>
<td>Schools</td>
<td>Global concerns</td>
</tr>
<tr>
<td></td>
<td>Stay at home</td>
<td>Global politics</td>
</tr>
<tr>
<td></td>
<td>Employment</td>
<td>USA COVID response</td>
</tr>
<tr>
<td></td>
<td>Parenting</td>
<td>Pandemic declaration</td>
</tr>
<tr>
<td></td>
<td>Quarantine</td>
<td>Transmission</td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>COVID statistics</td>
</tr>
<tr>
<td></td>
<td>Protests</td>
<td></td>
</tr>
</tbody>
</table>
Sentiment analysis - TextBlob

3 classes – based on polarity score
• Positive (>1)
• Negative (<1)
• Neutral (0)
## Emotion classification - NRCLex

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Maximum expressed topic</th>
<th>Least expressed topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>Lockdown</td>
<td>Transmission</td>
</tr>
<tr>
<td>Anticipation</td>
<td>Economy</td>
<td>Transmission</td>
</tr>
<tr>
<td>Disgust</td>
<td>Outbreak/Pneumonia</td>
<td>Transmission</td>
</tr>
<tr>
<td>Fear</td>
<td>Origin</td>
<td>Transmission</td>
</tr>
<tr>
<td>Joy</td>
<td>Parenting</td>
<td>Transmission</td>
</tr>
<tr>
<td>Sadness</td>
<td>Lockdown</td>
<td>Origin</td>
</tr>
<tr>
<td>Surprise</td>
<td>Global concern</td>
<td>Transmission</td>
</tr>
<tr>
<td>Trust</td>
<td>Stay at home</td>
<td>Transmission</td>
</tr>
<tr>
<td>Topic</td>
<td>Highest emotion</td>
<td>Lowest emotion</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Outbreak/Pneumonia</td>
<td>Fear</td>
<td>Surprise</td>
</tr>
<tr>
<td>Origin</td>
<td>Fear</td>
<td>Sadness</td>
</tr>
<tr>
<td>Travel restrictions</td>
<td>Fear</td>
<td>Surprise</td>
</tr>
<tr>
<td>Healthcare workers</td>
<td>Fear</td>
<td>Joy</td>
</tr>
<tr>
<td>Schools</td>
<td>Fear</td>
<td>Joy</td>
</tr>
<tr>
<td>Stay at home</td>
<td>Fear</td>
<td>Joy</td>
</tr>
<tr>
<td>Employment</td>
<td>Fear</td>
<td>Surprise</td>
</tr>
<tr>
<td>Parenting</td>
<td>Fear</td>
<td>Joy</td>
</tr>
<tr>
<td>Quarantine</td>
<td>Fear</td>
<td>Joy</td>
</tr>
<tr>
<td>Testing</td>
<td>Fear</td>
<td>Joy</td>
</tr>
<tr>
<td>Protests</td>
<td>Fear</td>
<td>Joy</td>
</tr>
<tr>
<td>Lockdown</td>
<td>Fear</td>
<td>Joy</td>
</tr>
<tr>
<td>Economy</td>
<td>Fear</td>
<td>Joy</td>
</tr>
<tr>
<td>Global concern</td>
<td>Fear</td>
<td>Joy</td>
</tr>
<tr>
<td>Global politics</td>
<td>Fear</td>
<td>Joy</td>
</tr>
<tr>
<td>USA COVID response</td>
<td>Fear</td>
<td>Joy</td>
</tr>
<tr>
<td>Pandemic declaration</td>
<td>Fear</td>
<td>Surprise</td>
</tr>
<tr>
<td>Transmission</td>
<td>Fear</td>
<td>Surprise</td>
</tr>
<tr>
<td>COVID statistics</td>
<td>Fear</td>
<td>Joy</td>
</tr>
</tbody>
</table>
Disgust is most expressed for Outbreak/Pneumonia which when linked with elements within it explains the ultimately leading xenophobia developed towards Chinese (real-world scenario).
Misinformation?

- False or inaccurate information, especially that which is deliberately intended to deceive.
- Fighting with misinformation is even tougher than dealing with COVID-19.
- Public engagement with misinformation related posts are raising day by day.
- Leads to confusion, anxiety, politicizing the situation, etc.
Social media

- Most powerful millennial tool
- Definition from Oxford: “Websites and applications that enable users to create and share content or to participate in social networking.”

Is this content credible?
Credible sources

- CDC (Centers for Disease Control and Prevention)
- WHO (World Health Organization)
- NIH (National Institute of Health)
- State Departments of Health, etc.
How to combat misinformation?

• Educating the people
  • How to identify
  • How to prevent spread
  • How to report
• Promoting reliable/ credible sources
References


White paper report on HealthDX (Global health data exchange) – PROJECT 2

Chilukuri, Susmitha

Supervisor: Mostafa, Javed
Advisor: Gibson, Amelia
Introduction

• COVID-19 pandemic made us realize the importance of global health data exchange and how far we are from achieving it.

• Global health exchange/Interoperability is essential for handling the global health issues in an effective way.

• The progress in the space of inter-health data exchange is slowed by the lack of an aggregated knowledge base that is fed in real-time by a community of practice.
Aim/Objective

The main aim of this paper is to emphasize the importance of global health data exchange with the establishment of HealthDX platform and systematic review of the pillars/layers it stand by.

Method: Systematic literature review to report the whole process of creating a platform like HealthDX and what key features it is built on and makes it unique.
• Multifaceted platform that brings stakeholders who are passionate about improving healthcare by using technology and capacity in an innovative way, focusing on health data exchange.

• 3 layers

• Focus is inter-organizational exchange that has a major impact on decision making at organizational, district, state, national and international levels.
Inter-organizational Exchange
Website

- https://healthdx.unc.edu/
- Using this as a platform for partnering with potential organizations and hosting a global health conference/webinar in March 2021.
Knowledge base (KB)

- Collection of information about a particular subject
- Changes with context
- 6 types
  - Internal KB
  - Hosted KB
  - Self-Hosted KB
  - Customer KB
  - Open-source KB software
  - External KB
## Database vs Knowledge Base

<table>
<thead>
<tr>
<th>Data base</th>
<th>Knowledge base</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organized collection of data</td>
<td>Collection of information on a particular subject</td>
</tr>
<tr>
<td>Typically represents facts in their basic form</td>
<td>Fully developed and ready to be applied</td>
</tr>
<tr>
<td>Supports storage and retrieval of data</td>
<td>Supports rapid search, retrieval and reuse of information</td>
</tr>
<tr>
<td>Designed mostly for concrete (Surface thinking) knowledge</td>
<td>Designed for abstract (Thinking in depth) knowledge</td>
</tr>
<tr>
<td>Collection of data representing facts</td>
<td>Information at higher level of abstraction</td>
</tr>
</tbody>
</table>
DIKW Pyramid

Knowledge sharing

Knowledge base

Information

Data

Knowledge management systems

Knowledge base systems

Informaton management systems

Wisdom
Characteristics of KB

• Comprehensive
• Accessible
• Intelligent
• Search
• Accurate
Structure/architecture of KB
Working model of HealthDX
Thank you!

Questions?