

aggregation

March 16, 2016

Contents

1	Setup	1
2	Exploring the data	2
2.1	Structure	2
2.2	Pandas Overview	3
2.3	Source and Target Countries	4
3	CAMEO Score aggregation	5
3.1	Goldstein mean	5
3.2	Goldstein sum	5
3.3	Goldstein counts (positive and negative)	6
3.4	Duvall and Thompson counts	6

1 Setup

We are working with the data with the help of python and the **pandas** library.

```
1 import numpy
2 import pandas as pd
3 import matplotlib
4 import matplotlib.pyplot as plt
5
6 d = pd.read_csv('events.northkorea.csv',sep='\t')
```

We are interested in the events where either the source or the target country is North Korea.

```

1 isSource = d['Source Country'] == 'North Korea'
2 isTarget = d['Target Country'] == 'North Korea'
3
4 d = d[isSource | isTarget]

```

2 Exploring the data

2.1 Structure

ICEWS event data have the following columns:

```
1 d.columns
```

```

Index([u'Event ID', u'Event Date', u'Source Name', u'Source Sectors',
       u'Source Country', u'Event Text', u'CAMEO Code', u'Intensity',
       u'Target Name', u'Target Sectors', u'Target Country', u'Story ID',
       u'Sentence Number', u'Publisher', u'City', u'District', u'Province',
       u'Country', u'Latitude', u'Longitude'],
      dtype='object')

```

A sample of the first five rows:

```
1 d[:5]
```

	Event ID	Event Date	Source Name \
0	927032	1/5/95	Kim Young Sam
1	927031	1/5/95	Kim Young Sam
2	927033	1/5/95	North Korea
3	927111	1/5/95	Kim Jong-Il
4	927112	1/5/95	Kim Jong-Il

	Source Sectors	Source Country \
0	Executive Office,Government,Executive,Center R...	South Korea
1	Executive Office,Government,Executive,Center R...	South Korea
2	NaN	North Korea
3	Parties,(National) Major Party,Government Majo...	North Korea
4	Parties,(National) Major Party,Government Majo...	North Korea

	Event Text	CAMEO Code	Intensity \
0	Make an appeal or request	20	3.0
1	Make statement	10	0.0

2	Use conventional military force	190	-10.0
3	Express intent to release persons or property	353	7.0
4	Praise or endorse	51	3.4

	Target Name \
0	North Korea
1	North Korea
2	Kim Young Sam
3	Foreign Affairs (United States)
4	Foreign Affairs (United States)

	Target Sectors	Target Country	Story ID \
0	NaN	North Korea	28237624
1	NaN	North Korea	28237624
2	Executive Office,Government,Executive,Center R...	South Korea	28237624
3	Government,Foreign Ministry,Executive	United States	28241798
4	Government,Foreign Ministry,Executive	United States	28241798

	Sentence Number	Publisher	City	District \
0	2	The Associated Press Political Service	Seoul	NaN
1	4	The Associated Press Political Service	Seoul	NaN
2	5	The Associated Press Political Service	Seoul	NaN
3	1	Reuters News	NaN	NaN
4	1	Reuters News	NaN	NaN

	Province	Country	Latitude	Longitude
0	Seoul-teukbyeolsi	South Korea	37.5683	126.9780
1	Seoul-teukbyeolsi	South Korea	37.5683	126.9780
2	Seoul-teukbyeolsi	South Korea	37.5683	126.9780
3	Washington D.C.	United States	38.9171	-77.0003
4	Washington D.C.	United States	38.9171	-77.0003

2.2 Pandas Overview

Pandas gives us the following summary

```
1 d.describe()
```

	Event ID	CAMEO Code	Intensity	Story ID \
count	258465.000000	258465.000000	258465.000000	258465.000000
mean	10250022.922086	66.694903	1.345124	11028691.885044

std	5476433.304771	133.020745	3.508172	10040802.923762
min	927031.000000	10.000000	-10.000000	423.000000
25%	5655885.000000	33.000000	0.000000	3484897.000000
50%	9768434.000000	42.000000	1.900000	7944549.000000
75%	14496613.000000	51.000000	3.400000	12585791.000000
max	21920682.000000	2042.000000	10.000000	37587232.000000

	Sentence Number	Latitude	Longitude
count	258465.000000	258465.000000	258465.000000
mean	2.928973	37.959548	102.849428
std	1.747469	8.477567	59.051402
min	1.000000	-90.000000	-175.200000
25%	1.000000	37.568300	116.397000
50%	3.000000	39.033900	125.754000
75%	4.000000	39.033900	125.804000
max	6.000000	70.383300	178.441000

2.3 Source and Target Countries

What is the most common source country? Target country?

```
1 d['Source Country'].value_counts()[:6]
```

```
North Korea      119607
United States    37246
South Korea      32951
China            19764
Japan            17772
Russian Federation 8596
Name: Source Country, dtype: int64
```

```
1 d['Target Country'].value_counts()[:6]
```

```
North Korea      154748
United States    26585
South Korea      25454
China            16024
Japan            12764
Russian Federation 6543
Name: Target Country, dtype: int64
```

Are there any missing data? Let's check the fill rate.

```
1 d.count()/d.shape[0] * 100
```

Event ID	100.000000
Event Date	100.000000
Source Name	100.000000
Source Sectors	50.529472
Source Country	96.794924
Event Text	100.000000
CAMEO Code	100.000000
Intensity	100.000000
Target Name	100.000000
Target Sectors	35.832705
Target Country	97.828720
Story ID	100.000000
Sentence Number	100.000000
Publisher	100.000000
City	81.482599
District	2.539996
Province	86.423307
Country	100.000000
Latitude	100.000000
Longitude	100.000000

dtype: float64

3 CAMEO Score aggregation

We would like to aggregate the CAMEO scores of all data per some time unit (day) into a single new one to give us our time series. The existing literature indentifies four popular ways to do that.

3.1 Goldstein mean

Using this method, Goldstein scores are averaged per unit time.

3.2 Goldstein sum

In this method, the Goldstein scores are summed per unit time.

3.3 Goldstein counts (positive and negative)

Using this method, counts are generated for positive and negative Goldstein scores per unit time. This can be thought of as "net cooperation" and "net conflict."

3.4 Duvall and Thompson counts

Similar to Goldstein counts, Duvall and Thompson counts categorize events by cooperation or conflict. Duvall and Thompson counts, however, place events into four categories:

- Verbal Cooperation: The occurrence of dialogue-based meetings, statements that express a desire to cooperate or appeal for assistance from other actors.
- Material Cooperation: Physical acts of collaboration or assistance, including receiving or sending aid, reducing bans and sentencing, etc.
- Verbal Conflict: A spoken criticism, threat, or accusation, often related to past or future potential acts of material conflict.
- Material Conflict: Physical acts of a conflictual nature, including armed attacks, destruction of property, assassination, etc.