aggregation

March 14, 2016

Contents

1 Setup

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

```
import numpy
import pandas as pd

d = pd.read_csv('events.northkorea.csv',sep='\t')
```

2 Exploring the data

2.1 Structure

ICEWS event data have the following columns:

```
d.columns
```

A sample of the first five rows:

```
d[:5]
```

2.2 Pandas Overview

Pandas gives us the following summary

```
d.describe()
```

2.3 Source and Target Countries

What is the most common source country? Target country?

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

```
d[’Target Country’].value_counts()[:6].plot(kind=’bar’)
```

3 CAMEO Score aggregation

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

3.1 Goldstein mean

3.2 Goldstein sum

3.3 Goldstein counts (positive and negative)

3.4 Duvall and Thompson counts