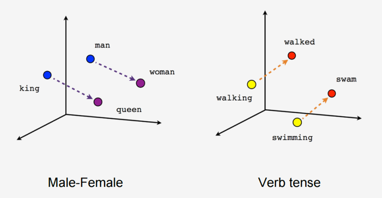
One-to-X analogical reasoning on word embeddings: a case for diachronic armed conflict prediction from news texts

Andrey Kutuzov, Erik Velldal, Lilja Øvrelid
Language Technology Group, University of Oslo, Norway
{andreku,erikve,liljao}@ifi.uio.no

What is wrong with standard word analogies?

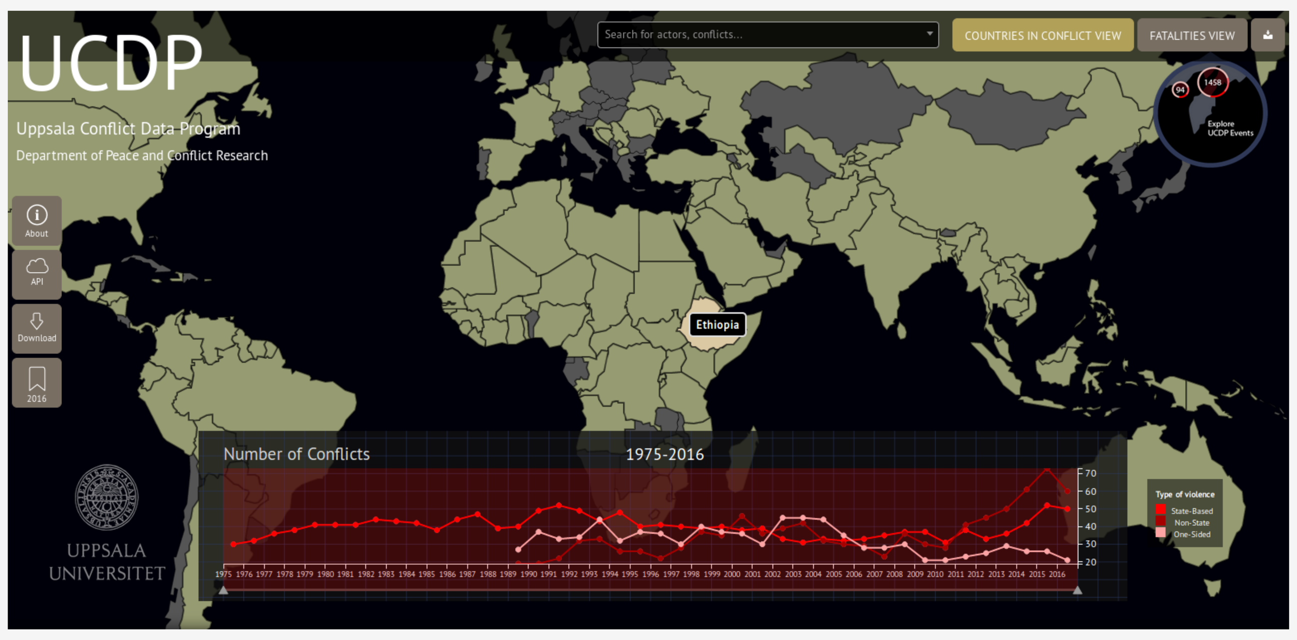


- Analogical inference ('word analogies') is used to evaluate word embeddings [Mikolov et al., 2013]
  - 'KING is to QUEEN as MAN is to ? (WOMAN)'
- A relational similarity task [Jurgens et al., 2012]
- Problem: exactly one best answer for each question:
  - WOMAN and GIRL cannot be both correct answers.

One-to-X analogies

- We extend analogical inference to include multiple-ended relations:
  - one-to-one ('*Jack* and *Jill* are friends')
  - one-to-many ('*Jack* and *Olaf* are also friends')
  - one-to-none ('*John* has no friends')
- For a vocabulary V, a relation z, and an entity $x \in V$,
- identify all pairs $x; i \in V$ such that z holds between x and i,
- providing as many correct answers as possible, and as few incorrect answers as possible.

Historical armed conflicts data



https://www.ucdp.uu.se/

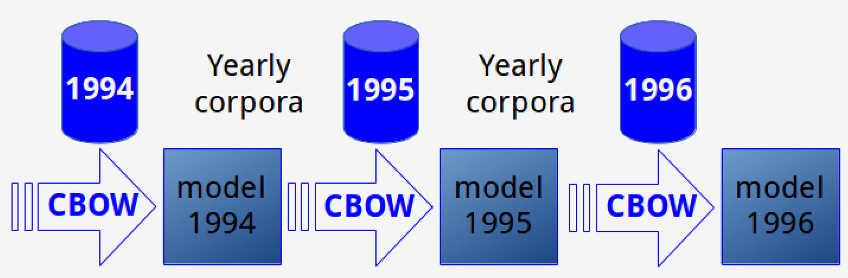
- We use one particular type of asymmetric semantic relations:
- a geographical location (country) and an insurgent group in an armed conflict against the government of the country:
  - several armed groups can operate in one location (one-to-many)
  - one armed group can operate in several locations (many-to-one)
  - some locations are peaceful: no armed groups there (one-to-none)
- Easily extended to diachronic setup: armed conflicts start and end.
- Historical armed conflicts data from the UCDP project [Gleditsch et al., 2002]
  - UCDP/PRIO Armed Conflict Dataset [Pettersson and Eck, 2018]
  - Example entry: '2016: Afghanistan: ["Taliban", "Islamic State"]'

UCDP data subsets

	Gigaword	NOW (News on Web)
Corresponding corpus	[Parker et al., 2011]	https://corpus.byu.edu/now/
Corpus size, tokens	4.8 billion	5.9 billion
Time span	1995–2010	2010–2017
Locations	52	42
Insurgents	127	78
Conflict pairs	136	102
New pairs share	37%	39%
Conflict locations share	46%	56%
Insurgents per location	1.65	1.50

Incremental diachronic word embeddings

Word embeddings retain enough structure to trace a relation after the model was additionally trained with new in-domain texts.

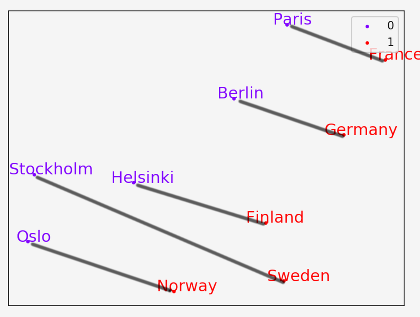


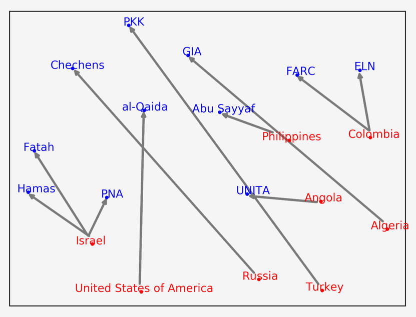
Diachronic (temporal) word embeddings with incremental training

The model M_{n+1} is initialised with the weights from the model M_n; if there are new words in the $n+1$ data which exceed the frequency threshold, then at the start of M_{n+1} training they are added to it and assigned random vectors.

Learning a 'projection/transformation' matrix

1. Apply 'semantic directions' (learned on the previous year data) to the next year.
2. If we know the 'Location: Insurgent' pairs from a time period n, we can find pairs with the same relation in $n+1$.
3. The input: gold pairs for the year n and their embeddings from the model M_n.
4. Linear projection $T \in \mathbb{R}^{p \times d}$ trained for each year pair ('2010–2011', '2011–2012'...)
   - p is the number of pairs, and d is the vector size





- Linguistically: T matrix is a *prototypical armed conflict relation*;
- Geometrically: '*average direction*' from locations to active insurgent groups in M_n.
- Optimal T is found by solving d normal equations (simple linear regression).
- For any location v, there is its 'armed conflict projection': $\hat{i} = v \cdot T$

Evaluation setup

1. Each yearly test set contains all locations (some peaceful).
2. Predict correct sets of insurgents for conflict areas and empty sets for peaceful areas.
3. 'Armed conflict projection' $\hat{i}$ produced for each location using T_n.
4. k nearest neighbours of $\hat{i}$ in M_{n+1} are predicted insurgents ('baseline').

Precision, recall and F1 score (with false negatives), averaged across all years in the test set.

Cosine threshold

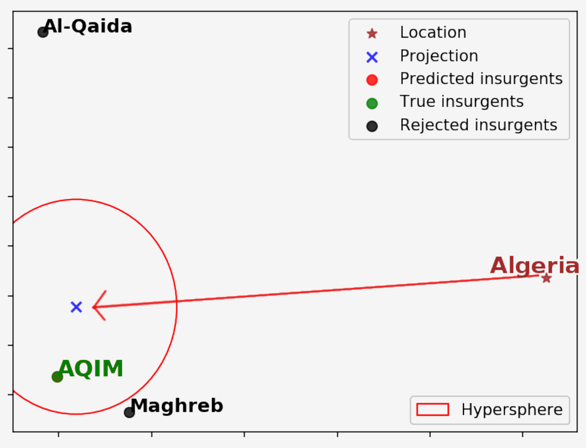
Problem: the 'baseline' system will always yield k incorrect candidates for peaceful areas.
Solution:

- real insurgents are closer to $\hat{i}$ than other nearest neighbours
- learn a hypersphere with the radius r as a cosine threshold:

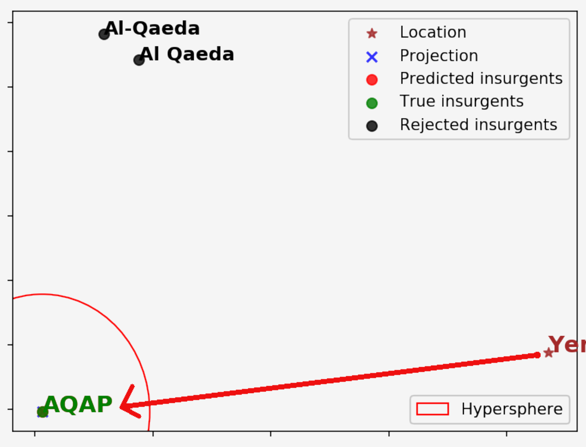
$$r = \frac{1}{p}\sum_{p=0}^{p} \cos\left(\hat{i}_p, g_p\right) + \sigma$$

- ...g_p is the insurgent in the p[th] pair, and σ is one stdev of the cosine distances in p

Keep only the candidates within the hypersphere inferred from the previous year.



Prediction of armed groups in Algeria, 2014



Prediction of armed groups in Yemen, 2011

Experiments ($k = 2$)

Projection matrix T_n and the threshold r_n are applied to the year $n+1$:

Dataset	Algorithm	Precision	Recall	F1
Gigaword	Baseline	0.19	0.51	0.28
	Cosine threshold	0.46	0.41	**0.41**
NOW	Baseline	0.26	0.53	0.34
	Cosine threshold	0.42	0.41	**0.41**

Summary

1. Word analogy task reformulated: multiple correct answers or no correct answer at all (one-to-X relations).
2. Temporal dataset of armed conflicts to evaluate one-to-X analogies.
3. Incremental word embeddings solve diachronic one-to-X analogies.
4. Learned cosine threshold can significantly improve the temporal one-to-X analogies performance by filtering out false positives.

Code, datasets, trained diachronic embeddings:
https://github.com/ltgoslo/diachronic_armed_conflicts