News As A Predictor

• • •

Brock Trevor Weipeng Henok

Goal

Look at events that have happened in the past or currently happening and design a model that could be used to predict future civil conflicts based on selected attributes.

Tools Used

BigQuery

GDELT Analysis Services

Jupyter

MATLAB

BigQuery

BigQuery is Google's fully managed, petabyte scale, low cost enterprise data
warehouse for analytics.
We used bigquery to select the events we were interested in and to do that we used
SQL.

GDELT

GDELT Project monitors the world's broadcast, print, and web news from nearly every
corner of every country in over 100 languages ... on the entire world.

The datasets that we analyzed were all from the GDELT database.

Conflict Zone Events Analyzed

Brexit   Egyptian Revolution   Syrian Civil War

Ukrainian Revolution   The First Chechen Rebellion

Somali Rebellion   Iraq Civil War

Philippine Revolution   Tulip Revolution

Bangladesh Rifles   South Sudanese Civil War

Tunisian Revolution   The Iran Student Protest

8888 Uprising in Burma   Kitchenware Revolution

Philippine Revolution   Failed Azerbaijan Revolution

Shabag Protests   Kivu Conflict

Non-Conflict Zone Countries Analyzed

Denmark	Austria
Portugal	New Zealand
Poland	Czech Republic
Switzerland	Canada
Japan	Bhutan
Finland	Sweden
Australia	Norway
Netherlands	Mauritius
Belgium	

The Process

Data Preprocessing

Attribute Subset Selection

Classification

Model Evaluation

Attribute Subset Selection

GDELT

    SQLDate

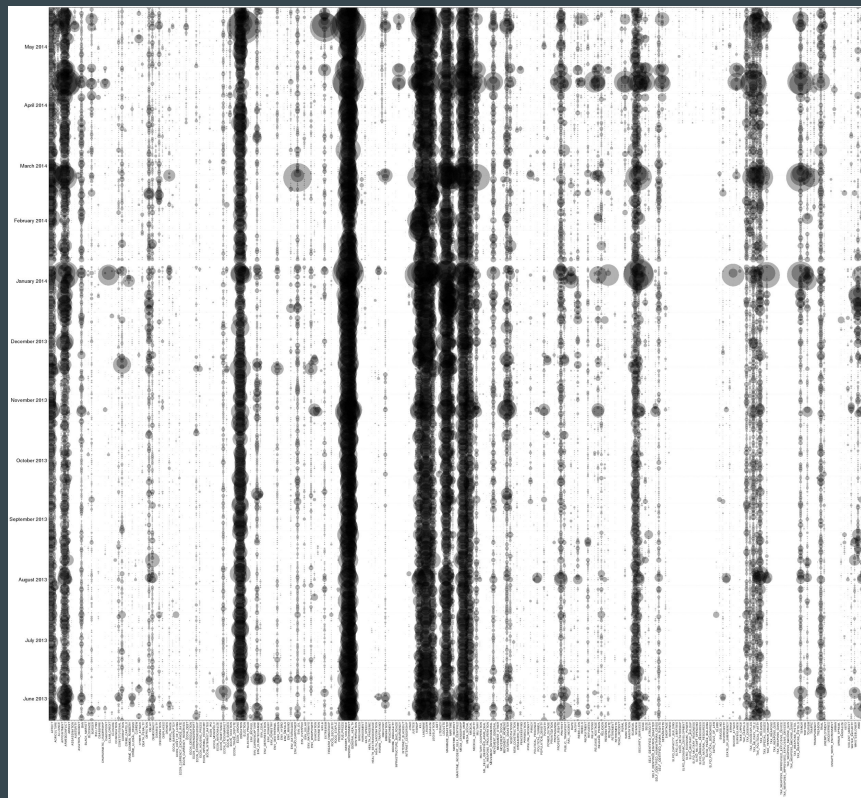   Actor1CountryCode

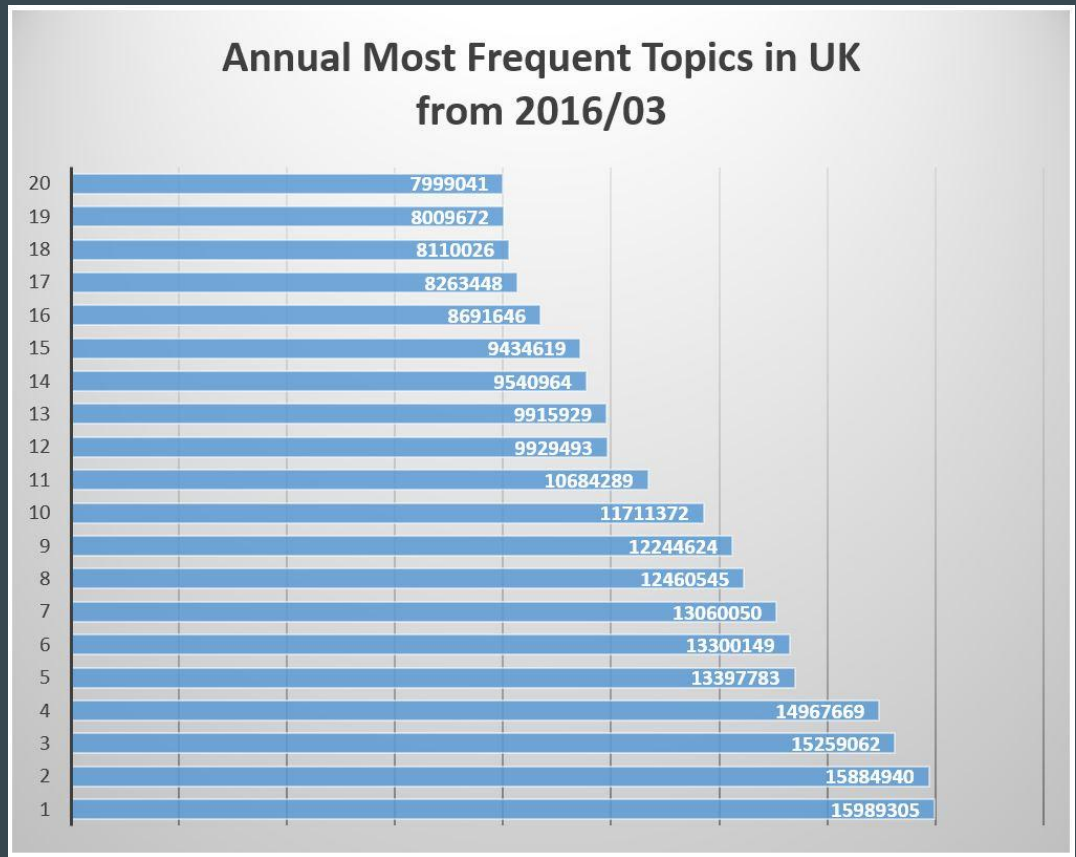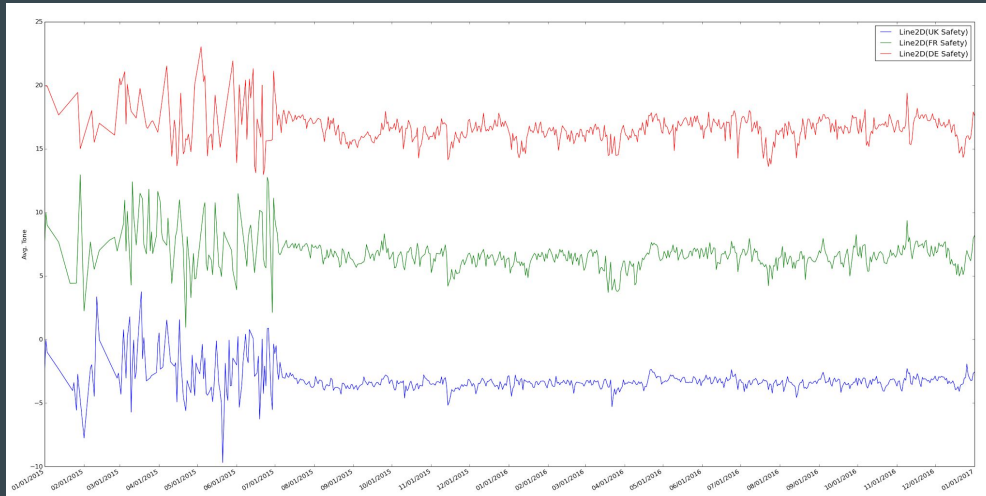   EventRootCode

GKG
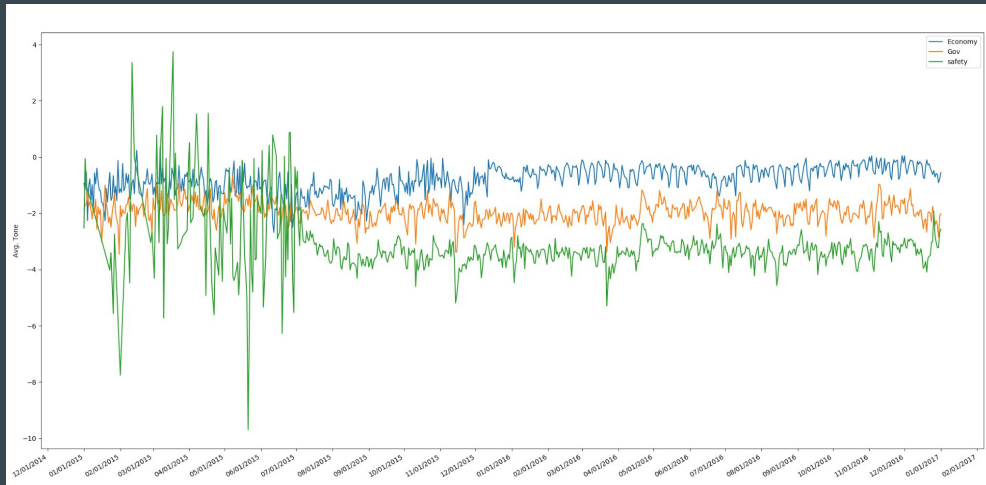
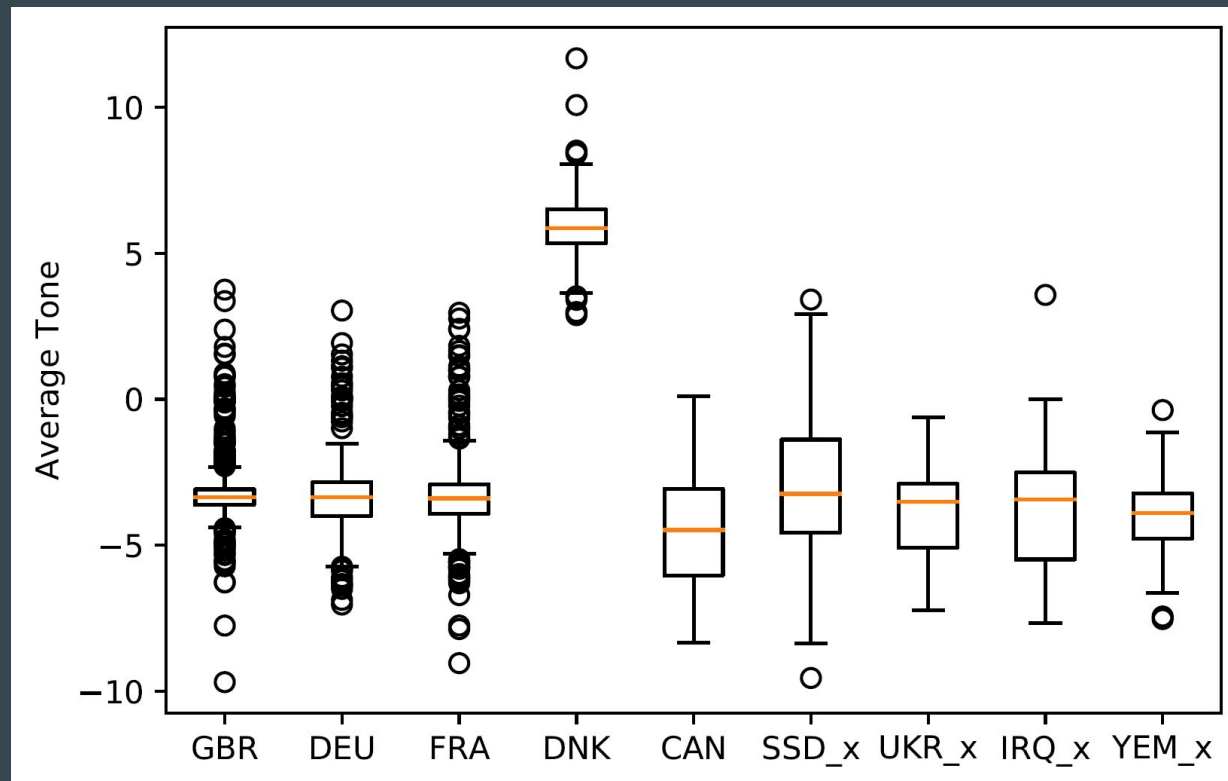   Date

    Locations

    V2Themes

    V2Tone

Theme Intensity Analysis

First Approach

Annual Most Frequent Topics in UK
from 2016/03

20	7999041
19	8009672
18	8110026
17	8263448
16	8691646
15	9434619
14	9540964
13	9915929
12	9929493
11	10684289
10	11711372
9	12244624
8	12460545
7	13060050
6	13300149
5	13397783
4	14967669
3	15259062
2	15884940
1	15989305

First Approach

Average Tones Result

Classification

Use Event Attributes

Randomly partition 24 countries (12 conflict, 12 peaceful) to training set and the remaining countries to a testing set

Do some data analysis

2 Classifications Techniques

     Correlation between testing data and the conflict and peaceful event categories

     Average normalized difference

Evaluation

Correlation Model

3328	2924
2672	3076

Normalized Difference Model

3445	671
2555	5329

	Correlation Model	Normalized Difference Model
Accuracy	0.5337	0.7312
Sensitivity	0.5547	0.5742
Specificity	0.5127	0.8882
Precision	0.5323	0.8370

Bonus

Model predicts that USA will not go through any major civil conflict in the next year!

Concluding remarks

Correlation of event distribution shape is inadequate for classification

Normalized difference provides better results

Further Work

  Model Improvement

    Ensemble Methods

    Bagging

    Boosting

    Combining Attributes