Systematizing Notes on Conflict Events

Cristina Mac Gregor | Gonzalo Pons | Darshan Sumant

Three main lines of analysis:

- Prediction of agents
- Relation extraction
- Topic analysis

Data: ACLED - Armed Conflict Location & Event

EVENT_ID_CNTY	EVENT_DATE	LOCATION	LATITUDE	EVENT_TYPE
YEM27135	05-January-2019	Baqim as Suq	17.397	Battle-no change \nof territory
YEM27136	05-January-2019	Jabal Marran	16.798	Battle-no change \nof territory
YEM27137	05-January-2019	Taizz-Ar Rawdah	13.593	Remote violence
YEM27138	05-January-2019	Taizz	13.580	Remote violence
YEM27139	05-January-2019	Taizz-Usayfirah	13.567	Battle-no change \nof territory

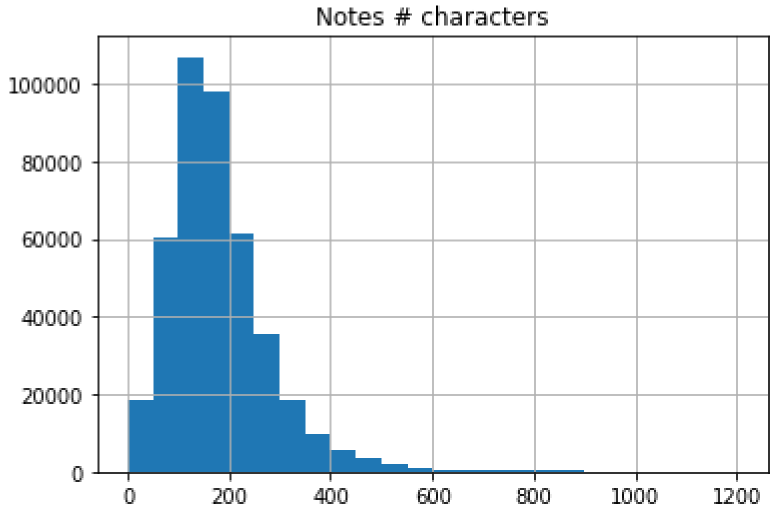
ACTOR1	ACTOR2	INTERACTION
Military Forces of Yemen (2012-)	Military Forces of Yemen (2016-) Supreme Polit...	11
Military Forces of Yemen (2016-) Supreme Polit...	Military Forces of Yemen (2012-)	11
Military Forces of Yemen (2016-) Supreme Polit...	Civilians (Yemen)	17
Military Forces of Yemen (2012-)	Military Forces of Yemen (2016-) Supreme Polit...	11
Police Forces of Yemen (2012-)	Militia (Ghazwan al Mikhlafi)	13

NOTES
20 Houthi soldiers were reportedly killed and ...
Anti-Houthi soldiers claim to have stopped an ...
At least 2 people (civilians) were killed and ...
4 Houthi fighters were killed and several othe...
Fighters loyal to a local militia leader "Ghaz...

Data: ACLED – Armed Conflict Location & Event

NOTES examples:

“26th Feb 2001- BBC Mon-Large military offensive all over the country sees 9 soldiers and 6 GIA killed” (101 characters)



*Max = 1,708

“A 40-year-old repentant who answered to the name of Hamid Doghman was assassinated this past Tuesday 4 March at about 2000 hours not far from downtown Zemmouri 16 kilometres east of Boumerdes by a militant group made up of between four and six elements. The attack targeted this ex-Islamist who had signed his repentance in August 1995.”

GOAL 1: PREDICT ACTORS

- Tokenize and vectorize the text input.
- Using existing software we initialize word embeddings from the text.
- Use SpaCy or Stanford NER to perform regular Name Entity Recognition and POS inputs
- Split data in training, development and test. Training will be the oldest (date) 60% of news. Random split of validation/test.
- Using these inputs, build a neural network to predict:
  - INTER1 and INTER2 OR INTERACTION.
  - FATALITIES
- Choose best model/weights with development set.
- Calculate precision, recall, and maximize Area Under Curve (AUC) measure in test set.

ACTOR1	INTER1	ACTOR2	INTER2	INTERACTION	FATALITIES	NOTES
GIA: Armed Islamic Group	Rebel Groups (2)	Military Forces of Algeria (1999-)	State Forces (1)	12	1	26th Feb 2001- BBC Mon-Large military offensive all over the country sees 9 soldiers and 6 GIA killed
Unidentified Armed Group (Algeria)	Political Militias (3)	Civilians (Algeria)	Civilians (7)	37	1	A 40-year-old repentant who answered to the name of Hamid Doghman was assassinated this past Tuesday 4 March at ...

GOAL 1: PREDICT ACTORS

“26th Feb 2001- BBC Mon-Large military offensive all over the country sees 9 soldiers and 6 GIA killed”

Name Entity Recognition:

STANFORD NER

Type: ORGANIZATION, Value: BBC

Type: ORGANIZATION, Value: GIA

STACY

26th	ORDINAL
BBC Mon-Large	ORG
9	CARDINAL
6	CARDINAL

POS:

[(26th, 'ADJ'), (Feb, 'PROPN'), (2001-, 'PROPN'), (BBC, 'PROPN'), (Mon, 'PROPN'), (-, 'PUNCT'), (Large, 'ADJ'), (military, 'ADJ'), (offensive, 'NOUN'), (all, 'ADV'), (over, 'ADP'), (the, 'DET'), (country, 'NOUN'), (sees, 'VERB'), (9, 'NUM'), (soldiers, 'NOUN'), (and, 'CCONJ'), (6, 'NUM'), (GIA, 'PROPN'), (killed, 'VERB')]

INTERACTION :
“MILITARY VERSUS REBELS”

FATALITIES: 1

GOAL 2: PREDICT RELATION & DIRECTION

ACTOR 1 **--- (ACTION)** → ACTOR 2

ACTOR 1 ← **(ACTION) ---** ACTOR 2

GOAL 2: PREDICT RELATION & DIRECTION

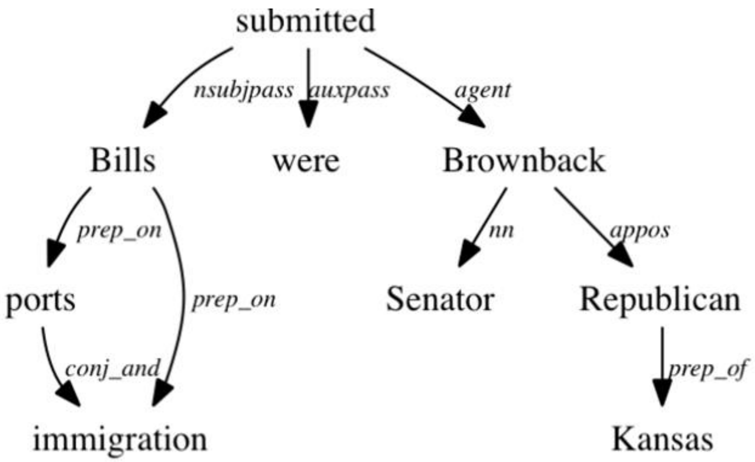
ACTOR 1 **--- (ACTION)** → ACTOR 2

ACTOR 1 ← **(ACTION) ---** ACTOR 2

RNN with handcrafted features (Keith et all):

- Syntactic dependencies
  - Different iterations of dependency paths that include ACTOR 1 & ACTOR 2:
    - With length 2 and with length 3
- N-gram features
  - Ngrams
  - POS windows centered on ACTOR

GOAL 2: PREDICT RELATION & DIRECTION



- Syntactic dependencies
  - Different iterations of dependency paths that include ACTOR 1 & ACTOR 2:
    - With length 2 and with length 3

GOAL 2: PREDICT RELATION & DIRECTION

[177] *'Troops have shot dead a commander of Algerias most radical guerrilla faction who had been sought for scores of killings, a pro-government newspaper said. LAuthentique said Hamou Eulmi, known by his nom de guerre Zinedine, was killed Tuesday near a mos'*

Troops shot guerrilla

[192] *'25 March 1999 The Globe and Mail Two girls aged 2 and 3 and a woman were among nine victims who had their throats cut overnight by suspected Islamic extremists at an isolated farm south of Algiers, security services said yesterday.'*

Civilians had throats cut by suspected extremists.

[1123] *'One militant was also killed Thursday in a clash with security forces in Beni Beshir, close to Skikda.'*

Militant killed by security forces

[19293] *'Violent fight between Seleka rebels and rioters from the Gobongo in Bangui'*[112212] *'Al Shabaab attack a military base. No reported casualties.'*

Al Shabab attack to military base

GOAL 2: PREDICT RELATION & DIRECTION

```
[('the', 677693),
 ('in', 492704),
 ('of', 456125),
 ('and', 324766),
 ('a', 293996),
 ('on', 230236),
 ('to', 166074),
 ('were', 143442),
 ('forces', 98638),
 ('by', 93749),
 ('killed', 87105),
 ('at', 85940),
 ('was', 80448),
 ('reported', 73558)
 ('protest', 73046),
 ('against', 73045),
 ('an', 66592),
 ('with', 63296),
 ('police', 60726),
```

```
[('forces', 98638),
 ('killed', 87105),
 ('reported', 73558),
 ('protest', 73046),
 ('police', 60726),
 ('area', 54486),
 ('district', 54456),
 ('two', 45376),
 ('fatalities', 45082),
 ('staged', 41551),
 ('city', 39068),
 ('al', 39049),
 ('injured', 38883),
 ('people', 35799),
 ('militants', 34736),
 ('village', 34615),
 ('town', 32373),
 ('clashes', 30781),
 ('attack', 30374),
```

GOAL 2: PREDICT RELATION & DIRECTION

```
[('the', 677693),
 ('in', 492704),
 ('of', 456125),
 ('and', 324766),
 ('a', 293996),
 ('on', 230236),
 ('to', 166074),
 ('were', 143442),
 ('forces', 98638),
 ('by', 93749),
 ('killed', 87105),
 ('at', 85940),
 ('was', 80448),
 ('reported', 73558)
 ('protest', 73046),
 ('against', 73045),
 ('an', 66592),
 ('with', 63296),
 ('police', 60726),
```

```
[('forces', 98638),
 ('killed', 87105),
 ('reported', 73558),
 ('protest', 73046),
 ('police', 60726),
 ('area', 54486),
 ('district', 54456),
 ('two', 45376),
 ('fatalities', 45082),
 ('staged', 41551),
 ('city', 39068),
 ('al', 39049),
 ('injured', 38883),
 ('people', 35799),
 ('militants', 34736),
 ('village', 34615),
 ('town', 32373),
 ('clashes', 30781),
 ('attack', 30374),
```

GOAL 2: PREDICT RELATION & DIRECTION

Main Challenges around Extracting Relations

1. we’re *predicting* the actor1 and actor2 items. Moreover, they’re not always fully embedded in the text.
2. no labelled data.
3. given the length of the document, we can find many verbs not related to the main *conflict* action [reported, described]

GOAL 3: TOPIC ANALYSIS & FINDING PATTERNS

Existing Methodology uses an Unsupervised Approach

1. Doc2Vec (based on 5+ million Wikipedia articles)
2. Pairwise cosine similarity between all pairs of (training) documents
3. Min Spanning Tree ($\boldsymbol{G_S}$) created for the entire dataset
    a. Nodes are articles
    b. Weight edges correspond to similarity between end-nodes
4. Markov Stability defines hard clusters based on $\boldsymbol{G_S}$

There's no way to predict what the Clusters or Topics would look like or mean!