Tracing armed conflicts with diachronic word embedding models

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

General overview

We employ diachronic word embedding models in the task of predicting the events of armed conflicts escalating or calming down in various geographical locations, spanning over 16 years (1994–2010).
The task is similar to that of detecting semantic shifts [Kulkarni et al., 2015], [Hamilton et al., 2016a], but focused more on subtle changes of perspective instead of full-scale meaning changes.

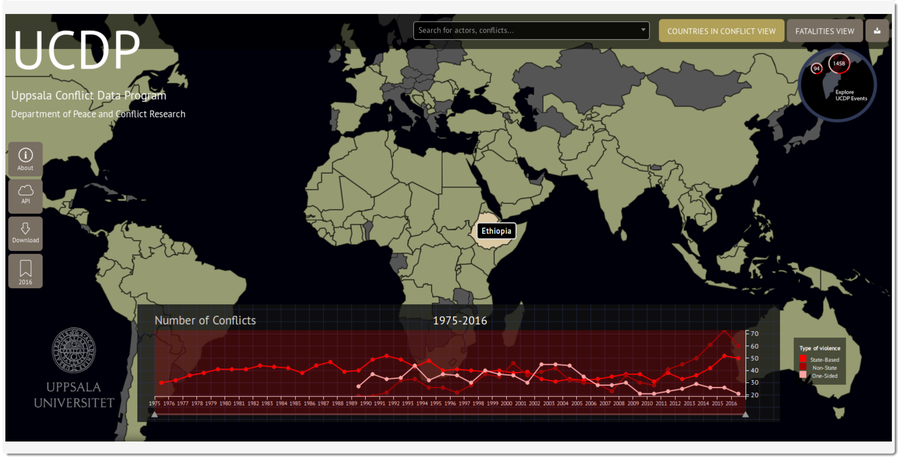
Classification task

We monitor changes in the local semantic neighborhoods of country names, applying it to the downstream task of predicting changes in the state of conflict:

1. Nothing has changed in the country conflict state year-to-year (class '**stable**');
2. Armed conflicts have escalated in the country year-to-year (class '**war**');
3. Armed conflicts have calmed down in the country year-to-year (class '**peace**').

Example: given two distributional models trained on news texts from 2002 and 2003, predict in what direction the conflicts state moved in these years in Senegal (the correct answer is 'it escalated')?

UCDP data



Data representation

Set of data points equal to the differences ($\boldsymbol{\delta}$) between the location's conflict state in the current year and in the previous year, 832 points in total (52 locations $\times$ 16 years).
As an example, for Congo, the transition from 2001 to 2002 was accompanied by the ending of armed conflicts. Thus, for the data point 'congo_2002', $\boldsymbol{\delta = 0 - 1 = -1}$. Then, there were no changes (each new $\boldsymbol{\delta}$ is 0) until 2006, when armed conflicts resumed with the intensity of 1. Thus, for the 'congo_2006' data point, $\boldsymbol{\delta = 1 - 0 = 1}$.
Then, $\boldsymbol{\delta}$ values were transformed to classes:

$$\boldsymbol{class} = \begin{cases} \boldsymbol{war} & \text{if } \boldsymbol{\delta \geq 0.5} \\ \boldsymbol{peace} & \text{if } \boldsymbol{\delta \leq -0.5} \\ \boldsymbol{stable} & \text{otherwise} \end{cases}$$

Gold standard

As a ground truth, we use the UCDP/PRIO Armed Conflict Dataset (`http://ucdp.uu.se/`) maintained by the Uppsala Conflict Data Program and the Peace Research Institute Oslo: a manually annotated geographical and temporal dataset with information on armed conflicts, in the time period from 1946 to the present [Gleditsch et al., 2002]. The resulting test set mentions 52 unique locations and 673 unique armed conflicts. The task was to predict the conflict state difference for these locations year-to-year.

Diachronic models

To train distributional word embedding models, we used the Continuous Bag-of-Words algorithm proposed in [Mikolov et al., 2013]. Models were trained on Gigaword [Parker et al., 2011] texts in 3 time-representation modes:

1. yearly models, each trained from scratch on the corpora containing news texts from a particular year only (separate);
2. yearly models trained from scratch on the texts from the particular year and all the previous years (cumulative);
3. incrementally trained models (incremental).

Methods

To actually detect semantic shifts for the word $\boldsymbol{w_q}$, one can either:

1. align two models ($\boldsymbol{M_{cur}}$ and $\boldsymbol{M_{prev}}$) using the orthogonal Procrustes transformation, and then measure cosine similarity between the $\boldsymbol{w_q}$ vectors in both models, as proposed in [Hamilton et al., 2016b];
2. alternatively, define a set of anchor words related to the semantic categories we are interested in, and then measure the 'drift' of $\boldsymbol{w_q}$ towards or away from these 'anchors' in $\boldsymbol{M_{cur}}$ compared against $\boldsymbol{M_{prev}}$.

The anchor words method can provide information about the exact direction of the shift. It can be quantified in 2 ways:

1. for each anchor, calculate its *cosine similarity* against $\boldsymbol{w_q}$ in $\boldsymbol{M_{cur}}$ and $\boldsymbol{M_{prev}}$ (Sim);
2. as above, but use the *position of each anchor in the models' vocabulary* sorted by similarity to $\boldsymbol{w_q}$ (Rank).

These methods produce two vectors $\boldsymbol{\vec{R}_{prev}}$ and $\boldsymbol{\vec{R}_{cur}}$, corresponding to the models $\boldsymbol{M_{cur}}$ and $\boldsymbol{M_{prev}}$, with the size is equal to the number of the anchor words. Then we can either:

1. calculate the *cosine distance between these 'second-order vectors'* (SimDist or RankDist);
2. element-wise *subtract* $\boldsymbol{\vec{R}_{prev}}$ *from* $\boldsymbol{\vec{R}_{cur}}$ to get the idea of whether $\boldsymbol{w_q}$ drifted towards or away from the anchors (SimSub or RankSub).

These features are then fed to SVM classifier.

Anchor words

(adopted from the search strings UCDP uses to filter news texts):
kill, *die*, *injury*, *dead*, *death*, *wound*, *massacre*.
Expanded list adds the nearest neighbors of these words in the CBOW model trained on the full Gigaword (26 words total).

Overall results

Approach	Separate	Cumulative	Incremental
Procrustes	0.15	0.24	0.29
	Basic word list		
SimDist	0.27	0.17	0.25
SimSub	0.31	0.26	0.26
RankDist	0.28	0.19	0.23
RankSub	0.26	0.22	0.21
	Expanded word list		
SimDist	0.25	0.18	0.23
SimSub	0.35	0.31	0.29
RankDist	0.24	0.20	0.28
RankSub	**0.36**	0.30	0.32

Macro-F1 measure of ternary classification

The test set is available

`http://ltr.uio.no/~andreku/armedconflicts/ucdp_conflicts_1994_2010_testset.tsv`

Best model detailed results

Class	Precision	Recall	F1
Peace	**0.13** (0.06)	**0.29** (0.06)	**0.18** (0.06)
Stable	0.80 (0.79)	0.58 (**0.82**)	0.67 (**0.80**)
War	**0.17** (0.12)	**0.33** (0.08)	**0.22** (0.10)

Detailed performance of the best model (results of weighted random guess baseline in parenthesis).
The 'shifting' classes War and Peace constitute 10% and 11% of the data points respectively.

Conclusion

- Tracing actual real-world events by detecting 'cultural' semantic shifts in distributional semantic models is a difficult task.
- The approaches proposed in the previous work for large-scale shifts observed over decades or even centuries are not very successful in this more fine-grained task.
  - [Hamilton et al., 2016b] report almost perfect accuracy for the Procrustes transformation when detecting the direction of semantic change. However, our time periods are much more granular and we attempt to detect subtle associative drifts (often pendulum-like) rather than full-scale shifts of the meaning.
- The proposed 'anchor words' method outperforms previous work approaches by large margin, but it still achieves a macro F1 measure of only 0.36 on the task of ternary classification ('stable', 'escalating', 'calming down').
- See also our forthcoming EMNLP'17 paper on tracing semantic relations in diachronic models.