Uncertainty Now!

A Data Driven Analysis of Coping Strategies in the COVID-19 Pandemic.

Tim Ruhe

Data Science Kolloquium, June 17th 2021
tim.ruhe@tu-dortmund.de
Jan-Hendryk de Boer, Daniel Ullrich, Timo Lüke, Isabelle Borucki, Elena Weber, Alexander Unser, Jithender Timothy, Maren Freudenberg, Cristian Cercel
Premise: There are certain coping strategies that people apply to deal with situations of inherent uncertainty (e.g. the COVID-19 pandemic, the European refugee crisis, the worldwide banking crisis...)

Question: Which of these strategies are actually applied (expressed in social media) and how does that correlate for example with elections, regions or socio-economic factors (net income)?
WORKFLOW

DATA DOWNLOAD, CLEANING

Keywords: #CORONA, #Coronavirus, #COVID19...

SHK

1,5 Mio

3K

Encoding (Isabelle, Alexander)

Expert Knowledge

Tweets

SHK

SHK

Strategies

AI/DATA (Tim/Jithender)

Geocoding (1,5 Mio)

Vectorizer

Word2vec (1,5 Mio)

Classifier

(ANN)

> STATISTIK

Results/Analysis

Strategies: Correlation with Election Results, Socio-Economic Factors, ...

SHK
Distribution of Tweets and Sanity Check

**Top 10 Constituencies:**

1. Berlin-Mitte 75 (23.5%)
2. Eisenach – Wartburgkreis – Unstrut-Hainich-Kreis (8.4%)
3. Hamburg-Altona 19 (5.5%)
4. München-Ost 218 (4.5%)
5. Frankfurt am Main II 183 (4.0%)
6. Köln I 93 (2.5%)
7. Hannover-Land I (1.6%)
8. Dresden I (1.6%)
9. Mainz (1.6%)
10. Stuttgart I (1.6%)

**Null Hypothesis:**
Tweets expressing certain coping strategies are distributed in the same way as all tweets.
Criticism and Information

Relative number of tweets per constituency

Criticism

GYF VI
Unsicherheit jetzt!
preliminary

283 Tweets

Information

Relative number of tweets per constituency

GYF VI
Unsicherheit jetzt!
preliminary

807 Tweets
Criticism and Information

Meaningful conclusions are challenging, due to the small numbers.

Relative number of tweets per constituency

Criticism

Information

GYF VI
 Unsicherheit jetzt!
preliminary

283 Tweets

807 Tweets
Source Stacking in Astroparticle-Physics

Different source with similar properties that are individually too weak for significant measurement.
Source Stacking in Astroparticle-Physics

Idea: Pile individual sources on top of each other to obtain a significant measurement.

Stack constituencies for example by direct mandate
Simulation

- Obtain probability $p$ of a tweet to stem from a certain constituency
- Obtain number of tweets $n$ observed for a certain coping strategy
- Sample $n$ tweets according to probability $p$
- Repeat $n_{\text{trial}}$ time

- Group tweets by criterion, e.g. direct mandate, compute expectation
- Compare expectation and observation
Possible Bias towards larger cities

Munich (North)

Hamburg (Mitte)

Berlin (Mitte)
Selected (and preliminary) Results

All constituencies

Constituencies above 2 sigma excluded
More Results on Information

SPD
GYF VI, Uncertainty now! preliminary

CDU
GYF VI, Uncertainty now! preliminary

- Observed Value
- 0.01 Quantile
- 0.05 Quantile
Summary & Outlook

Relative number of tweets per constituency

674713 Tweets

Information (Direct Mandates), GYF VI, Uncertainty now! preliminary

Expected vs. Observed for CDU, CSU, SPD