Mining Sentiments and Arguments in United Nations Security Council (UNSC) Speeches

"Exploring the UNSC political speech corpus"

Atreya Shankar, Juliane Hanel
Cognitive Systems (M.Sc.)

PM: Mining Sentiments and Arguments
University of Potsdam, WiSe 19/20
Prof. Dr. Manfred Stede

February 4. 2020

Table of Contents

Recap: Objectives

- **General objective:** Explore mining various components of the UNSC political speech corpus
- **Approach 1:** Using sentiment and subjectivity analysis
- **Approach 2:** Mining argumentation structure of speeches

What have we achieved so far?

- **Data preparation:** Organizing more than 65k unprocessed speeches including metadata
- **Research:** Wrapping our heads around state-of-the-art methods
- **Coding:** Implementing fundamental ideas and prototypes
- **Realizations:** How feasible are our objectives?

Sentiment Analysis - Methodologies

Sentiment Analysis:

- *Vader* for comparative sentiment analysis ([-1,1] from negative over neutral to positive)
- *TextBlob* did not perform well, a joint approach of the two frameworks is not feasible

Subjectivity Analysis:

- *TextBlob* for subjectivity analysis ([0,1] from objective to subjective)
- Possible extension with *MPQA* subjectivity lexicon (Wilson, Wiebe, and Hoffmann, 2005)

Sentiment Analysis

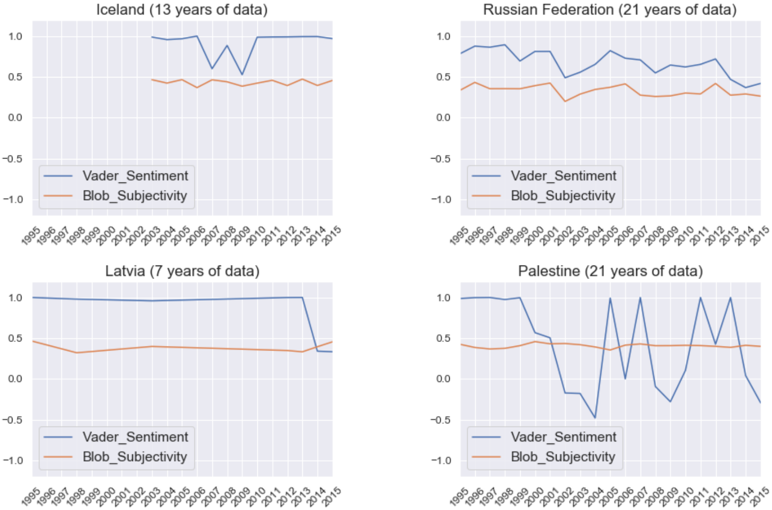

Figure 1: Examples of the sentiment analysis
⟶ **What is worth investigating?**

Example - Core Nations

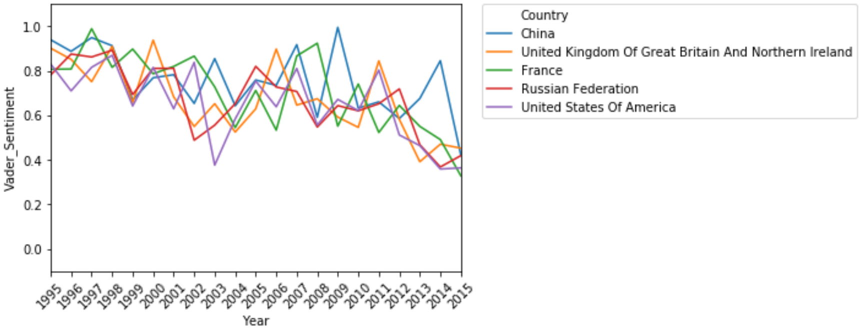

Figure 2: Sentiment of the UNSC core nations over time

Example - Core Nations

What is worth investigating?

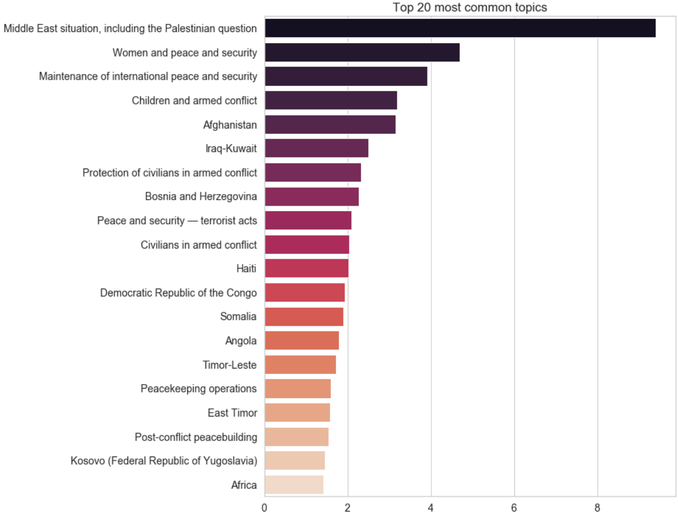

Figure 3: Relative frequency of topics (overall)

Example - Core Nations

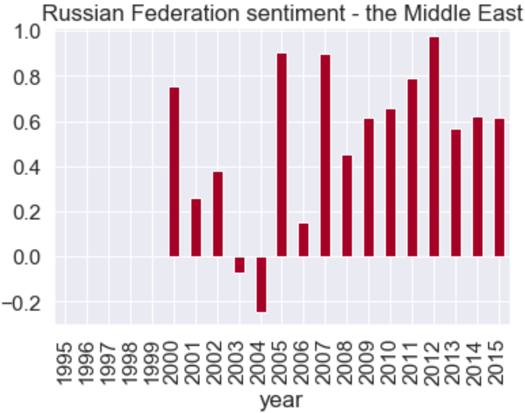

Figure 4: Russia as example for the sentiment in core nation speeches concerning the Middle East situation

⟶ 2002: *Road map for peace* was founded

Evaluation?

Positive Words	Negative Words
important	conflict
urgent	dangerous
acceptable	confrontation
support	stop
	violence

```
Mr. Denisov (Russian Federation)(spoke in Russian):
The Minister for Foreign Affairs of Russia
today discussed the Palestinian–Israeli conflict in
telephone conversations with the Palestinian Authority
and with the Minister for Foreign Affairs of Israel. He
emphasized, in the course of those discussions, that the
important thing now is to take urgent steps in order to
put an end to the current dangerous situation of
confrontation. Since, as we understood it, the main
objective of the draft resolution was to stop the
violence, it was, in principle, acceptable to us –
although in our view it should have been more
balanced. We therefore proposed to our Algerian
colleagues that they make several changes to the draft,
and since those proposals were partially taken into
account, we took the decision to support the draft
resolution.
```

Figure 5: Relevant words in Vader vs. manual sentiment annotation

Further Steps

- Deciding on evaluation method (Words, sentences, speeches? One or multiple opinions?)
- Deciding on the analysis scope (Core nations? Specific topics?)
- Improving performance of subjectivity analysis by integrating lexica
- Organizing and implementing the final ideas

Recap: Argumentation Mining

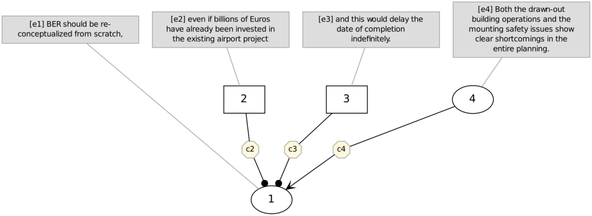

Figure 6: Example of a complete argumentative structure of a short text (Peldszus and Stede, 2015)

- Argumentation structure requires claims and premises; which are usually assembled into a tree structure
- Specify support/attack nature of claims and premises; *omitted in our project*
- Train argumentation classifier on US Election Debate corpus, apply on UNSC corpus

US Election Debate Corpus

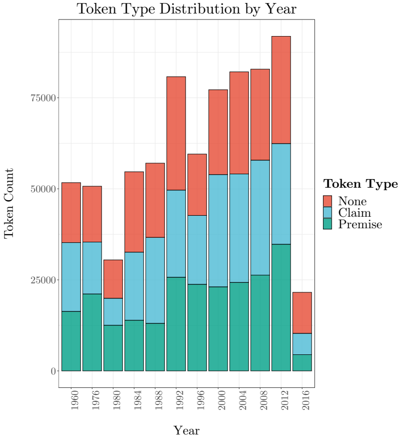

Figure 7: Token type distribution in the US election debate corpus by year

- US election debate corpus consists of 42 annotated presidential debate sessions from 1960 until 2016 (Haddadan, Cabrio, and Villata, 2019)
- Pre-processing of character spans yields token-based annotations
- Tokens can either be of class "None" (N), "Claim" (C) or "Premise" (P)
- **Task 1** $\Longrightarrow$ sequence tagging of tokens to N, C or P
- **Task 2** $\Longrightarrow$ argumentation structure via adjacency matrix

(AL)BERT as Input Encoder

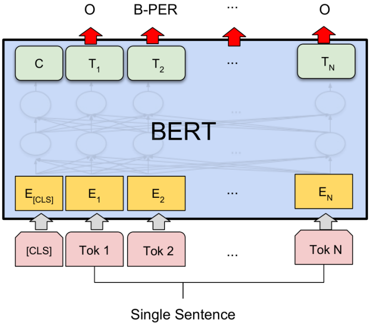

Figure 8: Sample schematic of using BERT for Natural Entity Recognition (NER) task (Devlin et al., 2018)

- **B**idirectional **E**ncoder **R**epresentations from **T**ransformers (BERT) and its variants have shown SOTA performance on various NLP tasks (Devlin et al., 2018)
- $\text{BERT}_{\text{base}} \implies \sim 108\text{M}$ parameters
- Unrealistic to fine-tune on single NVIDIA GeForce GTX 1080 Ti with 12 GB RAM (Uni-Potsdam available hardware)
- Alternative is the "Lite" version of BERT, known as ALBERT (Lan et al., 2019)
- $\text{ALBERT}_{\text{base}} \implies \sim 12\text{M}$ parameters
- $O(N^2)$ space complexity wrt. sequence length for attention mechanism

US Election Debate Corpus Pruning

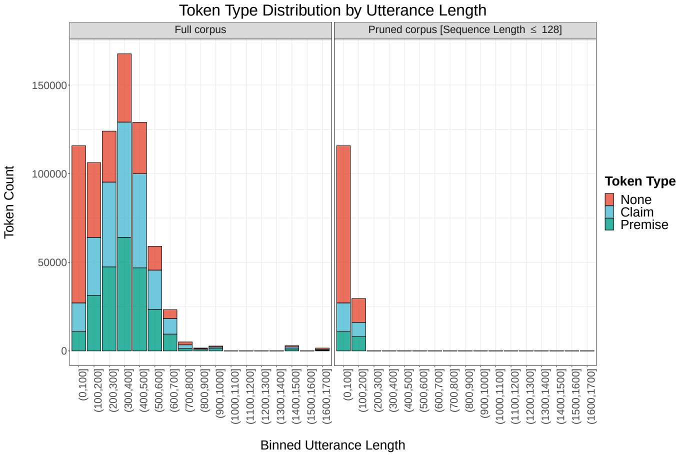

Figure 9: Token type distributions based on binned utterance length; based on full (left) and pruned corpus (right)

- Keep sequences where $N_{tokens} \leq 128$; corpus size reduced from 6559 $\longrightarrow$ 4594 utterances

Fine-tuning ALBERT

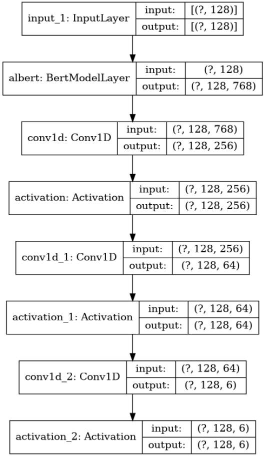

Figure 10: Sample graph for ALBERT encoder with 1D-Convolution decoder

- To fine-tune ALBERT on the argument classification task, we must decode ALBERT output to our supervised output
- The following architectures were tested through a grid-search with various learning profiles and normalization techniques; albeit with a fixed batch size of *48* to prevent OOM issues

i. 3 layer time-distributed dense (TD-Dense)

ii. 3 layer fully 1D-Convolutional (1D-Conv)

iii. 2 layer LSTM decoder (Stacked-LSTM)

Model Performance on Task 1 (Sequence Tagging)

Model	Train F_1	Test F_1	Test F_1[N]	Test F_1[C]	Test F_1[P]
TD-Dense	0.674	**0.643**	0.917	0.5174	**0.495**
1D-Conv	0.744	0.610	0.899	0.513	0.419
Stacked-LSTM	0.609	0.629	0.899	**0.569**	0.419

Table 1: Model performance summary based on train and test set F_1 scores

- Similar performances with $\sim 60\%$ F_1 scores on test set
- High F_1 scores across N-token classification, as it is a majority class
- TD-Dense shows the best overall performance; with best performance in P-token classification
- Stacked-LSTM has second-best overall performance; with best performance in C-token classification

Further steps

1. To prevent data loss from corpus pruning, use gradient accumulation in optimizers [1] to allow for smaller local batch-sizes and potentially longer sequence lengths; without affecting global batch-sizes
2. Improve train/test/validation data splits to be representative of overall corpus
3. Develop graph neural network to create explicit argumentation structure from annotated tokens $\Longrightarrow$ hypothesis is that multi-task setting will improve performances of both tasks
4. Run best classifier on short sequences in UNSC corpus to manually grade performance

[1] https://github.com/run-ai/runai

Bibliography I

Darmstadt, TU. *Subjective Verbs Lexicons*. URL: https://www.informatik.tu-darmstadt.de/ukp/research_6/data/sentiment_analysis/subjective_verbs_lexicons/index.en.jsp.

Devlin, Jacob et al. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv: 1810.04805 [cs.CL].

Groza, Adrian and Oana Popa (Sept. 2016). "Mining arguments from cancer documents using Natural Language Processing and ontologies". In: pp. 77–84. DOI: 10.1109/ICCP.2016.7737126.

Bibliography II

Haddadan, Shohreh, Elena Cabrio, and Serena Villata (July 2019). "Yes, we can! Mining Arguments in 50 Years of US Presidential Campaign Debates". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, pp. 4684–4690. DOI: 10.18653/v1/P19-1463. URL: https://www.aclweb.org/anthology/P19-1463.

Hutto, C.J. and Eric Gilbert (Jan. 2015). "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text". In:

Lan, Zhenzhong et al. (2019). *ALBERT: A Lite BERT for Self-supervised Learning of Language Representations*. arXiv: 1909.11942 [cs.CL].

Bibliography III

Peldszus, Andreas and Manfred Stede (2015). “An annotated corpus of argumentative microtexts”. In: *Argumentation and Reasoned Action: Proceedings of the 1st European Conference on Argumentation, Lisbon*. Vol. 2, pp. 801–816.

Potash, Peter, Alexey Romanov, and Anna Rumshisky (2016). *Here's My Point: Joint Pointer Architecture for Argument Mining*. arXiv: 1612.08994 [cs.CL].

Schönfeld, Mirco et al. (2019). *The UN Security Council debates 1995-2017*. arXiv: 1906.10969 [cs.DL].

Stab, Christian and Iryna Gurevych (2017). “Parsing argumentation structures in persuasive essays”. In: *Computational Linguistics* 43.3, pp. 619–659.

Bibliography IV

Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann (Oct. 2005). "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis". In: DOI: 10.3115/1220575.1220619.